ABSTRACT

BUHRMAN, GREGORY KALE. Application of the Multiple Solvent Crystal Structures Method to Analyze the Protein Binding Surface of H-Ras Protein (Under the direction of Carla Mattos)

H-Ras is a member of the small, monomeric GTPase protein superfamily. H-Ras functions as a 'molecular switch', using nucleotide dependent conformational changes to relay signals in a number of signal transduction pathways. Mutations in codons 12, 13 and 61 creates an oncogenic version of the protein which does not hydrolyze GTP, resulting in the constitutive activation of downstream effector proteins. Ras proteins participate in multiple protein : protein interactions in the cell, making Ras a good candidate protein to extend the Multiple Solvent Crystal Structures method (MSCS) to the analysis and prediction of protein binding surfaces. MSCS involves solving the crystal structure of the protein after soaking the protein crystal in a variety of organic solvent molecules. Replacing an aqueous solvent with an organic solvent affects the Ras protein structure in several ways. The disordered Switch II region of Ras is ordered in the presence of 2,2,2-trifluoroethanol or 1,6-hexanediol. Polar interactions that stabilize the ordered switch are enhanced in the presence of hydrophobic co-solvents. This suggests that hydrophobic solvents can be used in general to order short biologically relevant segments of disordered regions in protein crystals. We have used MSCS to study two crystal forms of active H-Ras bound to a nonhydrolyzable GTP analog (GMPPNP). We have also solved the structure of an oncogenic mutant of H-Ras (Q61L) in a non-canonical crystal form. This crystal form of H-Ras shows a new conformation for the flexible Switch II region that is not affected by crystal packing forces. This provides a structural explanation for the oncogenic properties of the Q61L mutation, showing that the Q61L mutation stabilizes a non-catalytic conformation of Switch II. MSCS analysis of Ras identifies the known Ras-effector binding domain as a site of protein: protein interaction and predicts a new protein binding site that is located in a large, solvent exposed pocket between Switch II and helix 3. In applying MSCS to the Ras protein, we show that by using polar organic solvent molecules as probes, we can identify binding sites that are highly charged and dynamic.

Application of the Multiple Solvent Crystal Structures Method to Analyze the Protein Binding Surface of H-Ras Protein

By

Gregory Kale Buhrman

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Molecular and Structural Biochemistry

Raleigh, North Carolina

2004

DEDICATION

I dedicate this dissertation to my wife, Tiffany Wynn Wagner. Without her love and support, nothing worthwhile that I've done in this life would have been possible, or meaningful. She has been my constant companion throughout my Masters and Ph.D. program. First, she moved with me to the little town of Emmitsburg, MD, where she worked as a Family Therapist for a residential drug treatment facility and I worked on a Masters degree at Hood College. Then, she left her job and family to follow me to North Carolina, so I could selfishly follow my dreams of learning protein crystallography. She has put up with my long hours and small income for as long as I can remember. She has had the strength of personality to support my dreams, while pursuing hers as well. Because of her, we have a lovely home, populated with one stubborn dog, one independent cat, one fish, and two amazing children, Eva Jane and Eliza Rae Buhrman. She is my compass and the lens through which I view my life. Without her encouragement and support, I never would have completed these studies. She believes in me when no one else does, and brings me back to earth when I get too full of myself. A friend once told us our life would be one of 'happy destitution'. That has certainly been the case, and I wouldn't trade away one minute of it. Thank you Tiffany, I love you with all of me.

BIOGRAPHY

I am the eldest son of Charles and Harriet Buhrman. My father was an electrical engineer, before an unfortunate accident left him partially paralyzed and unable to continue what looked to be a promising engineering career. He now works as a much loved and respected substitute teacher in Pennsylvania. My mother, when she finished raising us, returned to her teaching career as a Montessori school teacher. I was born in Bremerton, Washington, but spent most of my life growing up in Westminster, Maryland. I received my undergraduate B.S. in Molecular Biology from the University of Pittsburgh and a M.S. in Biomedical Research, from Hood College in Frederick, MD. The mentor for my Masters thesis, entitled 'Analysis of the DNA Binding Properties of Bacteriophage P1 C1 Repressor and Bof Modulator Proteins Utilizing a Comparative SELEX Technique', was Dr. Craig Laufer. Dr. Laufer provided me with my first real opportunity in science, hiring me as a full-time graduate student in the Biology Department. A small department, with only two other full-time graduate students, provided a unique opportunity for independent teaching and research and to develop close-ties with faculty, fellow graduate students and undergraduates. The positive experiences I had at Hood College encouraged me to pursue a Ph.D. at North Carolina State University under the mentorship of Dr. Carla Mattos in the Department of Molecular and Structural Biochemistry. My research at N.C. State has primarily involved using protein x-ray crystallography to investigate protein binding surfaces.

ACKNOWLEDGMENTS

First and foremost, I want to acknowledge my mentor, Dr. Carla Mattos. I new nothing of protein x-ray crystallography when I joined her lab, other than that I wanted to do it. Carla took me under her wing and patiently walked me through every aspect of protein crystallography. I especially enjoyed walking through structures with Carla. She has a refined appreciation for protein structure that is absolutely contagious. I also want to acknowledge Dr. Andre White and Dr. Bob Rose for many helpful conversations. Dr. White is the most meticulous crystallographer I've met. He taught me the intricacies of command-line Denzo and gave me a healthy respect for a detail-oriented approach to crystallization and data collection. Dr. Rose was not on my committee, but he has a natural enthusiasm for science in general, and was particularly supportive with the phosphatase project. I also need to acknowledge the rest of the Mattos Lab for their constant support and encouragement. Dr. Vesna Serrano got me started with H-Ras and my fellow graduate students, Nate Nicely and Michelle Dechene, were a constant source of help, fun and laughter. Also, thanks to Dr. Paul Swartz for so adeptly taking over the care and maintenance of the x-ray beam. Dr. Sharon Cambell kindly provided the clones and initial protein purification protocols for H-Ras. Finally, I want to thank the members of my committee for their support and valued criticisms of my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

The biological functions of many proteins depend on their ability to interact with other proteins to form complexes. Determining the region of a protein that participates in a protein-protein interface is an important first step to understanding how many proteins function. One common and informative approach is to solve the structure of the protein complex by NMR or X-ray crystallography. However, obtaining diffraction quality crystals of protein complexes is often a difficult task, even when diffraction quality crystals of the unbound proteins can be obtained. This is very often the case with transient protein hetero-complexes, which are made up of different proteins that can exist in both the bound and unbound state and have a range of binding affinities. Examples of transient protein hetero-complexes include enzyme-inhibitor complexes, protein-receptor complexes and signal transduction partners. The MSCS (Multiple Solvent Crystal Structures) method has already been used to successfully map the protein binding surface of several proteases which bind and cleave small, peptide substrates (Ringe & Mattos 1999). In this work, we extend the MSCS technique to mapping the protein binding surface of Ras, a protein which forms protein complexes with a variety of signal transduction partners.

I. History and Background of Ras Oncogenesis

In 1964, Harvey discovered a murine sarcoma virus that is responsible for tumor production in mice and rats (Harvey 1964). It was later shown that the viral gene responsible for tumor production encodes a protein referred to as p21 or Ras (Shih, Weeks, et al. 1979). In 1982, Furth et. al. isolated monoclonal antibodies that immunoprecipitated both p21

protein from cells transformed with the viral *ras* gene from Harvey and Kirsten murine sarcoma viruses as well as protein encoded by a cellular Harvey-*ras* gene (Furth, Davis, et al. 1982). Also in 1982, Channing Der showed that cellular homologs to the viral H-*ras* and K-*ras* gene are found in human bladder and lung carcinoma cell lines and are responsible for transformation of these cell lines (Der, Krontiris, et al. 1982). In 1983, Shimizu et. al. found that a third isoform of the *ras* gene, N-*ras*, was responsible for transformation of neuroblastoma cells (Shimizu, GLYdfarb, et al. 1983). These findings were consistent with the previous work of Harold Varmus and J. Michael Bishop, who demonstrated that the oncogene of Rous Sarcoma Virus (src) also has a cellular homolog. These cellular homologs of viral oncogenes are mutant forms of normal proteins involved in normal cellular growth and differentiation. Mutations at codons 12, 13 or 61 in Ras were identified as transforming mutations in the H, K and N – *ras* genes. These mutations are found in over 30% of cancers of the pancreas, colon, lung, thyroid and in leukemia, (reviewed in Bos 1989). This high incidence of mutant Ras in human cancer makes Ras a prime target for anti-cancer therapies, (reviewed by Cox & Der 2002).

II. The GTPase Cycle

H-Ras is part of a superfamily of small, monomeric GTPases that function as molecular switches in the cell. All GTPases undergo a guanine nucleotide dependent conformational change, switching from an active conformation when bound to GTP and an inactive conformation when bound to GDP (Wittinghofer & Pai 1991). The active form of the Ras GTPase interacts with downstream effector proteins, which in turn activate kinase cascades that activate transcription factors responsible for normal growth and differentiation.

GTP hydrolysis to GDP converts the protein to an inactive conformation that no longer interacts with downstream effector proteins, switching the signal pathway off. Oncogenic mutants of Ras have reduced GTPase activity which results in uncontrolled activation of the signaling pathways, leading to uncontrolled cellular growth and oncogenesis (Krengel, Schlichting, et al. 1990). The Ras GTPase cycle is a paradigm used to describe all small, monomeric GTPases (see Figure 1).

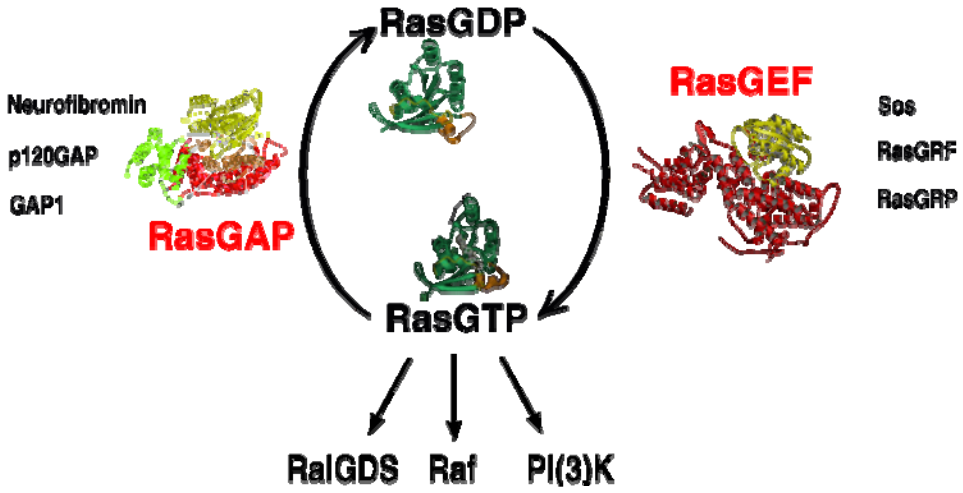


Figure 1. The Ras GTPase Cycle. Ras cycles between an active, GTP-bound form and an inactive, GDP bound form. The active form interacts with downstream effector proteins to activate multiple signal transduction cascades. The active form is converted to the inactive form by GTPase Activating Proteins (GAPs). The inactive form is converted to the active form by Guanine Nucleotide Exchange Factors (GEFs). This Figure is reprinted with permission from the Max Planck Institute of Molecular Physiology departmental web page found at: http://www.mpidortmund.mpg.de/departments/dep1/gtpase/gtpasedetail.php3.

The structures of Ras and several other GTPases have been solved, in the GTP (using nonhydrolyzable GTP analogs) and GDP bound forms, (reviewed by Corbett & Alber 2001). These studies have shown that despite average sequence homology, the structural conservation is very good. Because Ras was one of the first GTPases to be characterized structurally and functionally, Ras has served as a paradigm for characterizing all the GTPases and is considered the 'canonical' GTPase to which other GTPases are compared.

Acting alone, Ras would be an inefficient molecular switch. The intrinsic GTPase activity of Ras is relatively low (4.2×10^{-4} s^{-1}) (John, Schlichting, et al. 1989). The rate of GTP hydrolysis is accelerated by roughly 5 orders of magnitude by GTPase Activating Proteins (GAPs) (Gideon, John, et al. 1992). GAPs for Ras include the p120 GAP protein and neurofibromin (NF1). GAPs accelerate GTP hydrolysis by stabilizing the transition state of the hydrolysis reaction (Mittal, Ahmadian, et al. 1996). Converting the inactive GDP bound Ras to the active GTP bound form requires nucleotide release. Nucleotide dissociation rates for GDP and GTP are very low for free Ras (3.4×10^{-4} s^{-1})(John, Sohmen, et al. 1990). Nucleotide release is accelerated roughly six fold by Guanine Nucleotide Exchange Factors (GEFs). GEFs for Ras include SOS and RasGEF. GEFs accelerate nucleotide release by stabilizing an open conformation that does not bind nucleotide (Boriack-Sjodin, Margarit, et al. 1998). Nucleotide free Ras can then become activated by binding to GTP. GEFs are activated in response to activation of receptor tyrosine kinases and other signaling molecules.

Activated Ras interacts with multiple downstream effector proteins to activate a number of signal transduction pathways (Campbell, Khosravi-Far, et al. 1998). Some pathways have been fully elucidated, but much work remains to complete our understanding

of Ras signaling. One pathway that has been characterized is the Ras>Raf>MEK signal transduction cascade, see Figure 2. In this pathway, Ras binds to Raf kinase and brings Raf to the plasma membrane where Raf is activated and in turn, activates a Mitogen Activated Protein (MAP) kinase cascade. This MAP kinase cascade activates nuclear transcription factors, such as Elk-1, by phosphorylation. Elk-1 activates the transcription of genes responsible for cellular growth and differentiation. Other Ras activated downstream effector proteins include PI3Kinase, RalGDS (Guanine nucleotide Dissociation Stimulator), Byr2 (a MEKK/Raf homolog in *S. pombe*), AF6 (*drosophila*), RIN1 (Ras INterference), MEKK1 (Mitogen activated protein kinase Kinase Kinase 1) , PKC zeta and KSR Raf-related kinase, (Kinase Suppressor of Ras, *drosophila* protein with murine and human homologs), reviewed in (Joneson & Bar-Sagi 1997), (Herrmann & Nassar 1996) and (Marshall 1995). Putative Ras effector proteins have been identified by a number of methods, including coimmunoprecipitation, yeast two-hybrid experiments, genetic screens and sequence similarity with known Ras binding domains. Defining functional characteristics of Ras effector proteins include the ability to bind Ras in a GTP dependent manner, the ability to compete with known Ras binding proteins, inhibition or activation of known Ras signaling pathways, co-localization with Ras and an increase in activity concomitant with Ras activation. Putative Ras effector proteins which have been identified by homology or *in vitro* experiments need to be characterized by *in vivo* experiments which demonstrate the physiological relevance of the Ras –effector interaction. Defining putative Ras effector proteins remains an ongoing area of active research.

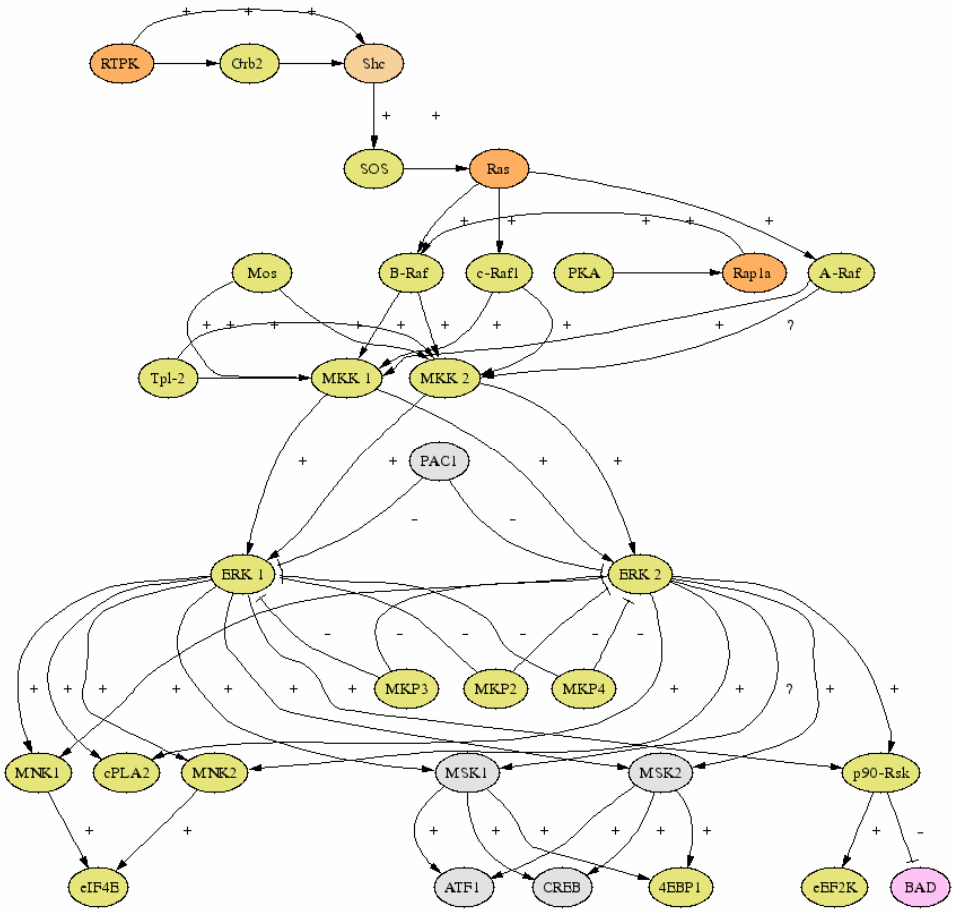


Figure 2. Ras involvement in the ERK1,ERK2, MAPK canonical pathway. Ras interacts with isoforms of Raf kinase, which in turn, activates either MKK1>ERK1 or MKK2>ERK2. Gary L Johnson, ERK1/ERK2 MAPK Pathway *Sci. STKE* (Connections Map, as seen December 2004), http://stke.sciencemag.org/cgi/cm/stkecm;CMP_10705.

Several of the complexes between downstream effector proteins and Ras have been characterized structurally, reviewed by (Corbett & Alber 2001). This structural information indicates several key concepts. Most effector proteins interact with a subset of residues in and around the switch regions. Binding of effector proteins often result in a disorder to order transition in the switch regions (primarily switch II). NMR experiments have shown that the active form of Ras, bound to either GTP or a GTP analog, is a highly dynamic protein which oscillates between a number of stable conformations in solution (Geyer, Schweins, et al. 1996). Downstream effector proteins often bind to a subset of the conformations adopted by activated Ras protein. Downstream effector proteins interact with Ras protein via several different binding motifs. Binding and activation of effector proteins are separate functions which may be due to distinct Ras-effector interactions.

III. The Ras Superfamily

Ras is the canonical member of a larger superfamily of GTPases, which is evolutionarily conserved in eukaryotes from yeast to man. The Ras superfamily includes over 100 human proteins, which can be sub-classified into the Ras, Rho, Rab, Arf, and GNA families, see Figure 3. The Ras subfamily alone includes at least 35 human proteins, reviewed in (Colicelli 2004). Sequence analysis of the Ras superfamily has defined sequence motifs in several regions of conserved residues, referred to as ‘G boxes’. The G1 box coincides with the phosphate binding loop, or P-loop. The G2 box primarily consists of Switch I residues. The G3-G5 boxes consist of conserved residues that interact, either directly or indirectly, with the guanine base or the conserved magnesium ion. The superfamily is functionally defined as GTP hydrolysis-coupled signal transduction relay

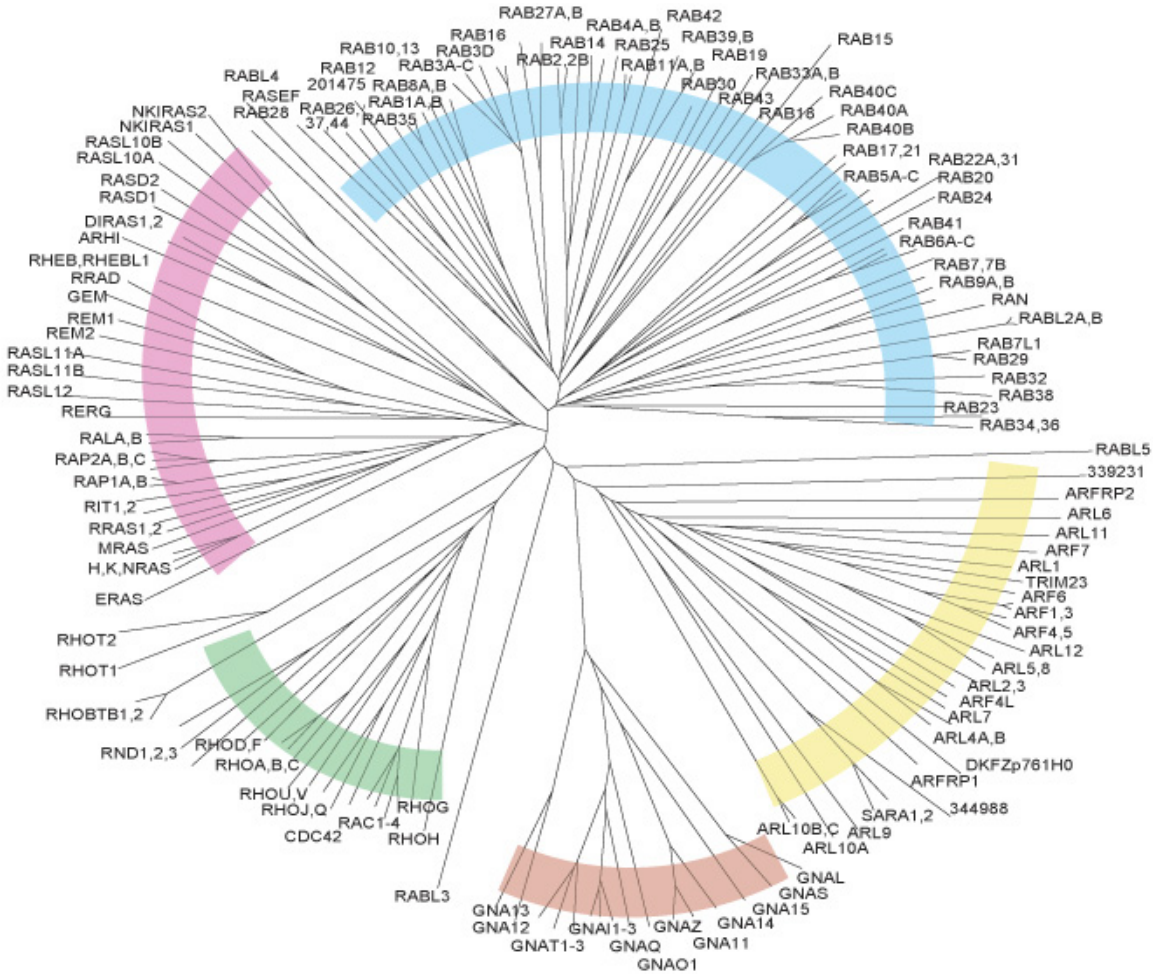


Figure 3. The Ras superfamily consists of five subfamilies, Rho (green) GNA (brown), ARF (yellow), Rab (cyan), and Ras (magenta), shown as an unrooted phylogenetic tree (Collicelli 2004).

proteins. These proteins play roles in a number of diverse cellular functions. An important aspect of the Ras superfamily is the significant degree of crosstalk that exists between family members. One example of crosstalk is seen in the ERK signal transduction pathway shown in Figure 2. Activated forms of both Ras and Rap can bind and activate the beta isoform of Raf kinase, (reviewed by Stork 2003). In this example, cross talk occurs via competition between GTPases for a common downstream effector protein. Another example of complex interactions between GTPases occurs between Ras and Ral. A number of Ral specific GEFs are downstream effector proteins which are activated by Ras. In this example, communication between Ras and Ral occurs via a common protein (RalGEF) that acts as a GEF for Ral, but is activated by Ras, even though the structural similarity of Ras and Ral is fairly high (Nicely, Kosak, et al. 2004). Another example of this kind of interaction involves Tiam1, a Ras binding protein, which also functions as a GEF for Rac. This complex behavior may be explained by the structure of the ternary complex of Ras-SOS-Ras-GTP, which illustrates what may be a common mechanism of GEF activation by active Ras (Margarit, Sondermann, et al. 2003).

IV. Structure and Function of Ras

The Ras superfamily is defined structurally by a number of conserved motifs that confer specific functionality to the GTPase proteins. The H-Ras protein is 189 residues long. The catalytic core of the protein, which was used in our studies, consists of residues 1-166. This C-terminal truncation of the wild type protein contains the conserved structural features of the GTPases and is functional as shown in biochemical assays (John, Schlichting, et al. 1989). The C-terminus is isoprenylated through a CAAX motif (where A is any aliphatic residue), which is necessary for membrane localization of the protein. It is disordered in all

structures of H-Ras and causes problems in bacterial over expression, purification and crystallization. Proteins with truncated C-termini are effective dominant negative inhibitors of the Ras>Raf>MEK>ERK signal pathway, binding efficiently to Raf kinase, but preventing membrane translocation and activation. Additionally, chimeric proteins of Raf plus the C-terminus CAAX motif of Ras are directed to the plasma membrane and activated in a Ras independent manner (Stokoe, Macdonald, et al. 1994). The C-terminus contributes to the interaction of H-Ras with the cysteine rich domain of Raf kinase, but does not result in any detectable conformational changes in H-Ras (Thapar, Williams, et al. 2004).

The catalytic core (1-166) consists of six beta strands, five alpha helices and 10 loop regions. Important structural features which will be discussed in greater depth include: Switch I (30-37), Switch II (59-72), the P-loop (11-14), the guanine nucleotide binding pocket (116-119,145-147) and the core and extended Ras effector regions (32-40, 20-48). Regions of interest are highlighted in the cartoon representation of the Ras structure, shown in Figure 4.

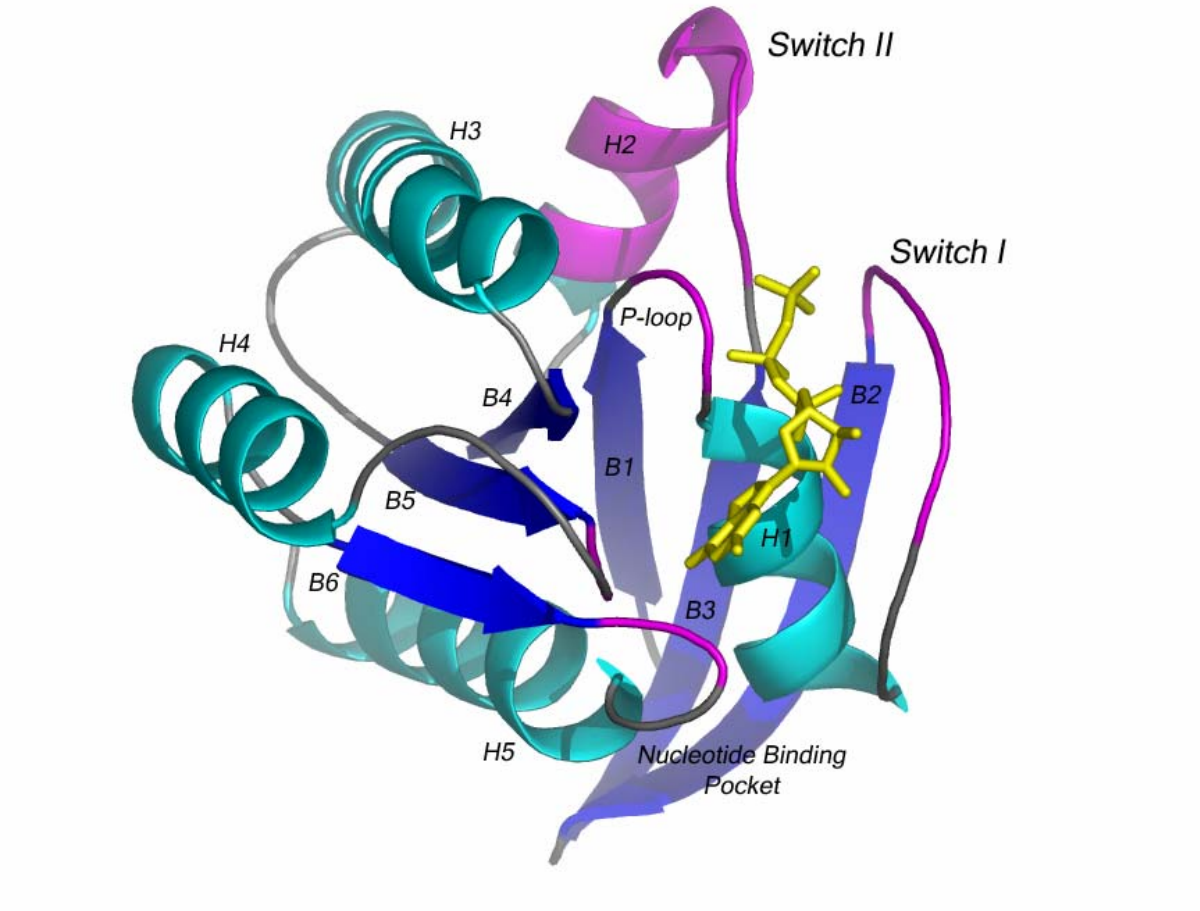


Figure 4. Structure of the catalytic domain of H-Ras bound to GMPPNP (a nonhydrolyzable GTP analog). The P-loop, Switch I, Switch II and Nucleotide binding pocket are colored magenta.

A. Switch I

The molecular switches are defined as residues that undergo a conformational change from the GTP to GDP bound state. Most literature references identify the Switch I of Ras as residues 30-37 (Asp Glu Tyr Asp Pro Thr Ile Glu). Importantly, the Switch I region coincides

with the core Ras effector binding region (32-40) and demonstrates how conformational changes from the active to inactive state change effector binding affinity. Switch I is an inherently flexible region that undergoes an order to disorder transition upon GTP hydrolysis. NMR experiments show that in the GTP bound state, Switch I residues oscillate between two stable conformations (Geyer, Schweins, et al. 1996). Molecular dynamic simulations of the GTP bound form of Ras also detect a large motion involving residues 27-37(Soares, Miller, et al. 2001). Recent studies indicate that the extent of conformational change that occurs upon GTP hydrolysis, and therefore the extent of Switch I, varies between GTPases. For example, the structures of both GDP and GTP forms of RalA shows that Switch I, when it is not stabilized by crystal contacts, is disordered in both the GDP and GTP forms (Nicely, Kosak, et al. 2004) . Because Switch I coincides with the core Ras effector binding region, changes in the conformational flexibility of different GTPases is likely to play an important role in determining differences in specificity and affinity between GTPases. Glu31 contributes to binding specificity with Raf kinase. Glu31 is a Lys in Rap and experiments have shown that this charge reversal is sufficient to change binding specificity between Ras and Rap for Raf kinase (Nassar, Horn, et al. 1996). The structure of Rap1a (E30D,K31E) double mutant bound to the Ras Binding Domain of Raf shows Glu31 interacting with Raf Lys84 and is taken as a mimic of the Ras-Raf interaction complex although no structure of Ras-Raf is currently available . Mutational analysis and molecular dynamic simulations also indicate that Asp33 may form a salt bridge with Raf Lys84. The D33A Ras mutant has a greatly reduced binding affinity in Ras-Raf interaction (Terada, Ito, et al. 1999). Tyr32 alternates between two stable conformations in the GTP bound form of Ras (Geyer, Schweins, et al. 1996). Mutation of Tyr32 to Phe, Ala, Ser, Glu or Lys reduces GAP

stimulated GTP hydrolysis but does not affect intrinsic GTP hydrolysis, demonstrating that Tyr32 plays a significant role in GAP catalyzed GTP hydrolysis (Fidyk & Cerione 2002). Pro34 may act to restrict the conformational changes that occur in switch I (Sprang 1997). Thr35 is a conserved residue that is linked directly to the gamma phosphate via its backbone amide group which hydrogen bonds to one of the gamma phosphate oxygens. GTP hydrolysis results in the loss of this hydrogen bond in the GDP bound form of Ras, allowing the Switch I region more conformational freedom in the GDP bound state. NMR studies of T35S and T35A mutant Ras indicates the entire Switch I region is disordered in both mutants, emphasizing the crucial role Thr35 plays in Switch I stabilization (Spoerner, Herrmann, et al. 2001). Furthermore, Thr35 is glucosylated by *Clostridiium sordelli* lethal toxin, trapping Ras in an inactive conformation that does not bind to downstream effector proteins (Geyer, Wilde, et al. 2003). Ile36 forms part of a hydrophobic binding pocket that interacts with Leu902 in the Ras-RasGAP complex (Scheffzek, Ahmadian, et al. 1997).

B. Switch II

Switch II, defined as a region which undergoes a conformational change between the GTP and GDP bound forms, comprises loop4 (58-64) and helix a2 (65-74). Switch II overlaps with the G3 box (blbbDxxGl), where l is a hydrophilic residue and b is a hydrophobic residue, identified by sequence analysis (Colicelli 2004). The structure of this region is very sensitive to the presence or absence of the gamma phosphate. Gly60 is a conserved residue that makes a hydrogen bond via its backbone amide with one of the gamma phosphate oxygens. Gln61 plays an important role in stabilizing the transition state of the GTP hydrolysis reaction. Mutation of Gln61 to Leu61 inhibits GTP hydrolysis and is

oncogenic. Tyr64 plays a role in binding to several downstream effector proteins and in binding RasGAP and RasSOS. Y64G/S65G or Y64G/Y71G double mutants show reduced binding to RasGAP NF1 and PI3Kinase (Moodie, Paris, et al. 1995). Mutation of A59, Q61 or E63 also inhibit interactions with RasGAP and NF-1 (Marshall 1993). Arg68 and Asp69 play important structural roles, stabilizing the backbone of Switch II by hydrogen bonding with backbone carbonyl oxygens. Met67 interacts with the Ras-GAP protein and other downstream effector proteins.

C. The P-loop

The diphosphate binding loop, or P-loop, which consists of residues 10-17 and lies between strand β1 and helix α1 overlaps with the G1 box identified by sequence analysis [aaaaGxxxxGK (S or T)], where a is an aliphatic residue, and is also referred to as the Walker A motif found in all purine nucleotide binding proteins (Colicelli 2004). Backbone amide groups of the P-loop (residues 14-17) stabilize the negative charge of the phosphates. The backbone amides of Gly12 and Gly13 form hydrogen bonds with the gamma phosphate, and mutation of either of these residues is oncogenic. Experiments have shown that these oncogenic mutants act by lowering the pKa of the gamma phosphate, making hydrolysis prohibitive (Schweins, Geyer, et al. 1996).

D. The Guanine Nucleotide Binding Pocket

The guanine nucleotide base is bound by two sets of conserved sequences 116-119 (NKXD), where X is Cys118 in H-Ras, and 145-147 (SAK). Sequence analysis defines the guanine binding pocket as the G4 box with the sequence motif [bbbb (N or T) (K or Q) xD]. Asn116 connects both groups of conserved residues as well as interacting with residues from

the P-loop, connecting all three nucleotide binding regions together via hydrogen bonds from the side chain amide of Asn116 to the side chain hydroxyl of Thr144 and the backbone oxygen of Val14 (Soares, Miller, et al. 2001).

E. Magnesium Ion Coordination

The negatively charged phosphate groups are partially stabilized by a positively charged magnesium ion. The magnesium ion is coordinated to the backbone phosphate oxygens of the beta and gamma phosphate groups as well as the side chain hydroxyls of Thr35 from Switch I and Ser17 from the P-loop. Two conserved water molecules also participate in Mg ion coordination. One water molecule makes a hydrogen bond with the backbone carbonyl of Thr58 and the side chain oxygen of a conserved Asp57. The other structurally conserved water molecule makes an additional hydrogen bond with the backbone carbonyl of Asp33 from Switch I. This magnesium ion is found in both GTP and GDP structures and is well conserved across the Ras superfamily.

F. GTP Hydrolysis Mechanism

Despite the large amount of structural and biochemical data, the exact mechanism of GTP hydrolysis, and the effect of oncogenic mutations on GTP hydrolysis is still an area of active research. Research has focused on both intrinsic and GAP catalyzed GTP hydrolysis. Because Ras proteins with a mutation at Q61 remain stuck in the GTP bound state, even in the presence of GAPs, a mechanism in which Q61 played a catalytic role in hydrolysis was widely adopted. More recent studies question this interpretation. In both intrinsic and GAP catalyzed GTP hydrolysis, the substrate GTP acts as a catalytic base to abstract a proton from a conserved water molecule, which then attacks the gamma phosphate of GTP, to create a

pentavalent phospho intermediate and GDP(Schweins, Geyer, et al. 1996). GAPs catalyze GTP hydrolysis by stabilizing the transition state of the hydrolysis mechanism. Mutations at codons 12, 13 or 61 lower the rate of both intrinsic and GAP catalyzed GTP hydrolysis, although they work in different ways. In GAP catalyzed hydrolysis, mutations at codons 12 or 13 result in steric hindrance of the 'arginine finger', Arg789 (Scheffzek, Ahmadian, et al. 1997) . In intrinsic hydrolysis, mutations at codons 12 or 13 reduce the pKa of the gamma phosphate, making generation of the pentavalent phospho intermediate less favorable (Schweins, Geyer, et al. 1996). A recent computational study suggests that mutation at codon 61 disrupts the ability of Ras to adopt a catalytic conformation (Shurki & Warshel 2004). One of the difficulties in elucidating the role of Q61 stems from the conformational flexibility of this region. Our structures of the wild type and Q61L mutant of H-Ras lend support to the notion that Q61L preferentially adopts a non-catalytic conformation.

G. Ras Effector Binding Domain

A Ras effector domain is defined as Ras residues which interact with downstream effector molecules. The core Ras effector domain, residues 32-40, is completely conserved in all three Ras isoforms (H, K, and N) and is well conserved among different Ras subfamily members. Comparative sequence analysis studies refer to this region as the G2 Box. The core domain is involved in most, if not all Ras-effector binding interactions and overlaps with Switch I residues 30-37, providing a structural explanation of how the molecular switch works. Biochemical and genetic experiments provide evidence that residues 20-48 constitute an extended Ras effector binding region, reviewed by (Campbell, Khosravi-Far, et al. 1998). A sequence alignment of the extended Ras effector binding region is shown in Figure 5.

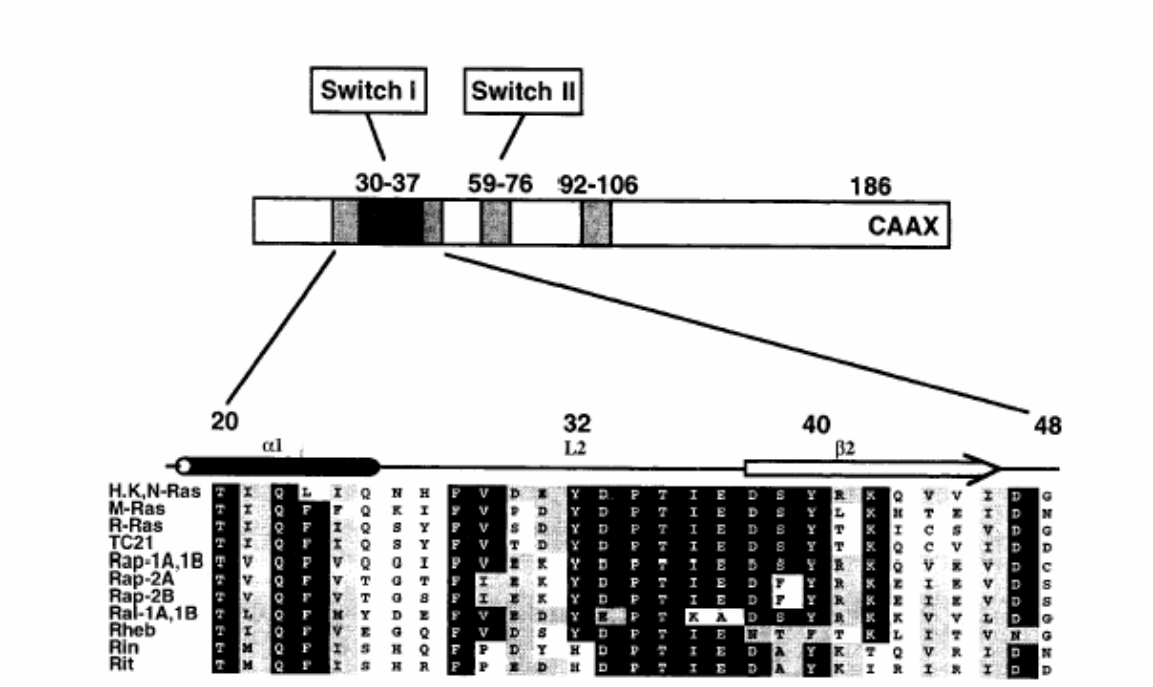


Figure 5. Sequence alignment of the extended Ras effector binding region. Figure reprinted with premission from (Campbell, Khosravi-Far, et al. 1998).

Several residues are well conserved in this region among members of the Ras subfamily, including Phe23, Phe28, Val29, Lys42 and Asp47. Mutagenesis experiments show that deletion of residues 22-43 blocks transformation of NIH3T3 cells by oncogenic Ras. This type of experiment was widely used to map important functional residues to the surface of Ras. Similar mutagenesis studies found effector residues between 26-48, with critical amino acids outside the core effector region being Val45 and Gly48. Experiments with RhoA-Ras chimeric proteins also implicate this region as a Ras effector domain. A RhoA-Ras chimera consisting of RhoA with Ras residues 23-48 was shown to be transforming, whereas smaller chimeras were not transforming, reviewed by (Marshall 1993).

Available crystal structures of Ras-effector complexes provide structural details about some Ras effector residues. However many of the available crystal structures are of minimum Ras binding domains that do not extend to interact with regions outside the core Ras effector domain. The studies with minimum Ras binding domains do not account for all interactions that are important in understanding the physiological interactions between Ras and effector proteins. For instance, mutations of residues 41 and 43 from the extended Ras effector domain, don't affect binding to the minimum Raf binding domain, but do interfere with binding to full length Raf-1 (Winkler, Johnson, et al. 1997). In an effort to probe the entire surface of Ras for potential protein: protein interactions that can guide future biochemical and genetic experiments, we use the multiple solvent crystal structures method (MSCS) to characterize the surface of active Ras bound to GMPPNP.

V. Characterizing Protein Binding Sites by Solvent Mapping

A. Multiple Solvent Crystal Structures (MSCS)

MSCS involves solving the structure of a protein crystal soaked in several, distinct organic solvents. The multiple structures are then superimposed for analysis. The organic solvents used vary in a number of properties (size, shape, charge polarity, dielectric, viscosity, hydrophobicity) that mimic the variety of side chains and /or ligands with which a protein can interact. Structural studies have shown that protein crystals soaked in organic solvents retain the same structure as the protein crystals in aqueous solution, provided they are appropriately crosslinked before transfer to the organic solvents. These studies also found that organic solvent molecules bind at a limited number of discrete sites on the protein

surface (Mattos & Ringe 2001) . MSCS analysis of Elastase with the organic solvents acetonitrile, acetone, dimethylformamide, ethanol, isopropanol, 1-2 hexenediol and trifluoroethanol show that organic solvent molecules cluster in the active site at the P and P' subsites . With few exceptions, organic solvent molecules that don't cluster are found at regions of crystal contacts and are considered to be experimental artifacts due to non-biological binding pockets that form as a consequence of crystallization. MSCS of Thermolysin with the organic solvents isopropanol, acetone and acetonitrile found similar trends (English, Groom, et al. 2001)(English, Done, et al. 1999). The results of these studies led us to believe that sites of protein: protein interaction can be experimentally mapped by this method. Thus the MSCS method is used here to further characterize the binding surface of Ras. Because the surface properties of Ras differ significantly from those of proteases, this study also served to test and expand the limits of the MSCS method.

Ras protein is very different from the few proteins that have been characterized by soaking the protein crystal in organic solvents. With the exception of Lysozyme, all of the proteins soaked in organic solvents are proteases. This includes Thermolysin, Subtilisin Carlsberg, Elastase and Chymotrypsin (reviewed by Mattos & Ringe 2001). These proteins are very stable, extracellular proteins. The protein recognition site consists of a pre-formed active site that binds the protein substrate via a lock and key mechanism. The protein binding site for the proteases is situated within large, protected active site clefts on the protein surface. Ras, on the other hand is an intracellular signaling protein involved in transient protein: protein interactions with multiple partners. The known effector sites are conformationally flexible and many effector residues undergo a disorder - to - order transition upon protein binding. Because of the flexible nature of Ras protein binding sites,

we could not assume that the protein structure would remain unaffected by the organic solvent molecules. Therefore, initial room temperature studies of Ras soaked in a small number of organic solvents were carried out to test this assumption. The results of these studies led us to conclude that the overall structure of protein: protein interaction sites on Ras was relatively stable in organic solvents, even in the case of the conformationally flexible switch regions (Buhrman, de, et al. 2003).

MSCS has also been used to assess the relative affinity of solvent binding sites. English et. al. solvent mapped Thermolysin in increasing concentrations of isopropanol (2,5,10,25,60,80,90 and 100 percent). These experiments found two sets of solvent binding sites. One smaller set of sites was identified at low concentrations of isopropanol (2-25 percent) and a larger set of sites was identified at higher concentrations of isopropanol (60-neat) (English, Done, et al. 1999). Interestingly, only one of the four active-site subsites of Thermolysin was occupied by isopropanol at low concentration, whereas all four subsites were occupied by isopropanol soaked at high concentrations.

B. General Characteristics of Protein Binding Sites

Protein binding sites can be structurally classified in a number of ways. They can be classified based on size, shape, charge, specificity, affinity, flexibility, persistence, and shape complementarity. Specific measurements include: solvation potential, residue interface propensity, hydrophobicity, concave / convex shape and accessible surface area. Transient protein: protein interactions occur in signal transduction pathways, protein inhibitor complexes and antibody:antigen interactions. Permanent protein: protein interactions occur primarily with structural or scaffolding proteins or large multiprotein complexes like the

ribosome or proteosome. Interfaces in transient hetero complexes tend to be more hydrophilic than permanent complexes or protein inhibitor complexes. Ras protein is an example of a protein which forms transient protein interactions with a high degree of hydrophilicity. For example, the Raps - Raf complex (PDB code: 1gua), in which Rap was mutated to resemble Ras, buries 2.5 charges over 17 residues, or 0.15 charges per residue. The average charge per residue distribution in protein interfaces is 2.87 charges over 32 residues, or 0.09 charges per residue. (Protein interface statistics are taken from the Surface Properties of INterfaces (SPIN) database found at: http://trantor.bioc.columbia.edu/cgi-bin/SPIN/) The Raps - Raf interface includes more hydrogen bonds and salt bridge interactions than average. The large degree of polarity at this protein interface is expected, since Ras is a monomeric protein that remains stable when its protein interaction surface is exposed to aqueous solvent. Because of the polar nature of the Ras binding site, a wide range of polar solvents were used in solvent mapping Ras.

Protein interfaces tend to be roughly circular, with an average size of 1971 $Å^2$ with a standard deviation of 1438 $Å^2$, involving 32.21 residues on average. At 1224 $Å^2$, the Raps - Raf complex has an interface which is roughly two thirds the size of the average protein interface, but well within one standard deviation of known protein interfaces. However, the Raps - Raf crystal structure consists of a truncated Raf kinase protein, and does not account for all of the biochemical binding data on full length Raf kinase binding. Although protein interfaces tend to be large and encompass a number of residues, often only a few residues contribute significantly to binding affinity. Scanning alanine mutagenesis studies on human Growth hormone and its receptor showed that although the overall binding interface between hGH and hGHbp was over 1300 $Å^2$, a large proportion of the binding affinity was due to a

few residues that form a hydrophobic core, which is surrounded by hydrophilic residues at the center of the protein: protein interface (Clackson & Wells 1995). These crucial residues describe the functional epitope within the overall protein interface. MSCS with Elastase and Thermolysin not only correctly identify the protein interface, but the hotspots in which organic solvent molecules clustered identify known functional epitopes within the protein interface. Protein interface analysis shows that Ras interacts with effector proteins via a number of distinct binding modes (Corbett & Alber 2001). Functional epitopes within these binding interfaces have been characterized by genetic, biochemical and computational methods. Kiel and Herrmann have carried out an extensive alanine mutagenesis study of the binding interfaces of Raps-Raf and Ras-RalGDS (Kiel, Serrano, et al. 2004). Like the Clackson and Wells study, Kiel and Herrmann define a minimum functional epitope within the larger protein interface. Unlike the hydrophobic residues that make up the functional epitope for hGH : hGHb, the functional epitope for the Ras protein interface consists of mainly charged residues (Kiel, Serrano, et al. 2004). Our MSCS work with Ras identifies functional epitopes within known protein binding sites and also predicts new binding site epitopes in a region distal to the effector region of Ras.

The common characteristics of protein-protein binding sites have been exploited by computational methods used to predict protein binding sites. ProMate is a program used to predict protein: protein binding sites specifically in transiently formed complexes like Ras: effector protein complexes, as opposed to more permanent protein-protein interactions, of the type found in structural protein complexes. ProMate uses a prediction algorithm that accounts for the interface properties and tendencies in known transient complexes. ProMate ranks protein interface sites based on a number of criteria, including the presence and size of

hydrophobic patches, secondary structure, amino acid type and whether or not a multiple alignment shows evolutionarily conserved positions (Neuvirth, Raz R., et al. 2004). ProMate roughly predicts the known effector binding domain of Ras. CastP predicts protein: ligand binding sites on the basis of binding pocket geometry(Binkowski, Naghibzadeh, et al. 2003). CastP correctly identifies some of the known protein binding sites on Ras.

C. Functional Epitopes within Ras Protein Binding Modes

Of the nine Ras superfamily protein binding modes identified by Alber *et al*., there is structural data from x-ray crystallography of Ras - protein complexes for three major binding modes (Corbett & Alber 2001). The six binding modes with no Ras structural data include (Ran-Importin Beta, Arf-ArfGAP, Rab-Rabphilin-3A, Rho-PKN/PRK1, Ran-RanBP2, Rac-p67) (Corbett & Alber 2001). Several are represented by only a single protein complex in the protein interface analysis. More data on Ras-related complexes is required to determine whether or not these complexes represent major binding modes for Ras, or are unique outliers. The three binding modes utilized by Ras include complexes of the Ras - Ras Binding Domain (RBD), Ras-GEF and Ras-GAP. PI3 kinase, Raf kinase and RalGDS all interact with Ras via a conserved Ras Binding Domain. The Ras Binding Domain consists of a ubiquitin fold (BBαBBαB) and all RBDs interact with Ras in a similar manner, forming an anti-parallel, intermolecular Beta sheet between the second beta strand of the RBD and the second beta strand of Ras (Herrmann 2003). Residues which are buried in the Raps :Raf protein interface include : Gln25 (40 % buried in monomer/71 % buried in interface), Ile 27 (48/61), Val29 (79/90), Glu31 (45/55), Asp33 (36/61), Ile36(17/65), Glu37(33/84), Asp38 (63/87), Ser39 (48/95), Tyr40 (77/97), Arg41 (50/91) and Met52 (88/94) (Statistics on

percent burial are from the SPIN database). Residues that are 30 % or more buried in the interface than in the monomer are designated as critical buried residues and include: Gln25, Ile36, Glu37 and Arg41. Interface residues identified by Kiel and Herrmann as having a large positive effect on binding affinity, (>1.0 kcal/mol) as measured by isothermal calorimetry, include: (Asp33, Ile36, Gln37 and Asp38). Gln25 was not measured and Arg41 had a negative effect on binding affinity due to changes in entropy. Overall, residues that are well buried correlate with residues that define the functional binding epitope in the Raps : Raf complex. A second binding mode utilized by Ras-GEFs is seen in the Ras-SOS complex (PDB code: 1BKD) (Boriack-Sjodin, Margarit, et al. 1998). This interface consists of Ras residues: Arg102 (59.2% solvent accessible in monomer /21 % solvent accessible in complex) ,Gln61 (20.5/4.4), Glu63 (46.0/ 14.0) , Tyr64 (39.8/ 2.3), Met67 (23.6/1.9), Asp69 (17.2/4.6), Gln70 (33.7/2.7), Arg73 (42.9/12.6), Tyr40 (26.3 /10.4), Tyr32 (24.2/4.0), Pro34 (27.9/5.1), Ile21 (14.8/2.6), Ser17 (11.2/0.5) (Solvent accessibility was calculated with CNS). Residues Arg102, Glu63, Tyr64, Gln70 and Arg73 are critically buried. In this Ras binding mode SOS inserts a helix between Switch I and Switch II, disrupting phosphate binding residues and stabilizing the nucleotide free form of Ras (Boriack-Sjodin, Margarit, et al. 1998). A third binding mode utilized by Ras-GAPs is seen in the Ras-GAP complex (PDB code: 1WQ1) (Scheffzek, Ahmadian, et al. 1997). Ras residues that are significantly buried in this interface include: Tyr32 (60.0% solvent accessible in monomer /2.9% solvent accessible in complex), Asp33 (26.3/4.2), Pro34 (18.0/1.0), Ile36 (34.4/3.4), Gln61 (18.6/0.0), Glu62 (42.2/18.2), Glu63 (42.2/5.5), Tyr64 (22.9 / 2.2), Met67 (41.1/19.2) and Lys88 (43.4/12.9). Tyr32, Ile36, Glu63 and Lys88 are critically buried in this interface. The Ras-RBD critically buried residues surround the Switch I, Ras effector region. The Ras-GEF

critically buried residues are located in Switch II / helix 3. The Ras-GAP critically buried residues overlap with both the Ras-GEF binding mode and the Ras-RBD binding mode. These three binding modes stabilize different structures of Ras as shown by root mean square deviation (RMSD) comparisons of Ras bound to these Ras binding proteins. Most of the critical interacting residues are polar, in contrast to the proteases, whose nonpolar binding surface was previously mapped by organic solvents. Solvent mapping of the Ras binding surface required the use of a variety of organic solvents, with a wide range of polarities.

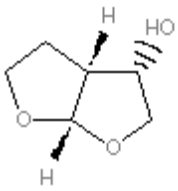
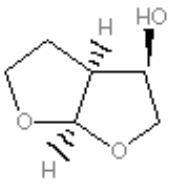
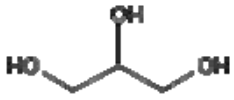
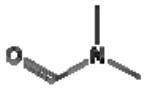
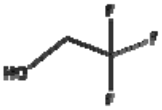
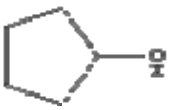
D. Organic Solvent Molecules

The organic solvent molecules used to solvent map Ras include 2,2,2 trifluoroethanol (TFE), N,N dimethylformamide (DMF), 1,6-hexanediol (HXD), hexane (HXN), glycerol (GLY), isopropanol (ISO), cyclopentanol (PEN) , R,S,R bisfuranol (BIR) and S,R,S bisfuranol (BIS). The chemical formula, molecular structure and various solvent properties for these organic solvents are listed in Table 1 and Table 2.

Table 1. Chemical properties of organic solvents used in MSCS of Ras.

Solvent	**Chemical Formula**	**relative polarity**	**dielectric**	**Solubility in water**	**logP**
2,2,2,Trifluroethanol	F3CCH2OH	n/a	26.5	Miscible (M)	0.652
dimethylformamide	C3H7NO	0.404	37	M	-1
hexane	C6H14	0.009	1.9	Not miscible	3.5
isopropanol	C3H8O	0.546	19.9	M	0.55
cyclopentanol	C5H10O	n/a	17	Slightly	0.71
glycerol	C3H8O3	0.812	43	M	-1.76
1,6 hexanediol	C6H14O2	n/a	n/a	M	n/a
bisfuranol	C6H10O3	n/a	n/a	M	-0.65

Table 2. Chemical Structure of organic solvents used in MSCS of Ras.

S,R,S-Bisfuranol 	R,S,R-Bisfuranol 	Glycerol 
Dimethylformamide 	2,2,2-Trifluoroethanol 	Cyclopentanol 
Isopropanol 	1,6-Hexanediol 	Hexane 

Because the Ras protein binding interface involves charged residues, we used a panel of organic solvents with a wide range of polarities, as measured by relative polarity, dielectric constant, Total Polar Surface Area (TPSA) and logP measurements of the octanol / water partition coefficient. The dielectric constants of the organic solvents used range from a low of 1.9 for hexane to a high of 43 for glycerol, with the relative order being: HXN < PEN < ISO < TFE < DMF < GLY. All of the organic solvents have a dielectric constant lower than water. TPSA measures the percentage of the molecular surface occupied by polar atoms (Ertl, Rohde, et al. 2000). The relative order of TPSA values for the organic solvents used is similar to the order of dielectric constant values for the organic solvents. Glycerol and bisfuranol have a large amount of electronegative surface area and hexane has no polar surface area. LogP measures the partition coefficient of a molecule between octanol and water(Eros, Kovesdi, et al. 2002). The partition coefficient is used as a standard measure of hydrophobicity. Hydrophobic molecules partition more in the octanol phase, while hydrophilic molecules tend to partition more in the water phase. Amino acids have a wide range of hydrophobicities, ranging from hydrophobic phenylalanine to hydrophilic arginine and glutamate. We used organic solvents with a similar range of hydrophobicities. LogP values range from a low of -1.76 for glycerol, which partitions primarily in the water phase, to a high of 3.5 for hexane which partitions almost exclusively in the octanol phase, with the relative order being: GLY < DMF < BIR < TFE < ISO < PEN < HXN.

With the exception of hexane and cyclopentanol, all of the solvents are miscible in water. Organic solvents that are miscible in water were soaked into the crystals at various solvent concentrations, whereas hexane and cyclopentanol were used at neat concentrations. The solvent concentration used for the final data collection was the maximum concentration

that could be achieved without damaging the crystal. Interestingly, similar maximum concentrations were found for organic solvents used in solvent mapping experiments with Elastase, Ras and Lysozyme (unpublished data). Prior to soaking the crystals in organic solvents, the crystals were crosslinked with glutaraldehyde to increase their stability in the organic solvents and prevent unwanted crystal damage. Crosslinking crystals with glutaraldehyde is commonly used in solvent mapping experiments, although the glutaraldehyde concentration and reaction time need to be optimized for different protein crystals. In solvent mapping experiments with Subtilisin Carlsberg, Fitzpatrick et. al. found a very narrow window for optimal crosslinking (Fitzpatrick, Steinmetz, et al. 1993). Crosslinking at 1% glutaraldehyde resulted in crystal damage during solvent soaks, while crosslinking at 2% resulted in crystals that did not diffract x-rays.

Trifluoroethanol (TFE) is widely used in protein biochemistry as a membrane mimetic and to stabilize protein structures (reviewed extensively in Buck 1998). TFE is particularly adept at stabilizing alpha helical structures in small peptides (Mishraa, Palgunachari, et al. 2001). The ability of small peptides to form alpha helices in TFE correlates with the helical propensity of the amino acids within the peptide sequences. TFE effects a β-sheet to α-helix transition in β-lactoglobulin protein (Kumar, Modig, et al. 2003) The apparent mechanism of helix stabilization involves the aggregation of TFE around the protein, resulting in water displacement and creation of a low dielectric local environment that favors protein: protein hydrogen bonds, along with disruption of native tertiary structure (Roccatano, Colombo, et al. 2002). In crosslinked Ras crystals, the disruption of native tertiary structure is diminished and a helical transition is not observed. Dimethylformamide (DMF) has been used in previous solvent mapping experiments and is known to affect

protein structure. Melnikova et. al. found that soaking two monoclonal antibodies in a range of DMF concentrations (5-40%) resulted in conformational changes around the antigen binding site. Interestingly, this effect was specific for DMF and was not seen with methanol, dimethyl sulfoxide or TFE (Melnikova, Odintsov, et al. 2000). DMF is very polar and has planar character, so it is used as a peptide bond mimic (Ringe & Mattos 1999). Hexane is a nonpolar organic molecule, with no alcohol moiety and has been used in previous solvent mapping experiments (Gao, Maldonado, et al. 1999), (Yennawar, Yennawar, et al. 1994). Hexane is used in computational biology, where computational hexane / water partitions are used to mimic the cytoplasm / membrane interface to examine protein structural changes at the interface. Hexane was used to mimic the non-polar, non-aromatic side chain residues. Cyclopentanol was chosen because its cyclic shape could mimic the non-polar, cyclic protein side chains. Cyclopentanol is an antagonist of ethanol induced disruption of L1 mediated cell-cell adhesion and may act by binding within a hydrophobic protein pocket (Wilkemeyer, Menkari, et al. 2002). Glycerol is widely used as a protein stabilizing agent in cryo-protection buffers, protein crystallization buffers and protein re-folding buffers and is known to stabilize protein structure in general (Sousa 1995). Refolding experiments with denatured Lysozyme have shown that soaking in neat glycerol allowed regeneration of the native state of the protein (Dong, Huang, et al. 2004). Glycerol is polar, effectively replacing water molecules that interact with protein structure. Two enantiomers, (3R,3aS,6aR)-hexahydrofuro[2,3-b]furan-3-ol and (3S,3aR,6aS)-hexahydrofuro [2,3-b]furan-3-ol , which will be referred to as R,S,R and S,R,S bisfuranol, were purified separately and donated to the Mattos lab by GlaxoSmithKline. R-S-R - bisfuranol is of interest as a base compound for

drug design, and is being used as the major substituent in a new class of HIV protease inhibitors (Koh, Nakata, et al.).

VI. Summary of Dissertation Work

In this dissertation, the following topics will be examined:

Chapter 2 contains the details of all experimental procedures used in the Ras work discussed in Chapters 3-6.

Chapter 3 will examine the bulk effects of three organic solvents, 2,2,2 trifluoroethanol, 1,6 hexanediol and isopropanol, on protein structure, in the context of room temperature studies of Ras crystals which form with the symmetry of space group P3(2)21.

Chapter 4 will discuss the structure of wild type and the Q61L oncogenic mutant of Ras bound to GMPPNP, which was solved from crystals which form with the symmetry of space group R32. This crystal form provided us with a Switch I in the conformation found in the Raps-Raf complex and gives a unique view of the Switch II region undistorted by crystal contacts.

Chapter 5 will discuss the results of MSCS applied to Ras-GMPPNP crystals which form with the symmetry of space group R32. The panel of organic solvents used in this study included: 2,2,2 - trifluoroethanol, isopropanol, glycerol, cyclopentanol, dimethylformamide, 1,6 - hexanediol, hexane and R,S,R bisfuranol. The high resolution data sets for these studies were collected at 100 K, using synchrotron radiation at a wavelength of 1.0 Angstrom.

Chapter 6 will examine the interactions of two enantiomers, R,S,R and S,R,S bisfuranol, with two proteins, H-Ras-GMPPNP (Ras) and Hen egg white Lysozyme (HEWL) using the multiple solvent crystal structure method (MSCS).

Chapter 7 will examine the structural rearrangements in the active site P-loop of Cdc25B phosphatase that occur in response to oxidation of the active site cysteine residue. This project was unrelated to the main topic of this thesis and therefore was not discussed in this background chapter. An introduction to this work is given at the beginning of Chapter 7. We have obtained a crystallographic time course, as the enzyme goes from its apo state, through the sulfenic (Cys-SO^-) intermediate to the stable disulfide. We have also obtained the structures of the irreversibly oxidized sulfinic (Cys-SO_2^-) and sulfonic (Cys-SO_3^-) Cdc25B in addition to the structures of two active site mutants, C473S and C473D, which reveal additional structural rearrangements of the active site P-loop.

CHAPTER TWO

EXPERIMENTAL PROCEDURES

Protein expression and purification

Plasmids containing the wt and Q61L mutant H-Ras 1-166 gene are over expressed in *E. coli* cells. Cells are grown in LB containing 50mg/mL ampicillin and over expression is induced with 0.2 mM IPTG. Cells are grown at 32 ^{O}C and induced for 5-6 hours from an OD_{600} of 0.5-0.7. Cells are pelleted by centrifugation and processed immediately or stored at -80 ^{O}C. Pellets are resuspended in resuspension buffer (9mL/ g cell paste). Resuspension Buffer: Buffer A + protease inhibitor cocktail. Buffer A: 20mM Tris pH 8.0, 50mM NaCl, 5 mM $MgCl_2$, 1mM DTT, 5 % glycerol, 20uM GDP. Protease inhibitor cocktail: 2ug/mL antipain, 1 ug/mL leupeptin, 1ug/mL pepstatin, 1ug/mL E-64, 1mM Pefabloc, 5 mM benzamidine. Cells are lysed by sonication and the insoluble fraction is pelleted by centrifugation. DNA is precipitated from the soluble fraction by slow addition of polyethylenimide to 0.02 %, while stirring at 4 ^{O}C and pelleted by centrifugation. Anion exchange chromatography is performed by FPLC with a HiLoad 26/10 Q Sepharose Fast Flow ion exchange column (Amersham Pharmacia). H-Ras elutes as a broad peak between 220 and 280 mM NaCl. Protein pooled from anion exchange is concentrated to 2 mLs and loaded onto a HiPrep 26/60 Sephacryl S-100 High Resolution gel filtration column at a flow rate of 1 ml/min in Buffer A. H-Ras elutes as a single peak centered at 150 mLs. Protein purity is assessed by SDS-PAGE. The GDP nucleotide bound to Ras was exchanged for GMPPNP using published procedures (John, Sohmen, et al. 1990). Protein from the exchange reaction is exchanged into stabilization buffer (20mM Hepes pH 7.5, 50mM NaCl,

5mM $MgCL_2$, 1mM DTT, 20 uM GMPPNP) concentrated to 15-20 mg/mL and used immediately for crystallization trials or stored as 50 ul aliquots at -80 ^{O}C.

Experimental Procedures for Room Temperature Studies

Crystals of H-Ras 166/ GMPPNP which were solved in the space group P3(2)21 were obtained by the hanging drop vapor diffusion method using published conditions (Scherer et al., 1989).

Crosslinking and transfer to organic solvents

Ras crystals were transferred to sitting drop plates containing a solution of 1% gluteraldehyde in the crystallization buffer: 64mM Hepes (pH 7.6), 25 % PEG 1450, 5 mM $MgCl_2$. The crosslinking reaction was allowed to proceed at room temperature for 30 minutes. Crystals were washed with 3 mL of the crystallization buffer without gluteraldehyde. This solution was exchanged with one containing buffer and an organic solvent: 64 mM Hepes pH 7.6 and either 50% isopropanol, 50% 2,2,2-TFE or 60% 1,6 hexanediol. The exchange was done in a stepwise fashion, increasing the solvent concentration by 10 % every five minutes until the final concentration is reached, and the crystals were soaked for two hours in the final solutions containing organic solvents at the concentrations specified above, before being mounted in quartz capillaries for data collection at room temperature.

Data Collection, Processing and Structure Refinement

Data were collected on a Mar345 phosphorimaging plate mounted on a GX-13 rotating copper anode running at 40 kV and 50 mA. The exposure time was 600 seconds per frame. The oscillation angle per frame was 1 degree and the crystal to detector distance was

170 mm for all crystals. The data were processed using the HKL package (Otwinowski and Minor, 1997). Initial electron density maps were calculated using a published model of Ras (PDB code 121P) with the Switch II residues, the nucleotide, the Mg^{2+} and all water molecules removed (Pai, Krengel, et al. 1990) . The models were refined with CNS (Brunger et al., 1998) using a maximum likelihood target function and all data. Water and organic solvent molecules were identified using Fo-Fc electron density maps contoured at the 3σ level and added during the course of refinement.

Experimental Procedures for Cryo Studies of wt and Q61L Ras

Crystallization conditions

Crystals solved in the space group R32 were grown by hanging drop vapor diffusion at 18 ^{O}C. Crystals of wt H-Ras 166 –GMPPNP form in 5-10 days and grew to an average size of 0.5 mm^3. The protein concentration in the drop was 8-10 mg/ml after addition of the reservoir solution. The reservoir consisted of 0.2 M $CaCl_2$, 20% PEG 3350 (PEG Ion Screen condition #7 from Hampton Research). Crystals of the Q61L mutant protein also form in 5-10 days. The protein concentration in the drop was 6-7 mg/ml after addition of the reservoir solution. The reservoir consisted of 800 ul (0.2 M $CaCl_2$. 24-26 % PEG3350) and 200 ul stabilization buffer with fresh DTT. Crystals of the Q61L mutant often grew from drops containing precipitated protein.

Data Collection and Processing

Crystals were frozen in liquid nitrogen directly from the crystallization drop. Initial screening and data collection was done in house with the Mar 345 image detection

system mounted on a GX-13 rotating copper anode running at 40kV and 50mA. Final data collection at a wavelength of 1.0 Angstrom was done at Ser-CAT ID-22 beamline at APS (Argonne, IL). Exposure time was from 1 -3 seconds with an oscillation angle of 1 degree and a crystal to detector distance of 120 mm. Frames were indexed and scaled with HKL2000 (Otwinowski and Minor, 1997).

Structure Refinement of aqueous control structure

A protein only model of H-Ras 166 (PDB code 1CTQ), with B-factors set at 30, was used as an initial search model for molecular replacement using the program Crystallography and NMR System (CNS) (Brunger, Adams, et al. 1998). The program CNS was also used for all reciprocal space refinement with 10% of the unique reflections set aside for the calculation of R_{free} (Brunger 1993). A fast direct rotation search gave a unique solution with a RF-function of 0.0877. The next best solution had a RF-function score of 0.0596. A translation search of the top five rotation peaks also gave a unique solution with a monitor value of 0.35 and a packing score of 0.5283. The next best solution had a monitor score of 0.158 and a packing score of 0.4480. The best solution was applied to generate a model used for 30 rounds of rigid body refinement at 2.5 Angstrom, followed by 30 rounds of rigid body refinement at 2.0 Angstrom. This was followed by a round of simulated annealing, energy minimization and group B-factor minimization in CNS prior to generation of 2Fo-Fc and Fo-Fc electron density maps. Manual rebuilding was done in O and successive rounds of energy minimization and individual B-factor minimization was done in CNS (Jones, Zou, et al. 1991). Residues 61-71 were removed from the initial model and rebuilt in O, along with the water molecules, ions and GMPPNP in successive rounds of

manual rebuilding. The final wt model was used to phase the Q61L mutant.

Crosslinking and soaking in organic solvents

Crystals were transferred to sitting drop plates containing 50 ul of stabilization buffer. Crystals were washed with stabilization buffer, with at least 3 buffer changes in the sitting drop. Stabilization buffer was exchanged with crosslinking buffer, which consisted of stabilization buffer + 1% glutaraldehyde and crystals were allowed to crosslink for 30 minutes to 1 hour, depending upon the size of the crystal. Crosslinking buffer was exchanged with stabilization buffer to remove unreacted glutaraldehyde prior to soaking in organic solvents. Organic solvents were added to the concentrations specified and crystals were allowed to soak from 90-180 minutes. For volatile solvents like hexane, cyclopentanol and isopropanol, drops were refreshed frequently to prevent crystal drying. After soaking, crystals were flash frozen in liquid nitrogen for data collection. In the case of cyclopentanol, crystals were transferred to stabilization buffer prior to freezing because of freezing and mounting problems that occurred with the neat solvent.

Structure Refinement of solvent soaked structures

The program Crystallography and NMR System (CNS) was also used for all reciprocal space refinement with 10% of the unique reflections set aside for the calculation of R_{free}. (Brunger, Adams, et al. 1998). (Brunger 1993). A protein only model of wt H-Ras 166 solved in the R32 space group was used for 30 rounds of rigid body refinement at 2.5 Angstrom, followed by 30 rounds of rigid body refinement at 2.0 Angstrom. This was followed by a round of simulated annealing, energy minimization and group B-factor refinement in CNS prior to generation of 2Fo-Fc and Fo-Fc electron density maps. Manual rebuilding was done in O and successive rounds of energy minimization and individual B-

factor refinement was done in CNS (Jones, Zou, et al. 1991). Residues 61-71 were removed from the initial model and rebuilt in O, along with the waters, ions, GMPPNP and solvent molecules in successive rounds of manual rebuilding. Solvent molecules were not added until the latter stages of refinement. Except for the two bisfuranol molecules, coordinate files for the solvent molecules were taken from previously published structures found in the Protein Data Bank. Coordinate files for the two enantiomers of bisfuranol were generated and energy minimized using Nano-CAD. The Hiccup server was used to generate topology and parameter files used for energy minimization in CNS from the coordinate files. The topology and parameter files were checked by energy minimization in X-plor, to make sure that no serious distortions would occur during CNS energy minimization runs.

Solvent Site Validation

Organic solvent molecules are small and can be mistaken for water molecules, alternate side chain conformations or spurious electron density artifacts. Often, the organic solvent molecules are not distinguishable until later rounds of solvent refinement. They were added into Fo-Fc electron density contoured at 3 σ during the last stages of refinement after several rounds of solvent refinement in which only water and ion molecules were added. Solvent molecules were kept only if the R-factor stayed the same or improved after solvent addition and if the 2Fo-Fc electron density stayed the same or improved after solvent addition. After the final round of solvent refinement, a 'solvent omit map' was calculated by omitting all water and solvent molecules, doing 1 round of simulated annealing in CNS, and calculating a 2Fo-Fc electron density map. Solvent molecules were kept only if significant 2Fo-Fc electron density remained. In many cases, while the 2Fo-Fc electron density clearly

justifies addition of a solvent molecule, some solvent molecules are not resolved enough to be precisely oriented. In these cases, while it is clear that a solvent molecule is bound, specific interactions at the atomic level cannot be determined. Table 5.1 ranks each solvent by B-factor and fits to 2Fo-Fc electron density. Electron density fits are calculated in CNS as the real space correlation coefficient of each molecule. In general, solvent molecules with B-factors higher than 55 and electron density correlation coefficients less than 0.80 are considered 'loosely bound' and the solvent binding site is not analyzed in detail.

DATA COLLECTION, REFINEMENT AND VALIDATION STATISTICS

Table 1. Data collection and refinement statistics. Ras structures solved at room temperature in aqueous mother liquor and in the presence of organic solvents.

Solvent % in water	**Aqueous**	**50% TFE**	**50% ISO**	**60% HXD**
Space Group	P3(2)21	P3(2)21	P3(2)21	P3(2)21
Unit cell	a = b = 40.2 Å c = 161.4 Å α = β = 90º γ = 120º	a = b = 40.3 Å c = 161.3 Å α = β = 90º γ = 120º	A = b = 40.1 Å c = 161.9 α = β = 90º γ = 120º	a = b = 40.4 Å c = 161.7 Å α = β = 90º γ = 120º
Temperature	RT	RT	RT	RT
Resolution (Å)	2.0	2.4	2.0	2.3
# reflections	10,789	5,814	10,423	7,088
Completeness (%)	99.0	95.2	95.9	96.4
R-factor/R-free (%)	20.0/24.5	18.7 / 22.5	21.2 / 24.8	20.1 / 25.0
Bond length (Å)	0.0051	0.0069	0.0052	0.0064
Bond angle (º)	1.12	1.28	1.13	1.15
Dihedral angle (º)	22.5	23.9	22.7	22.7
Average B-factor $Å^2$	20.3	33.4	27.1	32.1
# protein atoms	1269	1323	1273	1318
# magnesium atoms	1	1	1	1
# nucleotide molec.	1	1	1	1
# water molecules	99	35	107	47
# organic solv. mol.	0	1	3	1

Table 2. Data collection and refinement statistics for Ras/GMPPNP soaked in organic solvents.

Solvent (concen.)	**Crosslinked aqueous soak**	**GLY (70%)**	**PEN (Neat)**	**TFE (50%)**	**HXN (Neat)**	**HXD (60%)**	**BIS (20%)**	**DMF (55%)**	**BIR (90%)**
Space Group	R32	R32	R32	R32	R32	R32	R32	R32	R32
Unit cell	a = 88.73 b= 88.73 c = 135.05 Å a = b = 90º g = 120º	a = 89.28 b= 89.28 c = 134.70 Å a = b = 90º g = 120º	a = 87.63 b= 87.63 c = 134.24 Å a = b = 90º g = 120º	a = 88.32 b= 88.32 c = 133.91 Å a = b = 90º g = 120º	a = 88.33 b= 88.33 c = 135.13 Å a = b = 90º g = 120º	a = 88.79 b= 88.79 c = 135.09 Å a = b = 90º g = 120º	a=88.50 b=88.50 c=134.09 a = b = 90º g = 120º	a=88.53 b = 88.53 c = 133.09 Å a = b = 90º g = 120º	a =89.14 b = 89.14 c = 133.09 Å a = b = 90º g = 120º
Temperature	100 K	100 K	100 K	100 K	100 K	100 K	100 K	100 K	100 K
Resolution (Å)	1.7	1.6	1.4	1.9	1.52	1.7	1.6	1.7	1.7
# reflections	21,280	26,775	37,214	15,482	30,118	22,253	26,482	22,462	21,657
Completeness (%)	93.5	97.4	94.8	95.8	95.8	97.6	98.4	99.2	95.7
R-factor/R-free (%)	18.6/20.2	20.0/21.5	18.8/20.1	18.1/21.0	18.8/20.6	18.1/20.8	19.0/21.1	19.6/20.5	20.6/22.8
Bond length (Å)	0.01	0.004	0.0045	0.009	0.01	0.012	0.01	0.005	0.01
Bond angle (º)	1.44	1.1	1.13	1.43	1.5	2.1	1.5	1.12	1.5
# protein atoms	1307	1295	1320	1315	1292	1298	1296	1241	1223
# nucleotide atoms	32	32	32	32	32	32	32	32	32
# Magnesium molecules	1	1	1	1	1	1	1	1	1
# Calcium molecules	3	3	3	3	3	3	3	3	3
# water molecules	136	93	153	119	106	111	119	116	46
# solvent molecules	0	12	1	2	2	1	2	4	19

Table 3. Data collection and refinement statistics for Ras/GMPPNP crystals soaked in various concentrations of S,R,S bisfuranol (BIS) and R,S,R bisfuranol (BIR).

Solvent (concen.)	**BIS (20%)**	**BIR (20%)**	**BIR (50%)**	**BIR (90%)**
Space Group	R32	R32	R32	R32
Unit cell	a=88.50 b=88.50 c=134.09 a = b = 90º g = 120º	a = 88.02 b= 88.02 c = 133.68 Å a = b = 90º g = 120º	a = 88.04 b= 88.04 c = 133.56 Å a = b = 90º g = 120º	a =89.14 b = 89.14 c = 133.09 Å a = b = 90º g = 120º
Temperature	100 K	100 K	100 K	100 K
Resolution (Å)	1.6	1.6	1.5	1.7
# reflections	26,482	25,886	31,662	21,657
Completeness (%)	98.4	97.6	98.5	95.7
R-factor/R-free (%)	19.0/21.1	19.0/20.2	19.3/20.6	20.6/22.8
Bond length (Å)	0.01	0.004	0.004	0.01
Bond angle (º)	1.5	1.1	1.1	1.5
# protein atoms	1296	1308	1288	1223
# nucleotide atoms	32	32	32	32
# Magnesium molecules	1	1	1	1
# Calcium molecules	3	3	3	3
# water molecules	119	96	117	46
# solvent molecules	2	4	4	19

Table 4. Data collection and refinement statistics for hen egg white Lysozyme, crosslinked and soaked in 50% S,R,S bisfuranol (BIS) and 50% R,S,R bisfuranol (BIR).

Solvent (concen.)	**BIS (50%)**	**BIR (50%)**
Space Group	P4(3)2(1)2	P4(3)2(1)2
Unit cell	a=79.15 b=79.15 c=37.29 a = b = 90º g = 90º	a = 79.24 b= 79.24 c = 37.26 Å a = b = 90º g = 90º
Temperature	100 K	100 K
Resolution (Å)	1.6	1.55
# reflections	15,748	17,221
Completeness (%)	97.4	96.9
R-factor/R-free (%)	21.8/24.0	20.2/23.1
Bond length (Å)	0.004	0.004
Bond angle (º)	1.3	1.2
# protein atoms	1001	1001
# Chloride molecules	1	1
# Sodium molecules	1	1
# water molecules	68	137
# solvent molecules	3	3

Table 5. Data collection and refinement statistics for Ras/GMPPNP Q61L Oncogenic mutant.

Solvent (concen.)	**Aqueous**	**Hexane Neat**
Space Group	R32	R32
Unit cell	a=88.953 b=88.953 c=134.022 a = b = 90º g = 120º	a = 88.23 b= 88.23 c = 134.38 Å a = b = 90º g = 120º
Temperature	100 K	100 K
Resolution (Å)	2.0	1.4
# reflections	13,810	38,311
Completeness (%)	98.3	96.2
R-factor/R-free (%)	19.7/23.4	20.0/20.5
Bond length (Å)	0.005	0.004
Bond angle (º)	1.1	1.1
# protein atoms	1303	1,291
# nucleotide atoms	32	32
# Magnesium molecules	1	1
# Calcium molecules	3	3
# water molecules	128	100
# solvent molecules	0	2

$R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$, crystallographic R-factor calculated using 90% of the reflections against which the model was refined.

$R_{free} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$ calculated using the test set consisting of 10% of the total reflections, randomly selected from the original data set.

Table 6. Data collection and refinement statistics for Oxidation and reduction of Cdc25B

Time soaked in 50um H_2O_2 C473 state	O minutes reduced	20 minutes & 90 minute back-soak disulfide	20 minutes Cys-SO_1 sulfenic	30 minutes Cys-SO_2 sulfinic	80 minutes Cys-SO_3 sulfonic
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell	a = 50.23 Å b = 70.99 Å c = 73.90 Å $\alpha = \beta = \gamma = 90^o$	a = 50.20 Å b = 71.16 Å c = 73.69 Å $\alpha = \beta = \gamma = 90^o$	a = 49.86 b = 71.08 c = 74.30 Å $\alpha = \beta = \gamma = 90^o$	A = 49.27 Å b = 70.63 Å c = 73.76 Å $\alpha = \beta = \gamma = 90^o$	a = 49.70 Å b = 71.15 Å c = 74.43 Å $\alpha = \beta = \gamma = 90^o$
Temperature of data collection	100 K	100 K	100 K	100 K	100 K
Resolution (Å)	1.7	2.0	1.7	2.0	1.7
# reflections	29,364	17,645	29,267	16,387	28,852
Redundancy	7.0 (4.5)	6.7 (6.3)	6.9 (4.4)	3.1 (2.8)	7.2 (6.2)
R_{sym} (%)	6.7 (17.7)	9.0 (46.9)	7.6 (21.8)	8.0 (40.2)	6.2 (38.0)
Completeness (%)	98.6	95.8	98.4	91.1	97.1
Average I/σ	28.3 (7.5)	21.6 (4.5)	23.6 (6.5)	13.5 (2.1)	31.2 (3.8)
R_{work}/R_{free} (%)	17.6 / 19.4	20.8 / 24.5	18.7 / 20.7	20.7 / 23.1	18.6 / 21.0
RMS Bond length deviation from ideal geometry (Å)	0.009	0.005	0.012	0.013	0.012
RMS Bond angle deviation from ideal geometry(°)	1.5	1.3	1.5	1.5	1.5
Ramachandran statistics	97% in favored regions 100% in allowed regions	96% in favored regions 100% in allowed regions	96% in favored regions 100% in allowed regions	97% in favored regions 100% in allowed regions	97% in favored regions 100% in allowed regions
# protein atoms	1,427	1,427	1,428	1,429	1,430
# water molecules	186	110	171	137	198

The numbers in parenthesis describe the relevant value for the highest resolution shell.
$R_{sym} = \Sigma|I_i - \langle I \rangle|/\Sigma I$ where I_i is the intensity of the *i*th term observed and $\langle I \rangle$ is the mean intensity of the reflections.

Table 7. Data Collection and refinement statistics for active site Cdc25B mutants

CDC25B Active Site Mutants	C473D	C473S
Space Group	$P2_12_12_1$	$P2_12_12_1$
Unit cell	a = 49.89 Å b = 71.19 Å c = 73.87 Å $\alpha = \beta = \gamma = 90^o$	a = 50.25 Å b = 70.93 Å c = 74.06 Å $\alpha = \beta = \gamma = 90^o$
Temperature of data collection	100 K	100 K
Resolution (Å)	1.6	1.5
# reflections	34,010	37,967
Redundancy	7.1 (5.9)	4.6(3.5)
R_{sym} (%)	8 (35)	7 (24.4)
Completeness (%)	96.0	91.0
Average I/σ	32.8 (4.3)	17 (4)
R_{work}/R_{free} (%)	18.0 / 19.6	19.5/20.7
RMS Bond length deviation from ideal geometry (Å)	0.005	0.005
RMS Bond angle deviation from ideal geometry(º)	1.3	1.2
Ramachandran statistics	97% in favored regions 100% in allowed regions	97% in favored regions 100% in allowed regions
# protein atoms	1,429	1,427
# water molecules	216	183
# Sulfate molecules	0	1

The numbers in parenthesis describe the relevant value for the highest resolution shell.
$R_{sym} = \Sigma|I_i - \langle I \rangle|/\Sigma I$ where I_i is the intensity of the *i*th term observed and $\langle I \rangle$ is the mean intensity of the reflections.

CHAPTER THREE

Organic solvents order the dynamic Switch II in Ras crystals

Greg K Buhrman, Vesna deSerrano and Carla Mattos

Department of Molecular and Structural Biochemistry, North Carolina State University

Raleigh, North Carolina

Published in Structure (Camb). 2003 Jul;11(7):747-51.

ABSTRACT

Room temperature crystal structures of crosslinked H-Ras bound to GMPPNP were solved in 50% 2,2,2-trifluoroethanol, 60% 1,6-hexanediol and 50% isopropanol. The disordered Switch II region of Ras is ordered in the presence of 2,2,2-trifluoroethanol or 1,6-hexanediol. The overall backbone conformation of Switch II in these organic solvents is the same as in the Ras-GMPPNP complexes with RalGDS, PI_3 Kinase and RasGAP, indicating a biologically relevant form. Key polar interactions that stabilize the ordered switch are enhanced in the presence of hydrophobic co-solvents. These results suggest that hydrophobic solvents can be used in general to order short biologically relevant segments of disordered regions in protein crystals by favoring H-bonding interactions between atoms that are highly solvated and mobile in aqueous solution.

INTRODUCTION

The visualization of disordered regions in protein crystal structures has been a long-standing problem. This is particularly significant because in many proteins the most biologically relevant areas are found in short segments containing residues in multiple conformations. The Ras GTPase is an excellent example of a protein in this category and was chosen as a model to study the effects of organic solvents on short disordered segments within the context of the native protein structure. The crystal structure of Ras-GDP shows significant disorder in both switch regions, as indicated by lack of well defined electron density (Tong et al., 1989). Switch I is well ordered in the structure of Ras bound to GTP analogues (GMPPNP or GMPPCP), but Switch II still exists in multiple conformations (Ito et al., 1997). The Switch II of Ras bound to GTP

undergoes a disorder to order transition upon complex formation with protein binding partners (Menetrey and Cherfils, 1999).

In this study, the effects of high concentrations of organic solvents on the structure of Switch II of Ras are explored as a first step in the use of co-solvents for ordering short dynamic segments of proteins so that they may be visualized by X-ray crystallography. The study of proteins in organic solvents is well represented in the literature (Mattos and Ringe, 2001). Alcohols have received particular attention because they are soluble in water and have varying effects on peptide and protein structure (Buck, 1998). In general, alcohols are believed to denature proteins due to disruption of the hydrophobic core, while enhancing formation of local H-bonding interactions, which are optimized in α-helices. The effectiveness of alcohols in stabilizing α-helices increases with the length of the hydrocarbon and the extent of halogenation and decreases with the number of hydroxyl groups (Hirota et al., 1998). In the present study the denaturing effects of three alcohols were minimized by crosslinking Ras in the crystal. The ability of each alcohol to enhance local polar interactions within Switch II parallels its known efficacy in inducing helical structures in peptides. Where the presence of organic solvents induced an ordering of the switch, comparison with the well-ordered Switch II of Ras in published complexes with protein binding partners indicates a biologically relevant conformation.

RESULTS AND DISCUSSION

The crystal structures of the C-terminal truncated form of Ras (H-Ras 166) bound to GMPPNP were solved at room temperature in aqueous mother liquor, in 50% 2,2,2-trifluoroethanol (TFE), in 50% isopropanol and in 60% 1,6-hexanediol (Table 1, Data Collection, Refinement and Validation Statistics). The Ras protein was crosslinked in the crystal with gluteraldehyde before transfer to the organic solvents. The data were collected

at room temperature so that the effects of the organic solvents on the structure of Ras could be distinguished from those resulting from cryogenic conditions.

Isopropanol was included in this study because it is one of the smallest alcohols that can be easily distinguished from water molecules in electron density maps. The analysis of bound organic solvent is important in determining whether the observed effects are due to explicit binding of the molecules or due to bulk properties of the solvent. TFE was chosen as a representative of the halogenated alcohols and because of its extensive use in the study of peptides and proteins (Buck, 1998). 1,6-hexanediol represents a water miscible alcohol with a relatively long hydrocarbon chain.

The cross linked Ras crystals soaked in organic solvents showed no significant change in unit cell parameters that might have resulted from the change in solvent environment. The cross linking is random, with no indication of the crosslinker segments in the electron density maps. In the organic solvents, the overall structure of Ras is also the same as in aqueous mother liquor, with root mean square deviations in the range of 0.3 - 0.5 Å for all atoms. A comparison of each residue across the three solvent environments, however, revealed adjustments of charged or polar side chains that maximize polar interactions between protein atoms, while decreasing the solvent exposed surface. This effect is observed in the presence of TFE and of 1,6-hexanediol, but not of isopropanol and is mainly due to small rearrangements of side chains that are already ordered in aqueous solution. In a few instances significant conformational changes do occur with changes in solvent environments and there are cases where solvent exposed side chains that are disordered in aqueous solution become ordered in the presence of organic solvents. An example is the formation of a new salt bridge between Arg 135 and Asp 132 in 50% TFE that

is not present in aqueous mother liquor. In general, the trends observed for Ras in the presence of organic solvents center around the enhancement of local polar interactions between protein atoms, consistent with the effects observed for peptides in some alcohols and in particular TFE (Buck, 1998; Hirota et al., 1998).

The most significant result due to enhancement of local interactions in the presence of organic solvents is the ordering of the Switch II region of Ras in 50% TFE and in 60% 1,6-hexanediol. Figure 1 shows the electron density for residues 59-67 in the presence of TFE (Figure 1a), 1,6-hexanediol (Figure 1b) and in aqueous mother liquor (Figure 1c). In 50% TFE electron density is present for all of the backbone and most of the side chain atoms of Switch II. The situation is similar in the presence of 1,6-hexanediol, except for a break in the electron density at residue Ala 66. There is only scarce density for Switch II in aqueous solution. The general effect of the halogenated or hydrophobic alcohol is to shift the equilibrium from a dynamic sampling of several conformations (Ito et al., 1997) to a predominant structure that can be observed in the electron density maps. In contrast, in the presence of isopropanol this region remains largely disordered, with electron density similar to that seen for the structure in aqueous solution.

One molecule of TFE, one molecule of 1,6-hexanediol and three molecules of isopropanol are observed in the respective structures. Figure 2a shows the general positions of these organic solvent molecules on the surface of Ras and Figure 2b shows the electron density for each of the molecules. No explicit binding of TFE or 1,6-hexanediol molecules is observed at or near Switch II, suggesting that the ordering of the switch is due to the bulk properties of the co-solvents, rather than to an induced conformation stabilized by explicit interactions with bound solvent molecules. However, a definitive conclusion in this regard

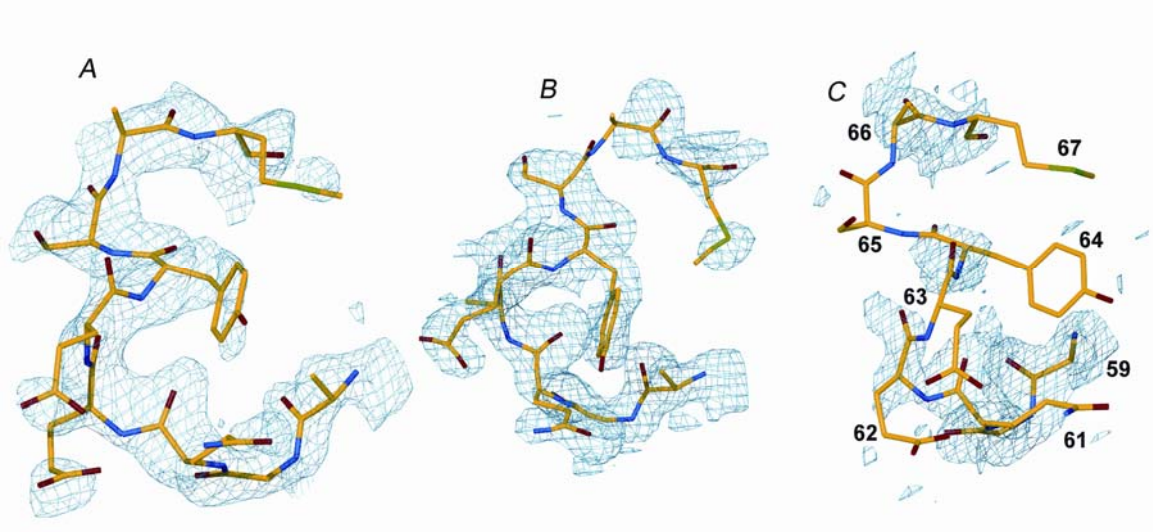


Figure 1. Electron density for the Switch II residues 59-67 in Ras crystals under different conditions: A) soaked in 50% TFE; B) soaked in 60% 1,6-hexanediol; C) in aqueous mother liquor. The electron density in Figure 1C is superimposed on the model of the Switch II in the presence of TFE. The Figures show 2Fo-Fc electron density maps contoured at 1σ.

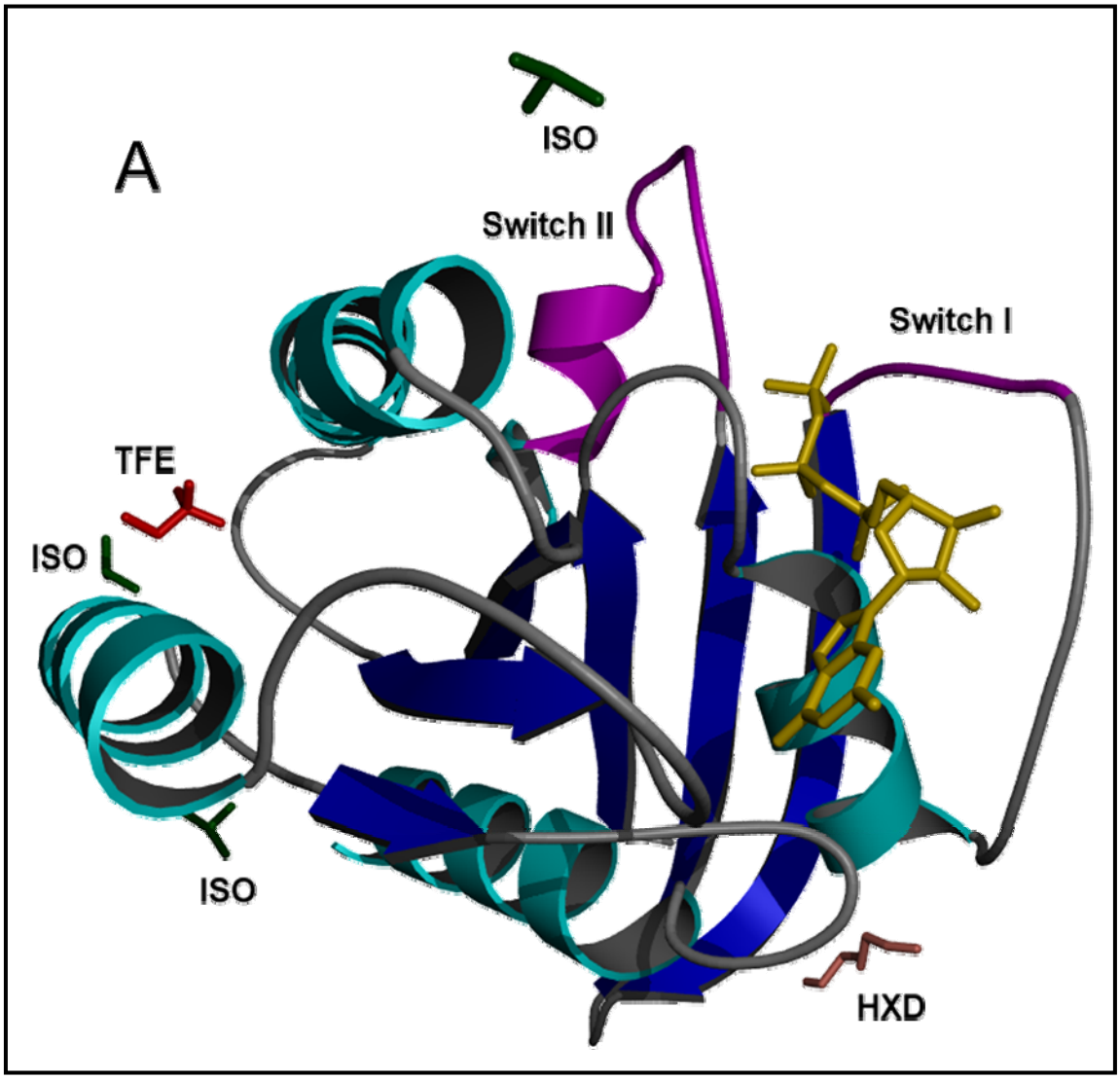


Figure 2. Organic solvents bound to Ras. A) Ribbon diagram of Ras showing general positions of the solvent molecules. The GTP analogue, GMPPNP, is in gold, the isopropanol molecules are green, the TFE molecule is red and the 1,6-hexanediol molecule is orange. B) Electron density for the organic solvent molecules (2Fo-Fc contoured at 1σ). This Figure was prepared using MOLSCRIPT (Kraulis, 1991).

can only be achieved with structures solved at high resolution (2 Å or better). As can be seen in Table 1, there is a correlation between the number of bound organic solvent molecules found in the electron density maps and the resolution at which the structures were solved, see Data Collection, Refinement and Validation Statistics. This correlation is also valid for the number of water molecules observed and is well documented in the literature (Carugo and Bordo, 1999). In the 2.3 Å and 2.4 Å room temperature structures presented here for Ras in the presence of 1,6-hexanediol and TFE respectively, only the most tightly bound solvent molecules are expected to appear in the electron density maps. Therefore, a detailed analysis of solvent structure is not warranted.

A comparison of the Ras-GMPPNP structures in 50% TFE and in 60% 1,6-hexanediol with those of Ras-GMPPNP in complex with the effectors RalGDS (Huang et al., 1998) and PI3K (Pacold et al., 2000) as well as with RasGAP (Scheffzek et al., 1997) shows that the overall backbone conformation of Switch II is the same in all five models (Figure 3). Some local shifts in main chain atoms occur primarily around residues Ser 65 and Ala 66, but the ordered forms of Switch II found in the five different environments have several features in common. The initial part of the switch wraps around Arg 68, which can H-bond to the backbone carbonyl groups of Ala 59, Gly 60, Gln 61 and Glu 62. At the other end of the switch, Asp 69 interacts with the backbone carbonyl groups of Ser 65 and Ala 66 in four of the models. In the Ras-GMPPNP/RasGAP complex, the H-bond between Asp 69 and Ala 66 in Ras is not present. In all five models, Tyr 71 and Met 72 (yellow in Figure 3) form a hydrophobic base that packs against the aliphatic portion of the Arg 68 side chain.

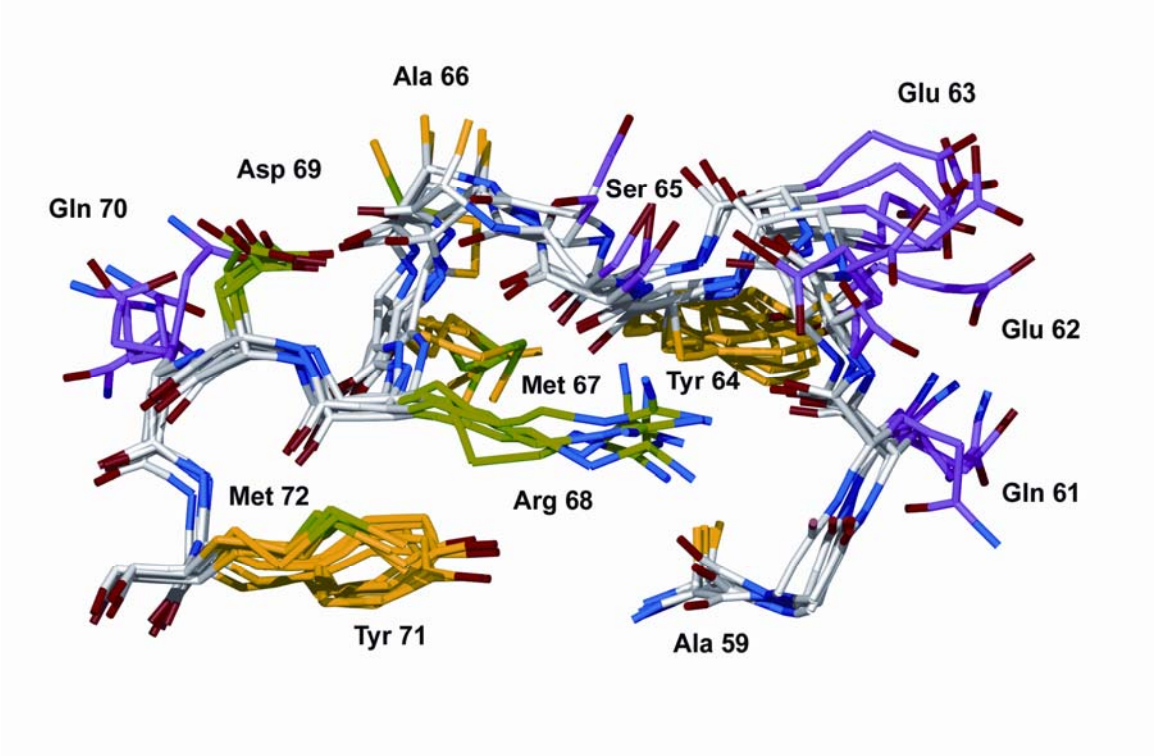


Figure 3. Switch II region in Ras from five models superimposed: in the presence TFE, 1,6-hexanediol and in complex with RalGDS, PI3Kinase and RasGAP. Non-polar side chains are in yellow and polar side chains found in multiple conformations are in purple. Arg 68 and Asp 69 are found in the same conformation in all models and are represented in green.

Residues Arg 68 and Asp 69 (green in Figure 3) are the only charged or polar residues in Switch II that do not interact with protein binding partners in any of the three complexes. Instead, these two charged residues and half of the Switch II main chain carbonyl groups provide a structural frame for the ordered switch, with strengthened polar interactions between protein atoms in Ras. There are several residues protruding from this frame toward the protein surface. Tyr 64 and Met 67 are hydrophobic residues that interact intimately with protein binding partners in the three complexes and are found in very similar conformations in all but the Ras-GMPPNP/RasGAP model, where Met 67 is in a different orientation. Five polar residues, Gln 61, Glu 62, Glu 63, Ser 65, and Gln 70, are found in unique conformations in each of the Ras structures (purple in Figure 3). While each protein binding partner or organic solvent does modulate the detailed features of Switch II, the partial shielding from water appears to be sufficient to induce the features in this region common to the ordered conformations. In each case there is an equilibrium shift from multiple solvated conformations in aqueous solution to a clearly favored conformation due to local enhancement of polar interactions within the switch in a less polar environment.

The effectiveness of the three alcohols used in the present study to order the Switch II in crosslinked Ras crystals correlates with the propensities of these alcohols to induce α-helices in free peptides. The halogenated alcohols and those with longer hydrocarbon chains have been established as more effective than smaller alcohols such as isopropanol (Hirota et al., 1998). The driving force in both the free peptides and Switch II of Ras appears to be the optimization of local polar interactions within the constraints of the system. In unconstrained peptides or proteins the native structure is denatured with the disruption of the hydrophobic core and α-helices are formed which optimize local backbone H-bonding interactions.

Crosslinking the protein in the crystal provides sufficient constraints so that the polar interactions must be formed within the native folded conformation. In general, the crystal structures of proteins solved in the presence of hydrophobic co-solvents may be an informative complement to the structures in aqueous solution, particularly when functionally interesting regions are at stake.

Accession Numbers

The coordinates and structure factors for the four structures have been deposited in the protein data bank with the following accession codes: Ras in aqueous solution, 1P2T; Ras in 50% TFE, 1P2S; Ras in 50% isopropanol, 1P2U; Ras in 60% 1,6-hexanediol, 1P2V.

CHAPTER FOUR

H-Ras 166-GMPPNP wt and Q61L mutant crystallized in the R32 Space Group Have a New Conformation for Switch II that is not distorted by crystal contacts.

ABSTRACT

Ras-GMPPNP crystals with symmetry of the space group R32 provides a structure of the activated form of Switch II undistorted by crystal contacts. In this new structure, Tyr32 is found in the closed conformation, which is shown to exist in solution by ^{31}P NMR spectroscopy. Switch II residues, which interconvert between multiple stable conformations in solution, are found in a conformation not previously seen in Ras GTPase structures, but which is similar to the Switch II conformation of Ran-GTP that is stabilized by Importin-β. The structure of the Q61L oncogenic mutant in a Switch II structure not affected by crystal packing contacts shows a new role for this oncogenic mutant in stabilizing a closed, non-catalytic form of Ras-GMPPNP through hydrophobic interactions with Switch I Tyr32 and Switch II Tyr64.

INTRODUCTION

Harvey-Ras (H-Ras) is the canonical member of a large superfamily of small monomeric GTPase proteins that function as 'molecular switches' in a number of signaling pathways in the cell (Barbacid 1987). Switch II, in particular, is inherently flexible and is disordered or adopts different conformations influenced by protein binding partners, solvent conditions and crystal contacts (Buhrman, de, et al. 2003). Mutations in codon 12, 13, or 61 convert the *ras* gene into an active oncogene. Ras proteins with these mutations have lowered intrinsic and GAP catalyzed GTPase activity. These mutant proteins are constitutively active,

resulting in unregulated cell proliferation and tumor formation. These mutations are so prevalent in human cancer that they are the most frequently occurring gain-of-function mutation detected in human tumors (Cox & Der 2002), (Bos 1989).

^{31}P NMR spectroscopy has identified two distinct, equally populated conformational states of Ras-GMPPNP that are in slow exchange on the millisecond timescale (Geyer, Schweins, et al. 1996). These states correlate to a solvent exposed, open conformation of Tyr32 and a closed conformation of Tyr32 that interacts via a hydrogen bond with the GMPPNP gamma phosphate. Furthermore, ^{31}P NMR spectroscopy showed that these states also exist in Ras-GMPPNP crystals having the symmetry of space group P3(2)21, although only the open form of Tyr32, which is partially stabilized by crystal contacts, is seen in electron density maps (Stumber, Geyer, et al. 2002). There is also ^{31}P NMR evidence that the open form of Tyr32 is stabilized by RasGAP, while the closed form is stabilized by the Ras binding domain of Raf kinase (RafRBD). Our new structure of wild type (wt) Ras-GMPPNP, solved from crystals having the symmetry of space group R32, provides a structure of Ras-GMPPNP with Tyr32 in the closed conformational state. For the rest of this chapter, I will refer to this structure as the R32 form or the closed form of Ras. The crystal structures of Ras-GMPPNP which were solved from crystals having the symmetry of the canonical space group P3(2)21, will be referred to as the P3(2)21 or open form of Ras. The closed conformation of Tyr32 is also seen in non catalytic crystals of Rap2A-GTP, the G12D oncogenic mutant of Ras-GMPPNP (PDB code: 1AGP) and the B and C chain of Ras-GDPCP (PDB code: 6Q21) (Cherfils, Menetrey, et al. 1997), (Pai, Krengel, et al. 1990), (Brunger, Milburn, et al. 1990).

RESULTS

The wild type (wt) and Q61L mutant of Ras-GMPPNP was purified as previously described, see Experimental Procedures. Crystal growth conditions, crosslinking, solvent soaking, data collection and refinement procedures are described in Experimental Procedures. The structure was solved in the R32 form using a protein only model (PDB code: 1CTQ) for molecular replacement. The calculated Mathews coefficient of 2.6 with a solvent content of 53 %, indicated one monomer in the asymmetric unit. Hexane soaked crystals of both wild type and mutant Ras diffracted to high resolution, 1.5 and 1.4 Angstrom respectively. This strategy resulted in an ordered Switch II conformation, which was rebuilt into Fo-Fc difference density contoured at 3 σ when it became clear, from the initial electron density maps, that the canonical Switch II conformation was not correct. Nucleotide, water molecules, hexane molecules and ions were included in the model based on Fo-Fc electron density maps contoured at the 3 σ level. Data collection and refinement statistics are shown in Tables 2.2 and 2.5.

In Chapter three, we investigated the ordering of protein structure by organic solvents. This is not investigated directly in these data sets, collected at cryogenic temperatures, because the ordering effect of organic solvents must be deconvoluted from the ordering of protein structure which occurs by kinetic trapping at low temperature and the cryo-stabilization effects of crosslinking and organic solvents. However, Switch II could only be rebuilt in it's entirety into 3 σ Fo-Fc difference density from crystals which had been stabilized by crosslinking and soaked in organic solvent.

There are a number of available structures of Ras bound to GTP or one of several GTP analogs. Most of these structures were obtained from crystals which formed with the

symmetry of the canonical P3(2)21 space group. Of the 22 structures listed in Table 1, four structures, 1LF0, 6Q21, 1AGP and 1IAQ were solved from crystals which formed with the symmetry of the space groups R32, P2(1), C2 and P2(1)2(1)2(1) respectively. In all of these crystal forms except those that form with the symmetry of the R32 and P2(1)2(1)2(1) space groups, the conformation adopted by the Switch II region is influenced by crystal packing forces, see Figure 1. In the P2(1)2(1)2(1) form, the Switch II region was too disordered to be modeled in the electron density maps(Spoerner, Herrmann, et al. 2001). In crystals that form with the symmetry of the P3(2)21 space group, Switch II is conformationally restricted and cannot adopt the conformation seen in our R32 form. The structure of Ras bound to GTP solved in the R32 space group (PDB code: 1LFO) is of the A59G mutant (Hall, Bar-Sagi, et al. 2002). Interestingly, this mutation results in a Switch II conformation distinct from the Switch II conformation we see in the wild type R32 form, see Figure 2. The Phi/Psi angles of Gly59 are in a generally allowed portion of the Ramachandran Plot. However, replacement of an alanine residue with glycine, which has less restrictive backbone dihedral angles, at the start of the Switch II region, may give the entire region greater conformational flexibility than is seen in the wt protein, allowing Ras-GTP A59G to adopt a distinct Switch II conformation. The switch regions in this mutant are found in a conformation intermediate

Table 1. Ras bound to multiple GTP analogs has been solved in crystals which form with the symmetry of space groups P3(2)21, C2, P2(1), R32 and P2(1)2(1)2(1).

PDB ID	bound nucleotide	Space Group
121P	GTO	P 32 2 1
1AGP	**GNP**	**C 2**
1CLU	DBG	P 32 2 1
1CTQ	GNP	P 32 2 1
1GNQ	CAG	P 32 2 1
1IAQ	**GNP**	**P21 21 21**
1JAH	GTO	P32 2 1
1LF0	**GTP**	**R 3 2**
1P2S	GNP	P 32 2 1
1PLJ	CAG	P 32 2 1
1QRA	GTP	P 32 2 1
1RVD	DBG	P 32 2 1
221P	GNP	P 32 2 1
5P21	GNP	P 32 2 1
621P	GNP	P 32 2 1
6Q21	**GTO**	**P 21**
521P	GNP	P 32 2 1

Structures from the Brookhaven Protein Data Bank shown in bold were solved in non-canonical space groups and are discussed in the text. They include: 1AGP (Pai, Krengel, et al. 1990), 1IAQ(Spoerner, Herrmann, et al. 2001) , 1LFO (Hall, Bar-Sagi, et al. 2002) and 6Q21(Brunger, Milburn, et al. 1990) . Abbreviations: GTO: Guanosine-5-[β–γ–methylene] triphosphate. GNP: Guanosine-5-[β–γ–imido]triphosphate. DBG:Diaminobenzophenone-β–γ–imido-GTP. CAG: Caged GTP.

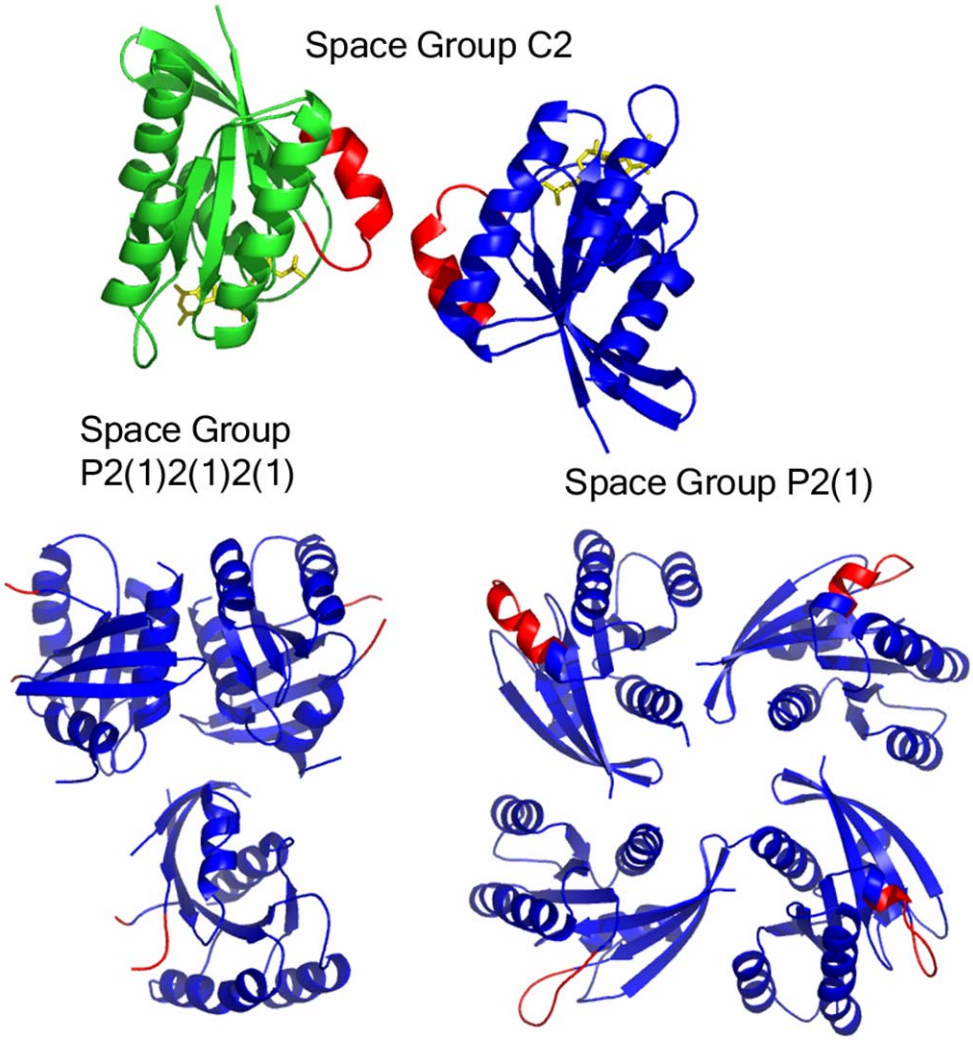
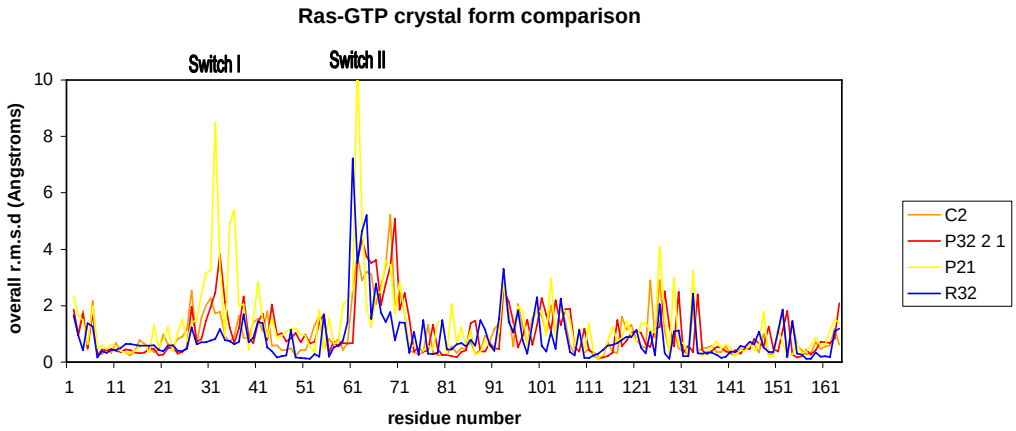


Figure 1. Three crystal forms of Ras bound to a GTP analog result in different local environments, and different structures of Switch II, colored red. All Figures in this chapter, were generated using the PyMOL Molecular Graphics System (DeLano Sientific, San Carlos, CA).

A.



B.

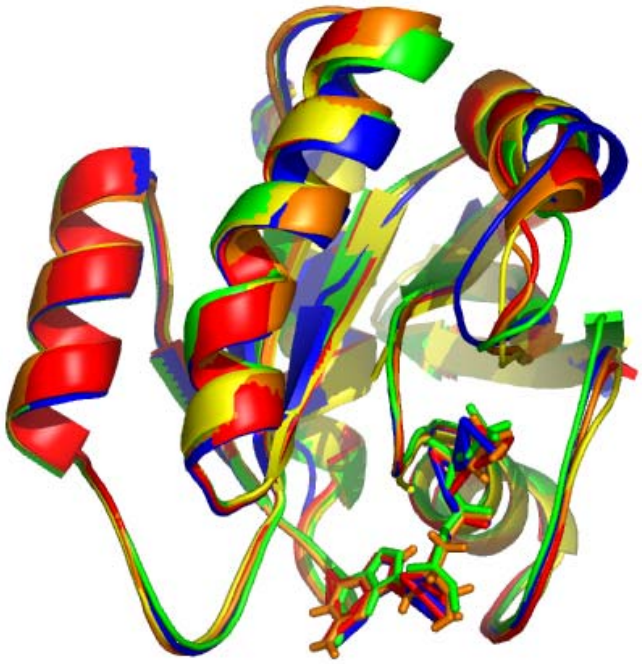


Figure 2. A. Models of Ras bound to a GTP analog from each crystal form were superimposed and the overall atom r.m.s.d. of each structure from the native R32 form was calculated for each residue and plotted in A. The largest differences occur in Switch I and Switch II. B. A cartoon representation, with the wild type R32 form shown in green and the same color scheme used in A, of the structures used to calculate the overall r.m.s.d.

between the GTP and GDP bound forms and had been previously observed in a molecular dynamics simulation study where the path between the two states of Ras was determined (Ma, J.P. and Karplus, M. 1997). The conformation of the switch regions in the Ras-GTP A59G mutant is therefore not representative of the active state of the wild type protein. The structure of Ras bound to GMPPCP (6Q21) that was obtained from crystals which formed with the symmetry of P2(1), contains four molecules of Ras in the asymmetric unit, which provide evidence of conformational flexibility within the Switch regions. Because of the noncrystallographic symmetry, the Switch II regions in each of the four molecules is in a unique local environment within the crystal. In each case, Switch II adopts a distinct conformation, modulated by crystal packing constraints. Of the four molecules in the asymmetric unit, only one, molecule C, has a well ordered Switch II. The structure of Ras-GMPPNP was also solved from crystals which form with the symmetry of space group C2 (1AGP). This protein is also a mutant form of Ras, where glycine 12 is replaced with aspartate. In this crystal form, Switch II is in direct crystal contact with the Switch II of a symmetry related molecule. In Figure 2, we plot the overall root mean square deviation (RMSD) of the native Ras-GMPPNP R32 form versus representatives of Ras solved from the other crystal forms of Ras bound to a GTP analog. Our structure of wt Ras bound to GMPPNP, in the R32 form, has a unique Switch II conformation that is undistorted by crystal contacts and is quite different from the Switch II conformations which, in other crystal forms, are stabilized by crystal packing forces. In the next section, we compare the R32 form with the P3(2)21 form, since the P3(2)21 form of Switch II is biologically relevant and is found in several Ras binding protein complexes.

Ras-GMPPNP R32 vs. P3(2)21 Space Group

Crystals of the R32 form of Ras were found in the presence of 0.2 M calcium chloride. The overall structure of the R32 form is very similar to the P3(2)21 form of Ras. The areas of significant differences include Switch I, Switch II and surrounding residues. In the R32 form, most of the crystal contacts occur in Switch I, residues 30-38. The Switch I residues in adjacent monomers of the unit cell interact via a 2-fold symmetry axis that forms around a calcium binding site between symmetry related Switch I residues, see Figure 3. This crystal form leaves Switch II free of any crystal packing constraints, allowing the Switch II region to adopt a unique conformation, not seen in other Ras crystal structures. The Switch II helix shifts closer to Switch I, allowing helix 3 to shift approximately 2 Å closer to Switch II, resulting in a shift of the loop preceding helix 3 (average Cα rmsd of 2 Å for residues 99-108). Interestingly, this is part of a larger region (residues 94-109) that also exhibits a shift in the Rap structure, relative to Ras. The differences in crystal packing between the canonical P3(2)21 form and the R32 form are shown in Figures 3 and 4. Tyr32 in the P3(2)21 form, makes a hydrogen bond with the gamma phosphate of a symmetry related GMPPNP molecule, stabilizing the open conformation of Tyr32. While the Switch II residues in the P3(2)21 form are not stabilized directly by crystal contacts, the proximity of a symmetry related Switch II results in a conformational restriction in this region, such that Switch II is unable to adopt a full range of conformations in the P3(2)21 form.

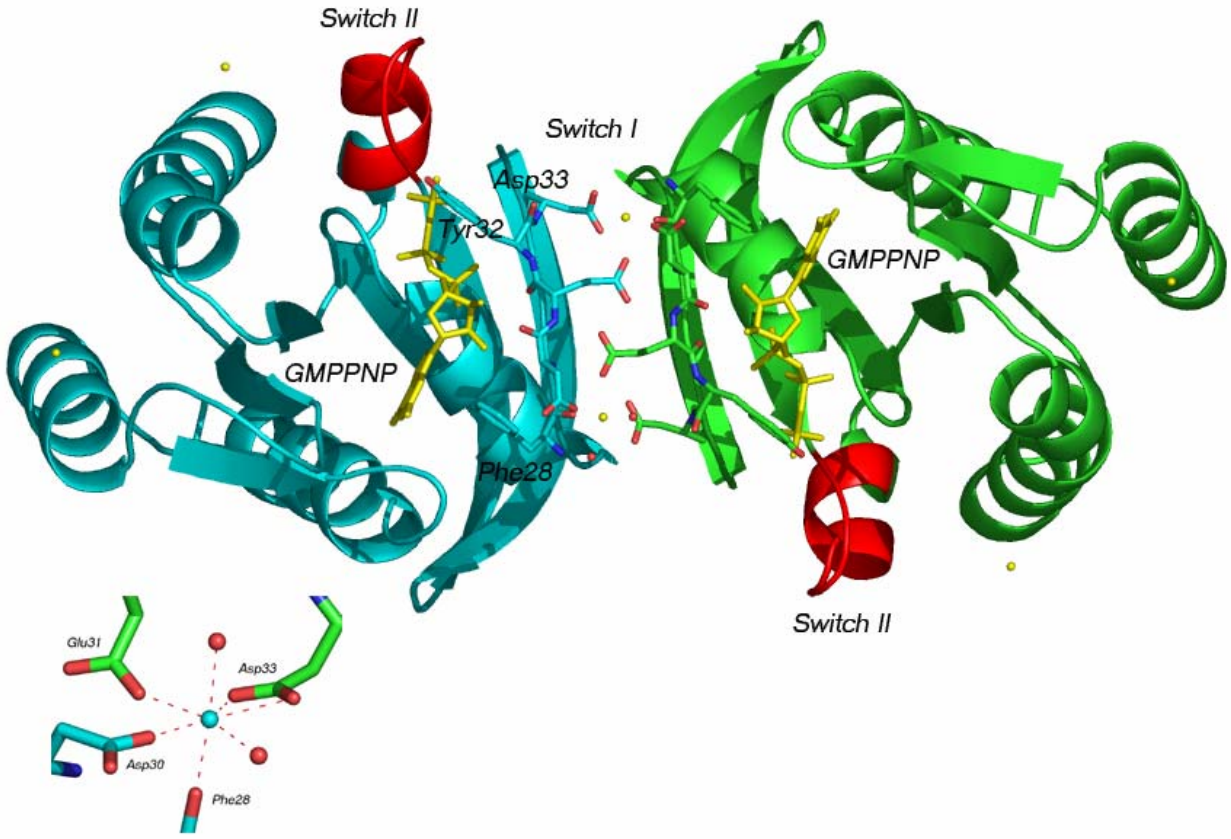


Figure 3. Crystal packing in R32 along a 2-fold symmetry axis between Switch I residues. A. Switch I residues 28-33 are shown in stick. Switch II residues 59-72 are shown in cartoon and colored red. The non-hydrolyzable GTP analog, GMPPNP, is shown in yellow. B. Closeup of Calcium binding site, with dotted lines indicating coordination. Coordination distances are all from 2.3-2.4 Angstrom.

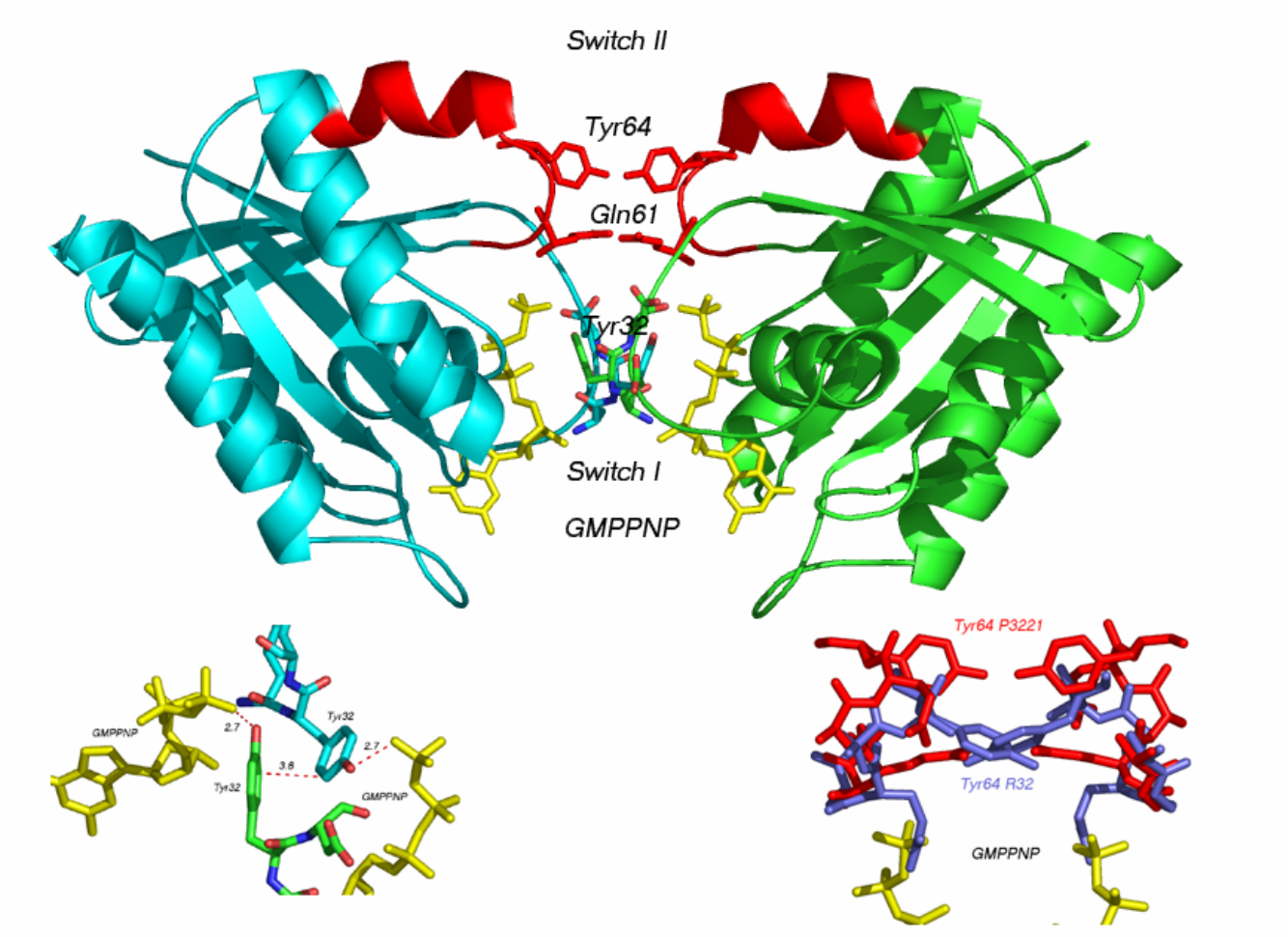


Figure 4. Crystal packing in P3(2)21 along a 2-fold symmetry axis between SwitchI and Switch II residues. A. Switch I residues 31-33 are shown in stick. Switch II residues 61 and 64 are shown in stick, with residues 59-72 colored red. The GTP analog, GMPPNP, is shown in yellow. B. Closeup of Tyr32 in the open conformation, hydrogen bonding to the gamma phosphate of a symmetry related molecule of GMPPNP. C. Closeup of Tyr64 in the P3(2)21, in red, and the R32 conformation, in slate. The R32 conformation was generated by applying the P3(2)21 symmetry operator to the wt Ras-GMPPNP structure solved in the R32 space group and superimposed on the Ras-GMPPNP structure solved in P3(2)21, to show that the Switch II conformations that Ras-GMPPNP can access in the P3(2)21 space group are restricted, and the R32 Switch II conformation is not allowed.

Calcium Ion binding Sites

The R32 form has several calcium binding sites. This is not surprising, given the crystallization buffer includes 0.2 M $CaCl_2$. Many of these binding sites are located at crystal contacts, contributing to formation of this crystal form of Ras. One calcium binding site is on a crystal contact between Sym Asp33, Sym Glu31 and Asp30. The backbone oxygen of Phe28 and 2 water molecules complete the coordination sphere (see Figure 3B, close up of calcium coordination). This site is located within a large interface between symmetry related Switch I residues. Five of the seven coordinating oxygen atoms belong to the protein. The coordination number is 7 and because Asp33 acts as a bidentate residue the coordination geometry can be considered as either a distorted octahedral, or a distorted pentagonal bipyramidal. By forming tight crystal packing constraints with residues Glu31 and Asp33, this crystal form stabilizes the closed conformation of Tyr32. The positively charged calcium ion enforces a Switch I conformation that is similar to what is seen in the Raps-Raf complex, where Lys84 from Raf kinase makes a salt bridge with the side chains of Glu31 and Asp33, shown in Figure 5A. A second calcium binding site lies on a 2-fold axis, formed by symmetry related Asp105 residues. Coordination is accomplished by symmetry related Asp105 side chain oxygens, the backbone oxygen of symmetry related Arg 102 residues and 1 water molecule. The coordination number is 7 and the geometry is distorted pentagonal bipyramidal due to Asp105 and SymAsp105 coordinating in a bidentate manner. (This site could also be considered a distorted octahedral with 1 missing oxygen.)

Switch I

Tyr32 adopts a closed conformation in the R32 form, and is positioned to make a good hydrogen bond (2.7 Å) with a gamma phosphate oxygen of GMPPNP. This conformation of Tyr32 would clash sterically with Arg789 of RasGAP, which stabilizes the open conformation of Tyr32, shown in Figure 5B. Arg789 is the "arginine finger" which is essential for GAP catalyzed GTP hydrolysis (Resat, Straatsma, et al. 2001). A large conformational change involving Switch I and Switch II residues would be necessary to transform Ras from the R32 form to the catalytically active structure seen in the Ras-RasGAP complex.

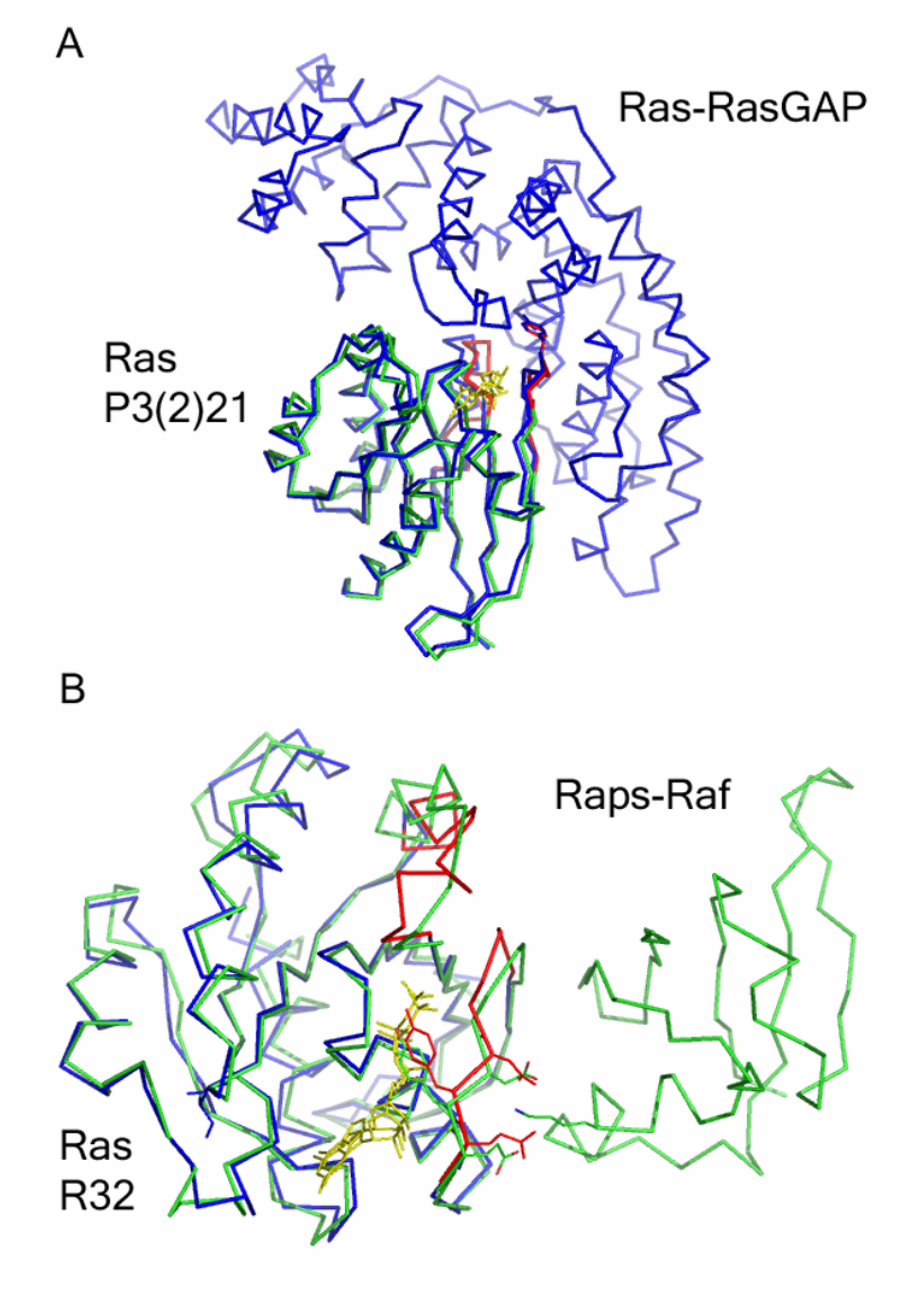


Figure 5. A. The Ras-RasGAP complex (1WQ1), in blue, is superimposed with a structure of Ras-GMPPNP solved in the P3(2)21 space group, in green, with the Switch regions colored red and Tyr32 shown in stick. B. The Raps-Raf complex (1GUA) in green, is superimposed with a structure of Ras-GMPPNP solved in the R32 space group, in blue, with the Switch regions colored red and residues 31-33 from Ras and Lys84 from Raf shown in stick.

Switch II conformation

Switch II is a highly dynamic, mobile region of the protein involved in many protein-protein interactions. One complication in analyzing the conformational flexibility of this region is crystal contacts. In most cases where a structure for Switch II is available, it is stabilized either by binding partners or by protein: protein interactions within the crystal. In cases where the Switch is stabilized by cryogenic conditions or organic solvents, crystal packing can still influence the conformations adopted by this region. In the R32 form, no symmetry related residues are within 10 Angstrom of Switch II, allowing Switch II to adopt a previously unobserved conformation. This conformation of Switch II is distinct from other conformations of Switch II that are stabilized by Ras effector proteins RalGDS and PI3Kinase, however it is similar to the Switch II conformation found in the Importin β-Ran-GTP complex, see Figure 6. Importin β inhibits the intrinsic and GAP catalyzed hydrolysis of Ran-GTP by stabilizing a non catalytic conformation of Switch II (Vetter, Arndt, et al. 1999). Inhibition of GTP hydrolysis by Importin β is an important aspect in the spatial control of nuclear import and cargo. Importin β that has released its cargo protein is transported by Ran-GTP from the nucleus to the cytosol, where Importin β is released to pick up a new cargo protein and GAP catalyzed GTP hydrolysis converts Ran to the GDP bound form. Importin-beta is the first protein identified in what is becoming a family of Ran-GTP binding proteins that inhibit intrinsic and GAP catalyzed GTP hydrolysis, (i.e. RanBP7 and RanBP8) (Gorlich, Dabrowski, et al. 1997)

The Switch II conformation in Ras complexes with binding partners RalGDS, PI3Kinase and RasGAP shares a common backbone architecture, stabilized by hydrogen bonds formed between Arg68 and backbone carbonyl groups, and common

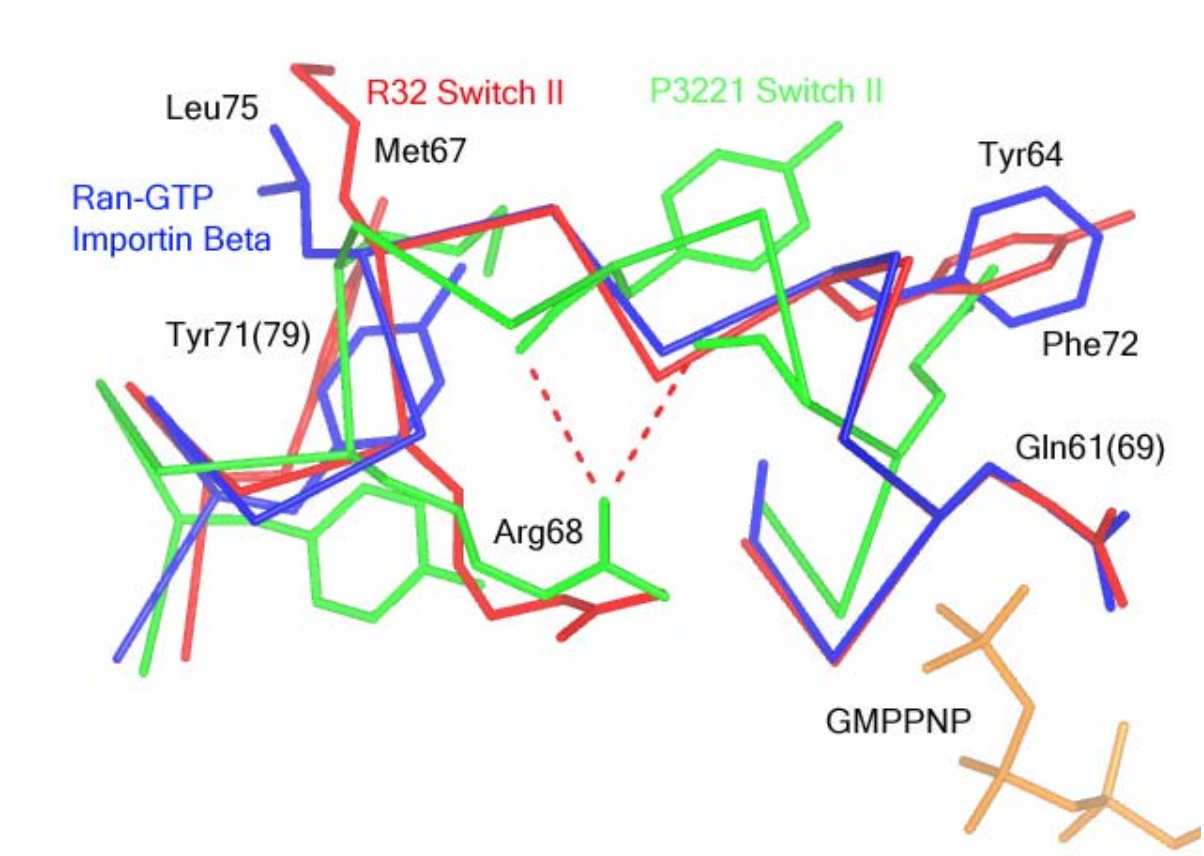


Figure 6. Switch II from R32 Ras-GMPPNP is similar to Ran-GTP : importin-beta and different from P3(2)21 Ras-GMPPNP. The Cα backbone trace is shown for residues 61-71, for Ras R32, colored red, and Ras P3(2)21, colored green. The Cα backbone for residues 69-79 of Ran-GTP is colored blue. Residues of interest are shown in stick. Even though the sequence conservation is not complete, the overall structure of the Switch is conserved between Ras and Ran.

conformations of nonpolar sidechain groups known to be important for binding affinity between Ras and Ras effector proteins (Buhrman, de, et al. 2003). In the R32 Switch II conformation, a different backbone architecture, which is not well stabilized by Arg68, forces alternate conformations of the nonpolar side chain residues, see Figure 6. Tyr71 adopts a buried conformation in most GTPases, acting to stabilize Arg68, but in R32, this residue adopts a solvent exposed conformation. Arg68 relaxes into a conformation which makes no hydrogen bonds to the carbonyl oxygen of Tyr64. Tyr64 extends towards Switch I and is sandwiched between hydrophobic residues Ile36 and Pro34. Tyr64 has steric clashes with the conformation of Gln61 seen in the P3(2)21 form. The function of Gln61 in GTP hydrolysis is to stabilize the transition state of GAP catalyzed GTP hydrolysis (Topol, Cachau, et al. 2004). The transition state of GAP catalyzed GTP hydrolysis is most clearly shown in the structure of the Ras-GDP- Aluminum Fluoride - Ras-GAP complex. The Switch II conformation seen in the R32 form, which by analogy to the Ran-importin β complex is in a non-catalytic conformation, is incompatible with the catalytic Switch II conformation stabilized by Ras-GAP.

Further evidence implicating the R32 form as a non catalytic conformation comes from analysis of the Ras-RasGAP complex, see Figure 7. Site directed mutagenesis studies have shown that RasGAP residue Leu902 affects catalysis by acting to stabilize a GTP hydrolysis competent conformation of Switch II. Mutation of Leu902 to Ile results in reduced catalytic activity, but not in reduced binding affinity (Ahmadian, Kiel, et al. 2003) In the R32 form, Tyr64 would have a severe steric clash with Leu902 of RasGAP.

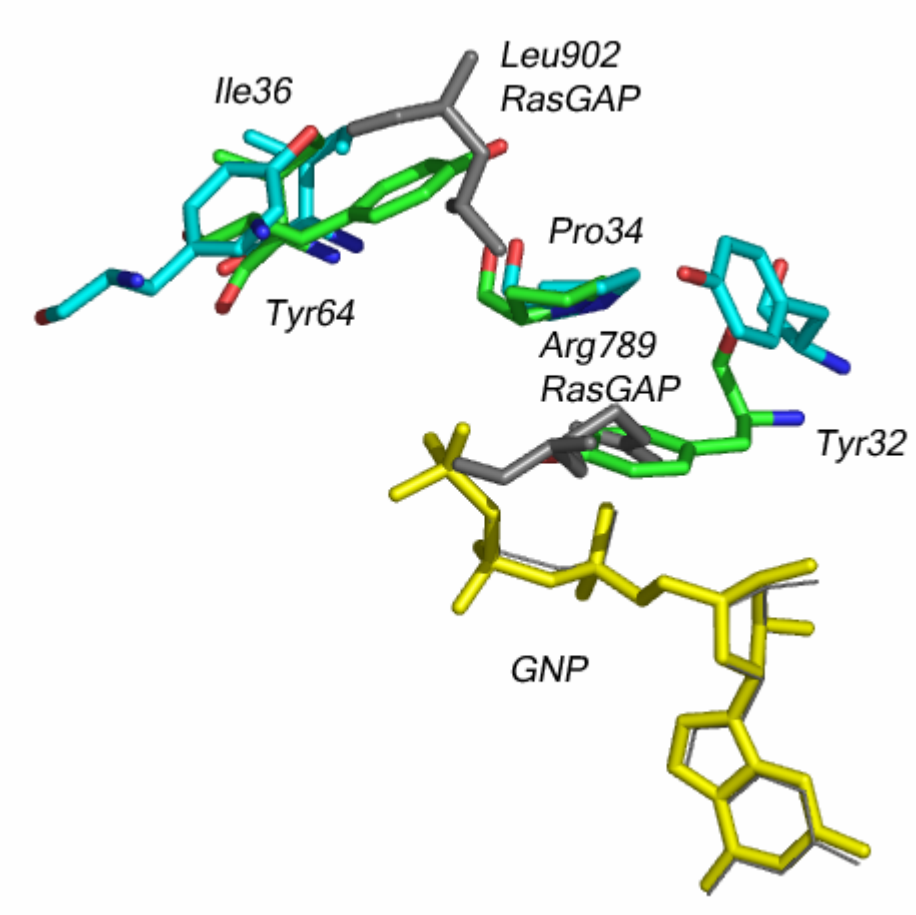


Figure 7. The R32 Switch is not in the proper conformation to interact with Ras-GAP. Ras residues are green, Ras bound to RasGAP (PDB code 1WQ1) residues are cyan, RasGAP residues are grey. Nucleotide from Ras is shown in yellow stick and nucleotide from Ras:RasGAP is shown as a gray line.

Q61L Ras-GMPPNP mutant solved in R32 space group

^{31}P NMR shows that the Ras Q61L constitutively active mutant bound to GTP also exists in the open and closed Tyr32 conformation, although the two conformations are only observed in wild type (wt) Ras bound to a nonhydrolyzable GTP analog , GMPPNP, but not in wt Ras bound to GTP (Geyer, Schweins, et al. 1996). Presumably, hydrolysis of bound GTP in solution interferes with the ability to detect these conformations by NMR. We have solved the crystal structure of the Q61L mutant of Ras-GMPPNP in the R32 form. In the R32 form, this mutation results in a contiguous hydrophobic patch between residues Tyr64, Pro34, Leu61 and the closed Tyr32, see Figure 8. The additional hydrophobic interactions provided by Leu61 may further stabilize the non catalytic conformation of Ras, when Tyr32 is in the closed conformation, decreasing the likelihood of productive interactions with Ras-GAP. Low B-factors for Leu61 and the surrounding residues and good electron density for this region support this interpretation, see Figure 9. By contrast, in wt Ras bound to GMPPNP, the Switch II region is more disordered, with higher B-factors and electron density that is not well defined, particularly for the Q61 side chain. This interpretation is in agreement with linear free energy relationship calculations showing that while many oncogenic mutants lower the rate of GTP hydrolysis by lowering the pKa of the gamma phosphate, Q61L mutant lowers the rate of GTP hydrolysis by a different mechanism which does not affect the pKa of the gamma phosphate (Schweins, Geyer, et al. 1996). This is also in agreement with recent quantum mechanics calculations which demonstrate that Q61L lowers GTP hydrolysis by not allowing formation of the catalytic conformation of Ras (Shurki & Warshel 2004)

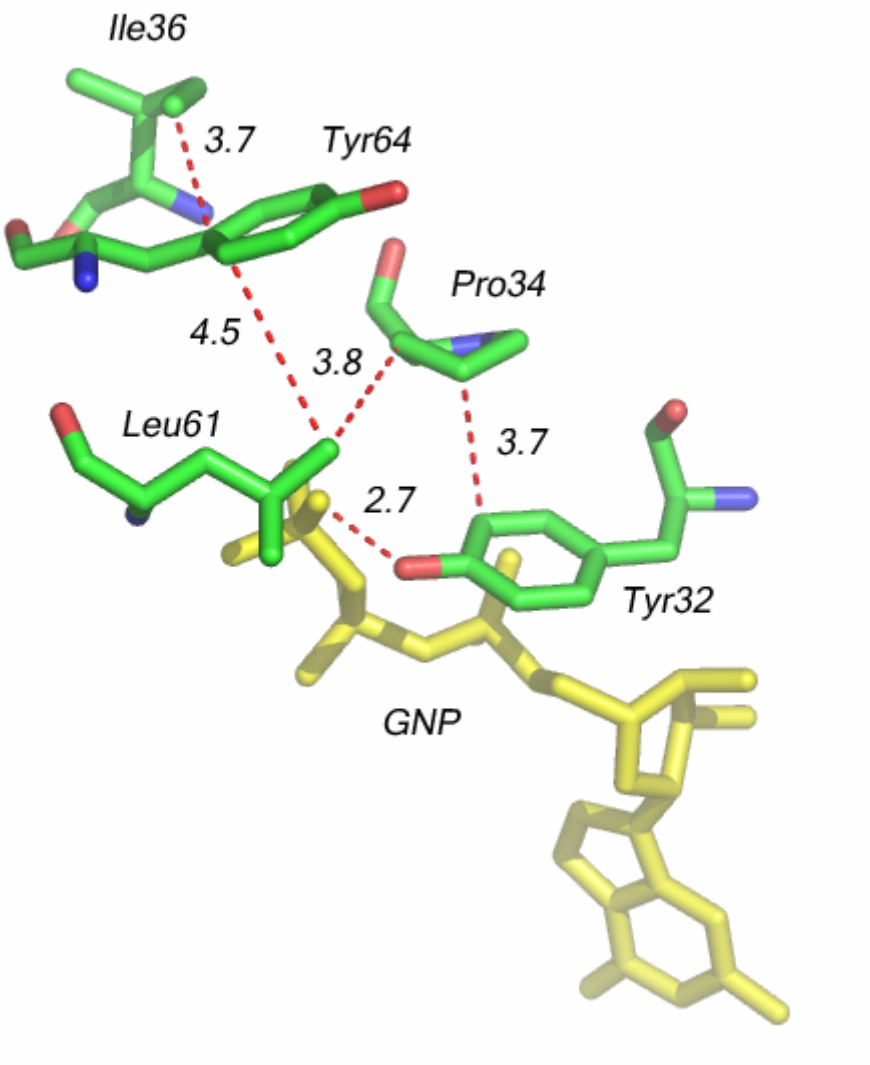


Figure 8. The Q61L mutant forms a contiguous hydrophobic patch with residues Ile36, Tyr64, Pro34,Leu61 and Tyr32.

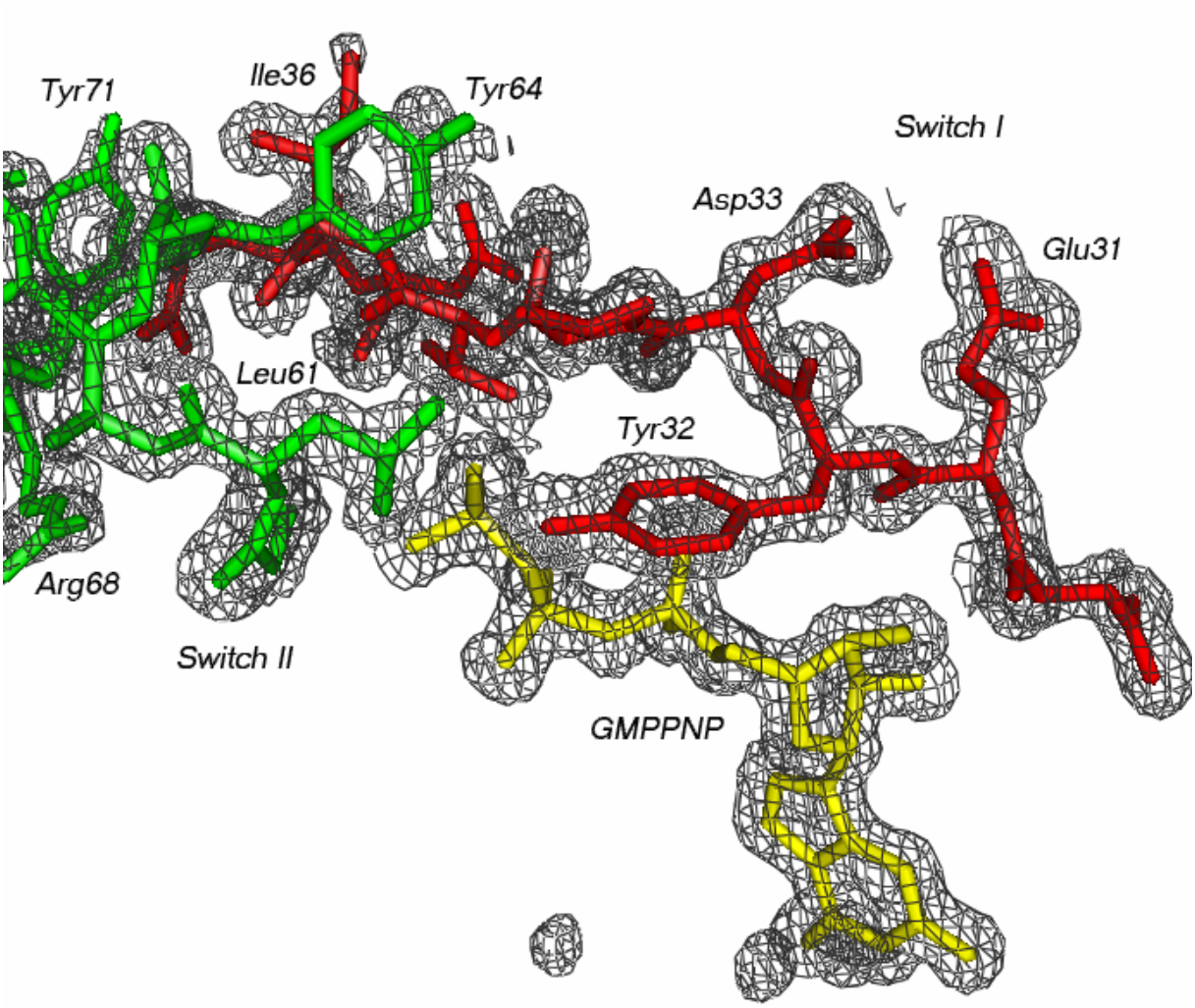


Figure 9. Electron density for Q61L mutant. Switch I residues are colored red. Switch II residues are green and the nucleotide is yellow. Electron density is from 2Fo-Fc maps contoured at 1.2 σ.

DISCUSSION

Several NMR and molecular dynamic studies indicate that GTPase binding proteins interact with GTPases by selecting various preexisting conformations of the free protein. More specifically, ^{1}H, ^{13}C and ^{15}N NMR studies of Ras-GMPPNP showed that residues 10-13, 21, 31-39, 57-64 and 71 interconvert slowly between 2 or more stable conformers (Ito, Yamasaki, et al. 1997). These studies complement the ^{31}P NMR work, showing interconversion between the open and closed forms of Tyr32 (Geyer, Schweins, et al. 1996). Ito et. al. propose that the conformational transition of Tyr32 seen in the ^{31}P NMR studies is part of a larger conformational interconversion that is centered on the gamma phosphate. The crystal packing evident in crystals solved in the R32 form that traps the closed conformation of Tyr32, provides a unique view of the Switch II of Ras-GMPPNP that is free from crystal packing influences and represents one of the stable conformations of Switch II that is sampled by Ras-GMPPNP.

Based on structural comparisons with Ran-GTP bound to Importin-β, a non catalytic crystal structure of Rap-GTP, and the Ras-RasGAP structure, we infer that the R32 form is in a non-catalytic conformation. Structural comparisons illustrate an important role for residues Tyr71, Tyr64 and Tyr32, all of which exhibit regional polysterism in Ras-GMPPNP protein in solution. We hypothesize that the packing interaction between Tyr71 and Arg68 is important in maintaining the Switch II effector/GAP binding conformation. Therefore, stabilizing a solvent accessible conformation of Tyr71, that does not interact with Arg68, may shift the equillibrium to the non catalytic conformation. This solvent exposed conformation is also seen in the Ran- Importin β complex. Tyr64 also shares a similar

conformation in Ran-GTP – Importin β complex and the R32 form of wt Ras-GMPPNP. The conformation of this residue influences the conformation adopted by Gln61, a crucial residue in GTP hydrolysis. Finally, Tyr32 in the closed conformation in all three structures (R32, Ran-GTP – Importin β (PDB code: 1IBR)) and the non-catalytic Rap2A-GTP (PDB code: RAP3) is positioned to hydrogen bond with a gamma phosphate oxygen.

Our structure provides evidence that the Switch II region of unbound Ras-GTP spends part of its time in a conformation unfavorable for GTP hydrolysis. This helps explain the low intrinsic hydrolysis rate of the GTPase, which is an important, functional characteristic of most small, monomeric GTPases. In many GTPases, intrinsic GTP hydrolysis controls the off-rate of the downstream effector protein, providing a 'timing mechanism' for the molecular switch. It is unlikely that Ras effector proteins which activate downstream signaling pathways in a Ras-GTP dependent manner would stabilize a GTP hydrolysis incompetent form of Ras. This would result in a signal that is potentially always on, which leads to cancer in oncogenic mutants of Ras. More likely, Ras binding proteins that stabilize the non catalytic form of GTPases in the Ras superfamily may function as transport proteins, like Importin β, or as storage proteins, providing a pool of activated GTPase that can readily activate downstream signaling pathways, simply by changes in specific protein: protein interactions.

The structure of wt Ras-GMPPNP in the R32 form provides an interesting context to understand the role of the oncogenic mutant Q61L. The role of Gln61 in GTP hydrolysis is the focus of much active research. Traditionally, Gln61 is thought to participate in GTP hydrolysis directly, by activating a conserved water molecule for nucleophilic attack of the gamma phosphate. However, a recent molecular dynamics study indicates that Gln61

primarily serves to stabilize the transition state of GAP catalyzed GTP hydrolysis (Topol, Cachau, et al. 2004). Unlike structures of the Q61L mutant solved in the Tyr32 open conformation of Ras-GTP (PDB code 721P), the structure of the Q61L mutant solved in the Tyr32 closed conformation shows that this mutant can form a contiguous hydrophobic patch with critical residues Tyr64 and Tyr32. Tyr32 is held in a solvent exposed conformation by the RasGAP protein which allows Arg789, the so-called arginine finger, access to the nucleotide phosphate groups. Stabilization of the buried conformation of Tyr32 by Leu61 may further stabilize a conformation of Ras-GTP that is insensitive to GAPs and unable to hydrolyze GTP.

CHAPTER FIVE

MSCS Analysis of H-Ras Identifies Known Protein Binding Sites and Reveals New Putative Protein Binding Sites

ABSTRACT

The surface of Ras was probed with multiple organic solvents (MSCS method). The solvents that were used included: 50 % 2,2,2 trifluoroethanol (TFE), 60% 1,6 hexanediol (HXD), neat cyclopentanol (PEN), neat hexane (HXN), 50% isopropranol (ISO), 90 % R,S,R bisfuranol (BIR), 55% dimethylformamide (DMF) and 70% glycerol (GLY). The MSCS work was done in Ras-GMPPNP crystals having the symmetry of space group R32, which represents the activated form of Ras that interacts with effector proteins such as Raf kinase.

INTRODUCTION

H-Ras is a member of a large group of proteins that form non-obligate, transient, hetero-complexes with other proteins as part of its functional role in the cell. These complexes can be difficult to characterize structurally. In many cases, even though a crystal structure of the unbound protein is available, efforts to crystallize the protein complex fail. Therefore, the development of alternative experimental or computational methods to analyze and predict protein-protein interactions is an important area of current biochemical research. In this chapter, we describe the use of the Multiple Solvent Crystal Structures method (MSCS) to predict areas of the Ras protein that are likely to be involved in protein-protein interactions. Prior to this study, MSCS has been used to study enzyme-substrate interactions of the proteases Chymotrypsin, Trypsin, Subtilisin and Elastase, reviewed by (Ringe &

Mattos 1999). The interactions characterized by MSCS in these studies involved small protein-peptide interactions that are highly specific and occur within a large active-site cleft. H-Ras is a good test case for extending MSCS to the study of larger protein complexes because it interacts with several binding partners and is part of a protein family for which there is a large amount of structural and biochemical data, reviewed by (Corbett & Alber 2001).

In this study, we have probed the entire surface of Ras using the Multiple Solvent Crystal Structures method (MSCS). MSCS is an experimental mapping approach that involves solving the x-ray structure of a protein in a variety of organic solvents at high concentrations (50% or greater) (Mattos & Ringe 2001). The x-ray structure of a number of proteins in one or two organic solvents have been solved, including Lysozyme (Wang, Zhu, et al. 1998) and the proteases Thermolysin (English, Done, et al. 1999), Subtilisin (Schmitke, Stern, et al. 1998) and Chymotrypsin (Yennawar, et al. 1994). The most extensive characterization of proteins in organic solvents has been done with the protease Elastase (Mattos & Ringe 2001). The MSCS characterization of Elastase illustrates several important trends in mapping protein binding sites by this method. In Elastase, organic solvent molecules cluster at the protein active site in known binding pockets. With few exceptions, other solvent molecule binding sites are at crystal contacts and do not exhibit the clustering effect.

Many of the Ras-complexes that have been characterized structurally are complexes of minimum binding domains only. Regions of Ras that may interact with binding partners outside of the minimum binding domains are overlooked in these studies. One example of this is the Ras-Raf kinase complex. The available crystal structure is of a Ras homolog, Rap

that was mutated to resemble Ras (Raps) bound to the minimum Ras binding domain of Raf kinase. The structure of this complex shows that the Raf binding domain of Raps consists of the core effector binding domain residues containing Switch I (Zeng, Treutlein, et al. 1999). However, NMR experiments with the full length Raf kinase demonstrate that Ras residues outside of the effector binding domain contribute to Raf kinase binding (Campbell, Khosravi-Far, et al. 1998). Using MSCS, we probe the entire surface of Ras for regions likely to participate in protein: protein interactions. We use the known structural information on Ras protein binding sites to validate the MSCS technique as applied to detecting protein binding sites and to identify organic solvent affinity hotspots as new putative protein binding sites on Ras. Ras is a small, monomeric protein in solution with the known protein binding surface dominated by polar, hydrophilic residues which interact well with water. This necessitated the use of polar organic solvents to successfully solvent map the Ras protein binding surface.

RESULTS

The surface of Ras was probed with multiple organic solvents (MSCS method). The solvents that were used included: 50 % 2,2,2 trifluoroethanol (TFE), 60% 1,6 hexanediol (HXD), neat cyclopentanol (PEN), neat hexane (HXN), 50% isopropranol (ISO), 90 % (3R, 3aS, 6aR)- hexahydrofuro[2,3-b] furan -3-ol (R,S,R-bisfuranol, BIR), 55% dimethylformamide (DMF) and 70% glycerol (GLY). H-Ras 166 bound to GMPPNP was purified and crystallized as previously described, see Experimental Procedures. Crosslinking, solvent soaking, data collection and refinement procedures are also described in Experimental Procedures. Data collection and refinement statistics are shown in Table 2 (Data Collection, Refinement and Validation Statistics). Representative electron density from 2Fo-Fc electron density maps contoured at 1 σ is shown for selected organic solvents in

Figure 1.

Global effects of organic solvent soaking

Since organic solvents have a lower dielectric constant than water, the effects of electrostatic screening are reduced in solvent soaked protein crystals. One result of this effect is strengthening intra-protein polar interactions, including hydrogen bonds and salt bridge formation. Polar surface residues are less solvent exposed and non-polar residues are more solvent exposed. In the room temperature studies with Ras-GMPPNP in the P3(2)21 form, there were examples of surface salt bridges that formed in the organic solvent soaked crystals, but not in the aqueous soaked crystals. Similar results are seen in the 100 K crystal structures. Comparing the structure solved from crosslinked crystals in aqueous solution to the one in the presence of hexane, the most non-polar organic solvent used, we see several examples of these effects. The solvent accessible surface area of each residue from the crosslinked, aqueous control and from hexane was calculated in CNS. The result is given as the ratio of solvent exposure in hexane to solvent exposure in water. In general, polar residues are less solvent exposed in hexane vs. water and non-polar residues are more solvent exposed in hexane vs. water. One example is Lys88 (16 $Å^2$ solvent exposed sidechain in hexane/ 22 $Å^2$ solvent exposed sidechain in water), which is solvent exposed and makes no interactions with the protein atoms in aqueous solution, but forms a salt bridge with Asp92 in hexane. Another example is Arg128 (15/18), which forms a salt bridge with Asp132 in hexane. A nonpolar sidechain Leu56 (6/4) adopts a more solvent exposed alternate conformation in hexane. Other nonpolar residues (Met1 (16/10), Phe28 (2/0), Val45 (13/11), Ala66 (61/35), Tyr157

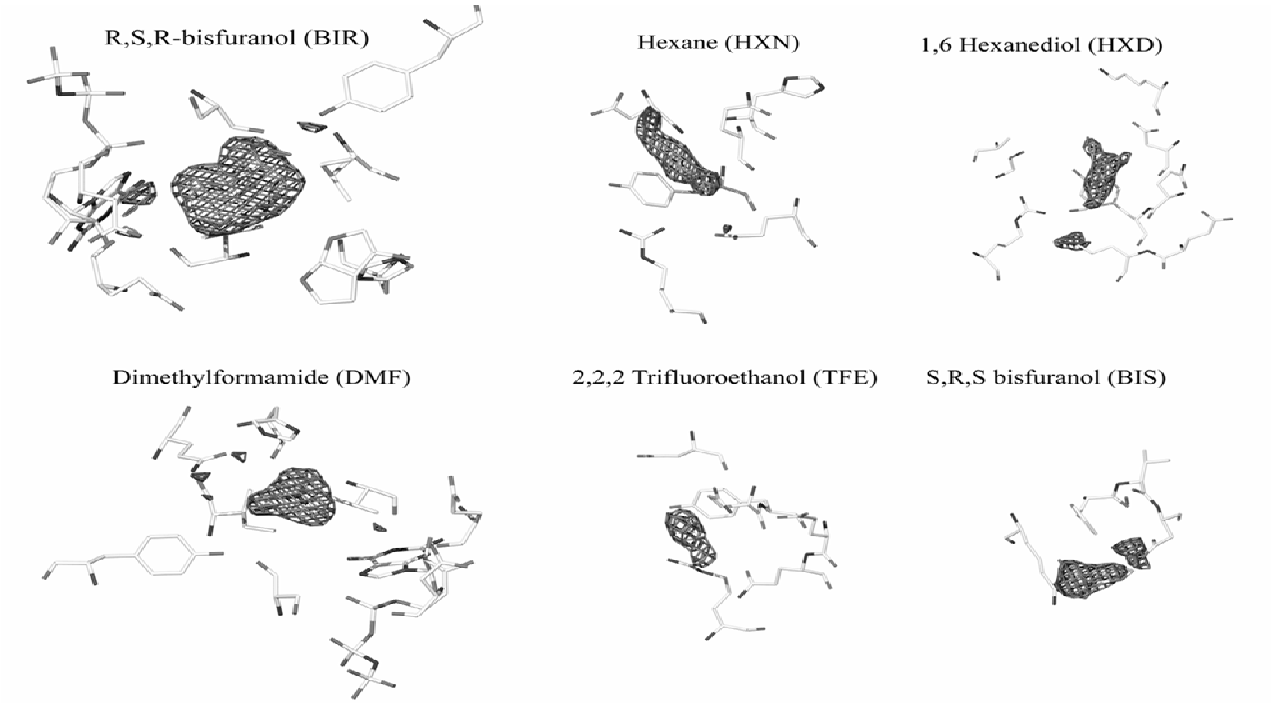


Figure 1. 2Fo-Fc electron density contoured at 1.0 Sigma for selected solvent molecules. Protein residues within 6 Angstrom of the solvent molecule are also shown. This Figure was generated with Pov-Ray.

2/0), Ala18 (11/0), Ala122 (58/54) and Ala146 (3/0)) adopt more solvent exposed conformations in hexane vs. water, although the changes in solvent exposure are due to small changes in structure. The only hydrophobic residues that adopt a less solvent exposed conformation in hexane vs. water are Ala70 and Ala121.

Solvent Binding Sites

Solvent molecules are bound at multiple sites on the surface of Ras. However, clusters of solvent molecules are found in four distinct regions of the Ras protein, the extended Ras effector domain (Ile21,Val29,Tyr40), Switch II (Met67), AH1(Tyr96) and AH2 (Leu133). All of the organic solvent molecules found on the surface of Ras are superimposed on the Ras protein structure in Figure 2. The electrostatic surface of Ras, with the solvent binding sites labeled with asterisks and the important functional regions labeled by name, is shown for comparison. Many solvent molecules, including BIR, PEN, DMF and GLY cluster loosely around the polar Ras effector binding domain, which is located around Switch I and between L2 and L6 in Figure 2C. This region is known to be involved in protein: protein interactions with several protein binding partners. Tight clusters of solvent molecules were found in two additional regions, referred to as affinity hotspot 1 (AH1) and affinity hotspot 2 (AH2), labeled with asterisks in Figure 2B. HXD, HXN, TFE and DMF cluster at AH1, located between Switch II and helix 3. BIR, GLY, HXN, TFE and DMF all cluster at AH2, located between helix 3 and helix 4. Solvent molecules which do not cluster are mainly found at crystal contacts, however symmetry related crystal contacts do play a role in several of the multiple solvent binding sites. The electrostatic solvent accessible surface maps show that the clustering of organic solvents occurs in surface pockets, grooves and channels. Hydrophobic residues interact with organic solvents in some binding sites.

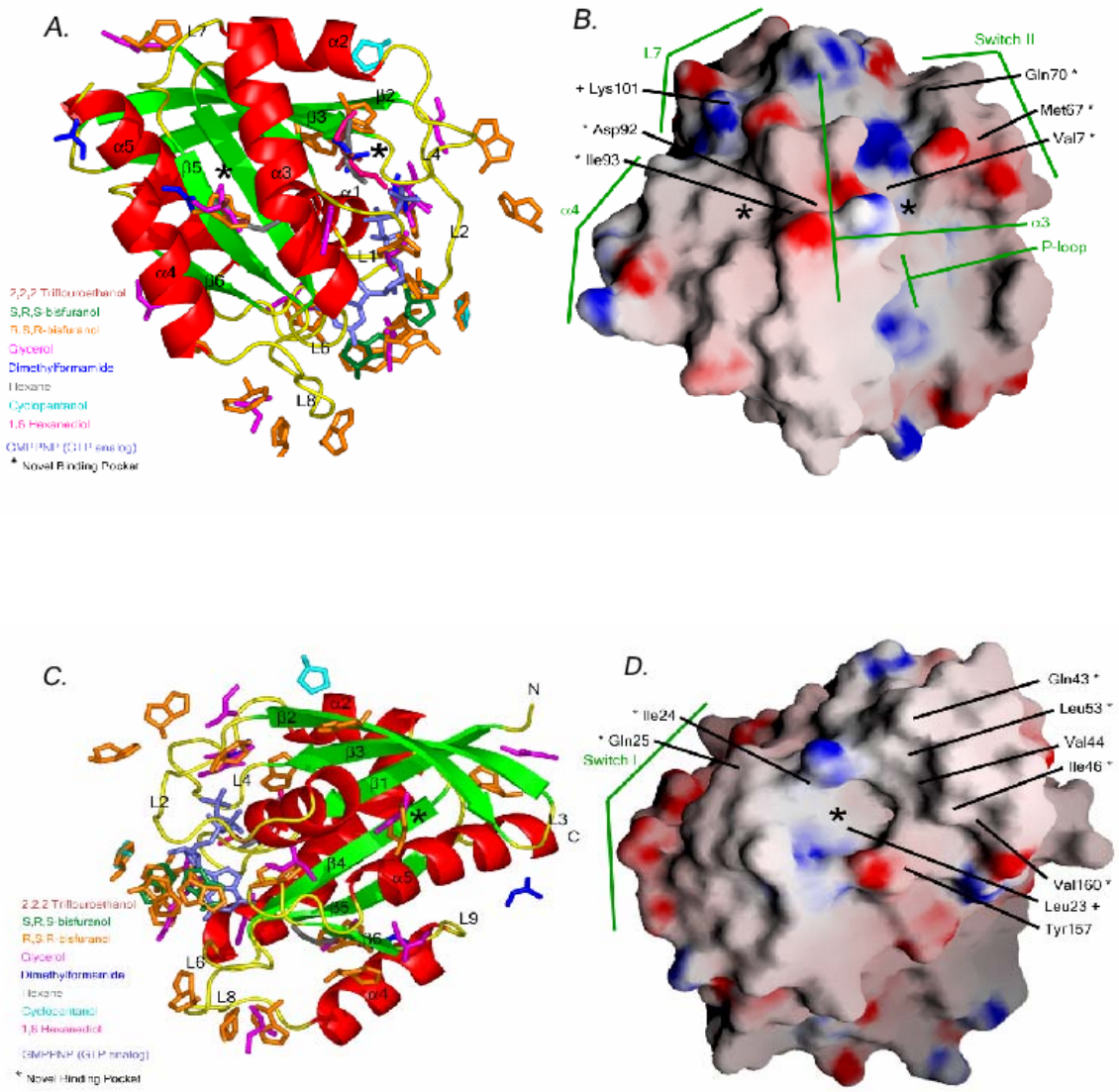


Figure 2. Organic solvent molecules superimposed on the structure of H-Ras/GMPPNP with the electrostatic surface of H-Ras shown for side by side comparisons, shown in two rotations. The electrostatic surface representation is the solvent accessible surface generated with GRASP (Nicholls A. et. al., 1991)

Polar interactions and hydrogen bonding between protein and organic solvent atoms occur at several solvent binding sites, especially within the polar Switch I region. In the Switch I solvent binding sites, the side chains adopt a similar conformation in all of the MSCS structures. On the other hand, solvent binding sites AH1 and AH2 involve inflexible hydrophobic residues and flexible polar residues that adopt multiple conformations depending on the solvent.

Affinity Hotspot One, SwitchII/Helix3

The non catalytic conformation of Switch II in the R32 form of Ras creates a binding pocket that is accessible by multiple organic solvent molecules and can be utilized in protein: protein interactions. Importin β binds to Ran-GTP at this site, utilizing a conserved Lys68 to make a salt bridge with Asp107 (Gln99 in Ras) (Vetter, Arndt, et al. 1999). To the best of our knowledge, the Ran-GTP-Importin β complex is the only structurally characterized complex with the Switch II conformation seen in the R32 form of Ras and is the only structurally characterized complex to utilize AH1 as a site of protein: protein interaction. Tyr96 forms the hydrophobic base of AH1 between Switch II and helix 3. In Figure 3, organic solvent molecules other than R,S,R - bisfuranol and glycerol are superimposed on the Ras structure. The AH1 hotspot, located between Switch II (colored magenta) and helix 3 (labeled) has solvent molecules DMF 1510, HXD 1870, HXN 1620 and TFE 1252 superimposed at this site. No symmetry related residues are within 10 Å of the solvent binding site, so crystal packing does not affect this affinity hotspot. This site is surrounded by polar residues that are either flexible, disordered or adopt multiple conformations in the solvent soaked structures. Arg68, Glu62, Lys88 and Gln99 are partially disordered in several MSCS structures. Arg102, Gln95 and His94 adopt alternate conformations, depending on the

solvent used. In Figure 4, a superposition of several solvent soaked structures shows that side chains Gln95 and His94 adopt alternate conformations depending upon the solvent. Interestingly, Gln95 and His94 adopt distinct conformations in the isopropanol, glycerol and the crosslinked control structure, solved in aqueous solution, none of which have solvents bound at AH1. In the solvent soaked structures which do have a solvent molecule bound at AH1, Gln95 is displaced, which in turn displaces His94, indicating a potential mechanism for crosstalk across both sides of helix 3.

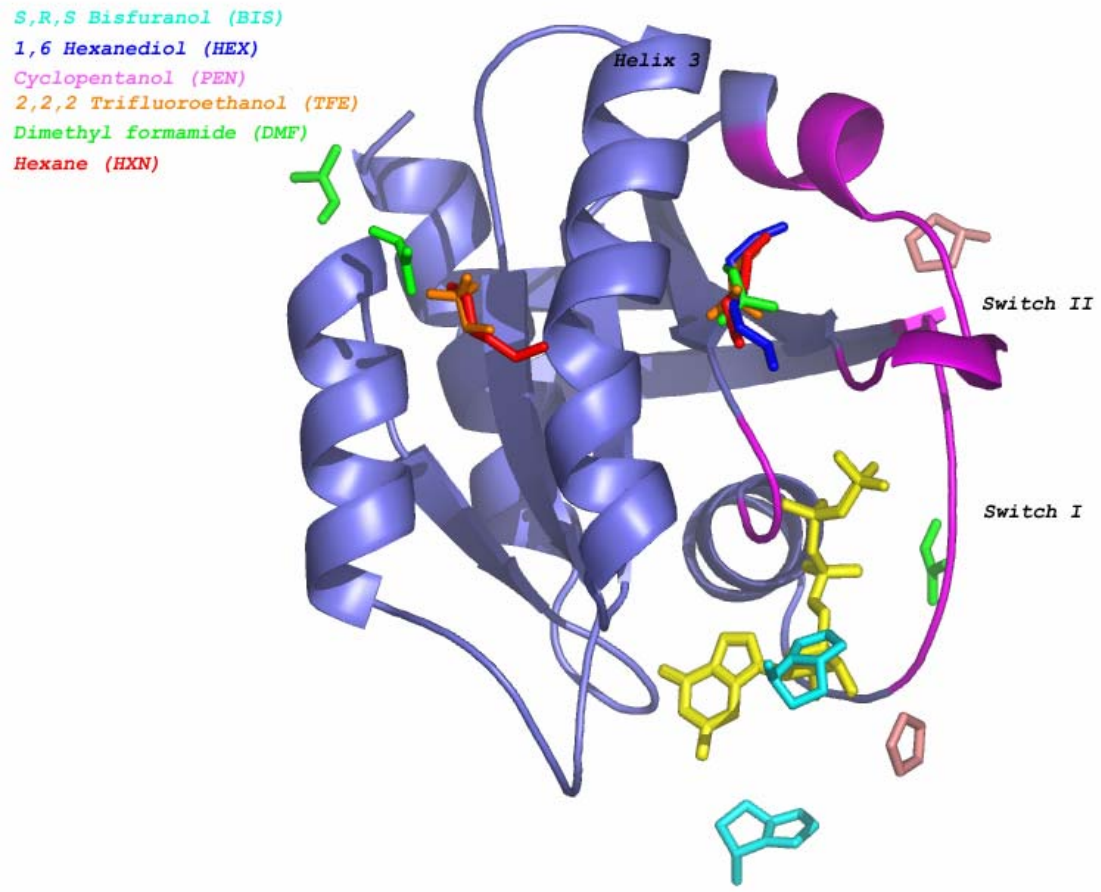


Figure 3. Organic solvent molecules superimpose at the affinity hotspot between Switch II and helix 3. The affinity hotspot between Switch II and helix 3 is bound by hexanediol, TFE, DMF and hexane, colored blue, orange, green and red.

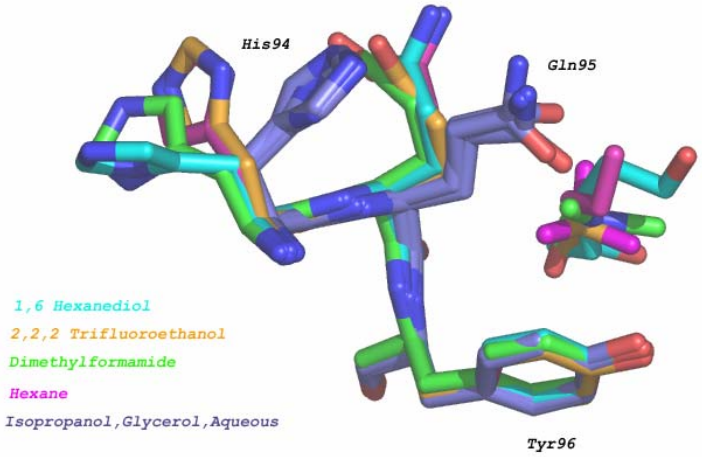


Figure 4. Conformation flexibility induced by organic solvent molecule binding at the Switch II / helix 3 affinity hotspot. Gln95 and His94 from the isopropanol, glycerol and aqueous structures, which do not have a solvent molecule bound in the affinity hotspot, adopt a similar conformation, shown in blue. Gln95 and His94 from the hexane, DMF, TFE and hexanediol structures, colored magenta, green, orange and cyan, adopt alternate conformations in response to bound solvent molecules.

In Ran-GTP, His94 is replaced by Pro102, which kinks the middle of helix3, making this binding site accessible to Importin β. This region is also important as a specificity determinant between Ras, Ral and Rap GTPases (Nicely, Kosak, et al. 2004). In our structure, the hydroxyl group of Tyr96 makes a hydrogen bond with a water molecule that, in turn, forms a hydrogen bond with Arg68, a critical residue involved in stabilizing Switch II. Tyr96, which forms the base of this binding site, adopts a flat buried structure in all three GTPases except for Ral-GMPPNP A. The crystal structure of Ral-GMPPNP has two monomers (A and B) in the asymmetric unit. In Ral-GMPPNP B, a Switch II which is ordered by crystal contacts helps to stabilize the flat Tyr96 (Phe107 in Ral). Ral-GMPPNP A has a disordered Switch II, in which Tyr96 is flipped approximately ninety degrees relative to the other structures. This binding site is sensitive to the conformation adopted by Switch II, which undergoes an order to disorder transition upon GTP hydrolysis. Therefore, proteins which utilize this binding site are likely to do so in a GTP dependent manner, which is a distinguishing characteristic of true Ras effector proteins (Campbell, Khosravi-Far, et al. 1998).

Affinity Hotspot Two

AH2 is located on the opposite side of His94 between helix 3 and helix 4 and is bound by TFE1251, HXN1621, DMF1513, BIR 1216 and GLY 1999. Tyr137 and Leu133 make favorable hydrophobic interactions with solvent molecules in this binding site. A symmetry related Leu120 also makes favorable hydrophobic interactions with the organic solvent molecules in this site. Because crystal contacts form part of the solvent molecule binding site, this hotspot is somewhat artificial and may be an artifact of the experimental technique. However, it is interesting to note that virtually all of the differences in the

sequences of RalA and RalB map to this pocket, supporting the role of this region as a binding site for protein: protein interactions. In addition, this region is of interest because many of the residues in this large hydrophobic pocket, which includes; Ile93, Phe90, Leu113 and Phe82, are important tree determinant residues that distinguish the Ras, Rap and Ral GTPases (Nicely, Kosak, et al. 2004).

Switch I Solvent Binding Site

Only five of the nineteen BIR molecules bind outside of the effector region. Specific interactions between BIR molecules and the extended Ras effector domain are shown in Figure 6. Three BIR molecules (1203, 1204 and 1206) make specific hydrogen bonds with Switch I residues Asp30, Tyr32 and Thr35 respectively. BIR 1205, not shown, is sandwiched between Switch I and the nucleotide. BIR 1203 is found within a hydrophobic pocket formed by Phe28. The hydroxyl of BIR 1203 makes a hydrogen bond with Asp30. Lys147 (not shown) provides a partial positive charge for favorable electrostatic interactions with O1 and O2 of BIR 1203. BIR 1204 and PEN 2000 bind near Tyr32 (4.2 Å from the side chain). Tyr32 is important for binding to most Ras Switch I binding proteins and is significantly buried in the Ras-SOS and Ras-GAP protein complexes (Hall, Yang, et al. 2001). BIR 1206 makes a favorable hydrophobic interaction with Ile36 and the hydroxyl group of 1206 is hydrogen bonded to the backbone carbonyl of Pro34. BIR1206 is 4 Å from the position of Tyr64 in other structures, while BIR 1226 (not shown) superimposes with Tyr64. The entire Switch II region is disordered in the BIR soaked structure, probably due to steric hindrance between the preferred Tyr64 conformation and the BIR 1206 and 1226 molecules.

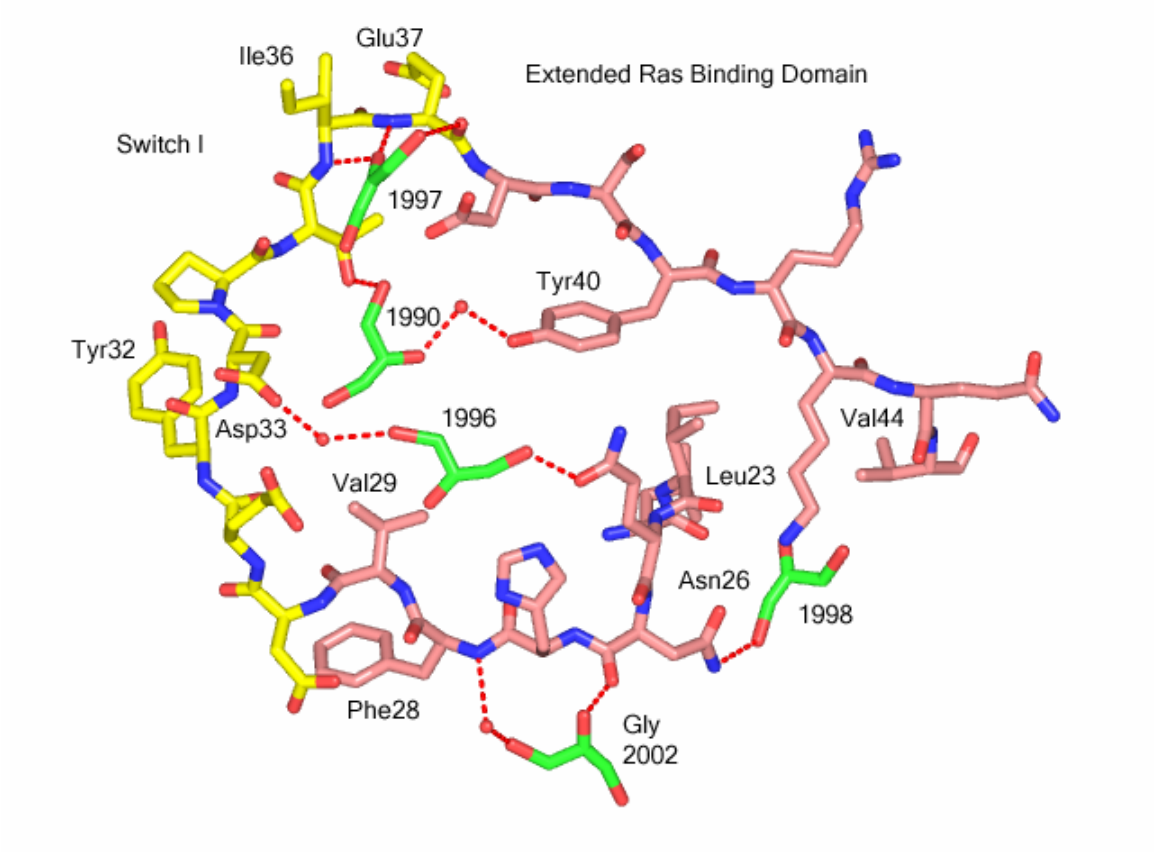


Figure 5. Switch I and Extended Ras protein binding domain with interacting glycerol molecules. Hydrogen bonds are drawn in red. Note the hydrogen bonds between glycerol molecules and protein atoms mediated by water molecules. All least squares superpositions of structures shown in this and other Figures were calculated using LSQMAN (Kleywegt and Jones, 1994). All Figures in this chapter, with the exception of Figure 1, 2C and 2D were generated using the PyMOL Molecular Graphics System (DeLano Sientific, San Carlos, CA).

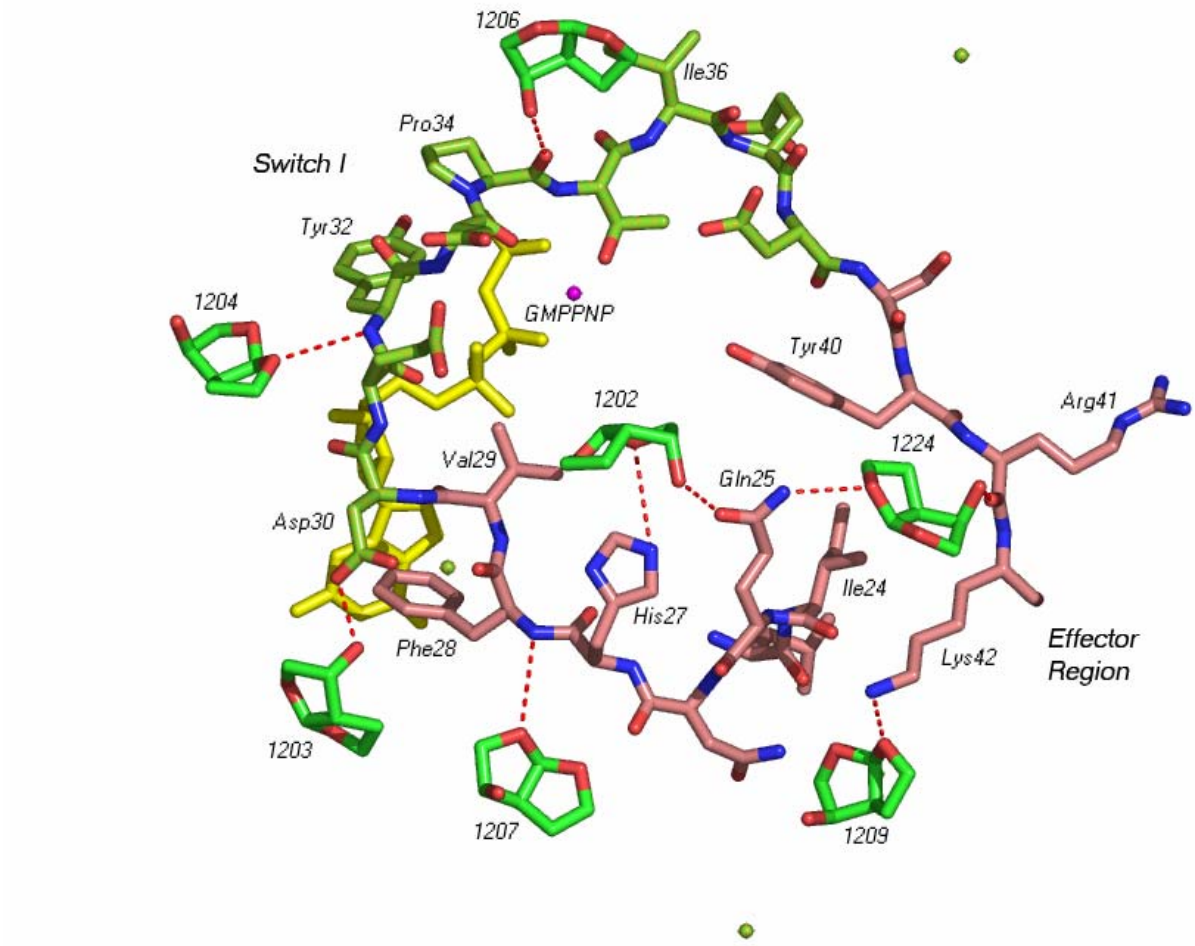


Figure 6. Selected R,S,R-bisfuranol molecules that make specific hydrogen bond interactions within the extended Ras effector binding domain. Switch I residues 30-38 (light green) and effector region residues 23-42 (pink) and bisfuranol (green) are shown in stick, with dashed lines indicating hydrogen bonds. The GTP analog GMPPNP is shown in yellow.

Bisfuranol molecules which interact with the extended Ras effector domain, colored pink in Figure 6, include 1202, 1207, 1209 and 1224. BIR 1202 is located on the opposite face of Switch I, positioned for polar interaction with His27 and a hydrogen bond with Gln25 and making favorable hydrophobic interactions with Val29. BIR 1202 overlaps with DMF 1511 and GLY 1996, which are all found within the hydrophobic pocket formed by Ile21 (not shown) and Val29. BIR1202 is near a two-fold symmetry axis, such that 1202 interacts with symmetry related residues Tyr40 (3.2 Å), Ile21 (4 Å) and Val29 (4 Å) as well as with a symmetry related molecule of 1202. BIR1207 forms a hydrogen bond with the backbone amide of Phe28 and GLY 2002 is also found at this position. GLY1998 and BIR1209 superimpose in a binding pocket formed by Leu23, see Figure 6. Leu23 has been identified as a switch of function residue by mutagenesis studies. Mutation of Leu23 to Phe23 is sufficient to change the phenotype of cells from polar to eyelash morphology, where a phenylalanine in Ral and Rap alters the binding pocket important for the polar phenotype, (reviewed in Nicely, Kosak, et al. 2004).

R,S,R-Bisfuranol and Glycerol

The solvent soaking experiment with R,S,R-bisfuranol (BIR) and glycerol (GLY) produced high resolution data sets that allowed identification of more organic solvent molecules than other, less polar, solvents. Ras crystals tolerated high concentrations of R,S,R-bisfuranol (90%) very well, diffracting to 1.5 Å, see Table 2.2. BIR molecules displace a number of water molecules. Only 46 water molecules are present in the final x-ray model, most of which are buried. Nineteen solvent molecules were located in the electron density maps, whereas only a few solvents (1-4) were located in the electron density maps from other solvent soaks. Unlike organic solvents which order the dynamic Switch II region,

BIR displaces Switch II residues important in ordering this region (Tyr64 and Arg68), resulting in a disordered Switch II. Many BIR molecules bind in a large, electropositive groove found near Switch I. This is not suprising, given the electronegative character of the BIR solvent molecules. Table 2 ranks the organic solvent molecules based on B-factors and fits to electron density and lists buried protein residues. Only well ordered BIR molecules and BIR molecules that are found at multiple solvent binding sites are discussed further.

Glycerol (70%) is another polar solvent for which a large number of binding sites (12) were determined. Seven of these binding sites overlap directly with BIR binding sites (see Table 2). Only three glycerol molecules (1991, 1992 and 1995) are found at unique binding sites. Like the BIR molecules, GLY molecules are more ordered at some binding sites than others. The well ordered GLY molecules, with a density correlation coefficient greater than 0.80, include GLY 1990, 1991, 1996 and 1997. GLY 1990, 1996 and 1997 are found in the Switch I binding pocket, see Figure 5. These glycerol molecules make multiple hydrogen bonds with protein backbone and side chain atoms. Several glycerol molecules are hydrogen bonded to water molecules, which in turn are hydrogen bonded to the protein. Unlike the bisfuranol molecules, which displace most of the surrounding water molecules, glycerol molecules, with three hydroxyl groups available to act as hydrogen bond donors and acceptors, substitute for water molecules and maintain the water mediated hydrogen bonding network surrounding the protein. Unlike bisfuranol, glycerol has a stabilizing effect on Switch II, as seen from electron density fits in that region. The entire main chain was built into electron density, with side chain residues 62, 63, 67, 69, 70 and 73 disordered. The two residues that play a large role in stabilizing this region (Tyr64 and Arg68) are well-ordered in the glycerol soaked structure.

Table 1. Solvent evaluation data.

		Electron Density Fit (2Fo-Fc)	
Solvent Molecule	B factor	Corr.Coef.	nearest residues
BIR 1202	**48**	**92**	**Ile21 * # (4)**
BIR 1203	**48**	**88**	**Phe28**
BIR 1206	**52**	**85**	**Ile36 * #**
BIR 1224	**50**	**84**	**Ile24 (3)**
BIR 1222	**56**	**82**	**Asn86**
BIR 1214	**60**	**83**	**Tyr96 (1)**
BIR 1219	**58**	**82**	**Arg123**
BIR 1211	**61**	**81**	**Pro34 ***
BIR 1217	**61**	**81**	**Gln95 (1)**
BIR 1221	**59**	**81**	**Cys118**
BIR 1226	**63**	**80**	**Ile36**
BIR 1204	**61**	**79**	**Tyr32 * #**
BIR 1209	**56**	**78**	**Leu23 (3)**
BIR 1207	**59**	**76**	**Gln22**

Table 1. (continued)

BIR 1220	**60**	**75**	**Ala121**
BIR 1212	**59**	**75**	**Lys117**
BIR 1205	**57**	**75**	**GNP**
BIR 1210	**64**	**73**	**Leu120**
BIR 1213	**63**	**71**	**Asn86**
GLY1990	**43**	**88**	**Ile21 *#,Val29 *(4)**
GLY1991	**63**	**85**	**Asp154**
GLY1992	**65**	**76**	**Glu76, Glu3**
GLY1993	**58**	**71**	**Lys88,Asn86, Gln61**
GLY1994	**63**	**80**	**Arg123**
GLY1995	**58**	**75**	**His94**
GLY1996	**20**	**86**	**His27,Ile21* # (4)**
GLY1997	**53**	**80**	**Asp38*#, Ile36*#**
GLY1998	**58**	**67**	**Leu23**
GLY1999	**56**	**70**	**Leu133,Tyr137 (2)**
GLY2002	**55**	**71**	**Gln22, Phe28**
GLY 2003	**45**	**80**	**Leu120, Lys117**
HXN1620	**33**	**77**	**Tyr96 (1)**
HXN1621	**25**	**80**	**Tyr137 (2)**

Table 1. (continued)

BIS 1028	**40**	**78**	**Phe28**
BIS 1032	**41**	**78**	**Tyr32,Lys117,Asn85**
DMF 1510	**45**	**83**	**Tyr96 (1)**
DMF 1511	**21**	**94**	**Ile21*#,Val29* (4)**
DMF 1512 ^	**30**	**75**	**Gln165,Glu162**
DMF 1513	**41**	**77**	**Tyr137 (2)**
PEN 2001	**36**	**73**	**Tyr71**
PEN 2000	**38**	**79**	**Tyr32*#**
HXD 1870	**38**	**72**	**Tyr96 (1)**
HXD 1871 ^	**25**	**83**	**Ile142**
TFE1251	**44**	**75**	**Tyr137 (2)**
TFE1252	**47**	**61**	**Tyr96 (1)**

*Residues that are buried > 10 % in known Ras : protein complexes

Residues with measured effects on binding affinity

^ Solvent molecules on a symmetry axis with occupancy < 1.0.

Solvent molecules in P3(2)21 vs. R32 Space Group

A limited MSCS analysis was done at room temperature, with crystals of Ras-GMPPNP that formed with the symmetry of space group P3(2)21, see Chapter 3. Organic solvent soaks at room temperature in the P3(2)21 form include: 2,2,2-trifluoroethanol, isopropanol and hexanediol. This limited study did not identify any affinity hotspots and only identified one molecule of TFE, one molecule of hexanediol and three molecules of isopropanol. The polar residues bisfuranol, glycerol and dimethylformamide that map the hydrophilic Ras protein binding site were not included in the initial room temperature studies. Isopropanol was used as an organic solvent in MSCS studies of Ras-GMPPNP performed at 100K with crystals that formed with the symmetry of the R32 space group. Unfortunately, the isopropanol solvent was not amenable to cryogenic conditions, resulting in numerous ice rings that compromised the high resolution data and no solvent molecules were identified in this data set. Two molecules of hexanediol and TFE were identified in the R32 form that were not identified in the P3(2)21 form. The higher resolution of the R32 data sets may have allowed these additional solvent binding sites to be resolved, see Table 2.2. Crystal contacts present in the R32 form that are not present in the P3(2)21 form may account for differences in one of the solvent binding sites. A common solvent binding site was found between Switch II and helix3. Switch II adopts a unique conformation in R32 form of Ras crystals, forming a pocket between Switch II and helix 3, which is larger and deeper than the pocket found in the P3(2)21 form of Ras. This may account for differences in the way particular types of solvent molecules interact within this pocket in the two crystal forms.

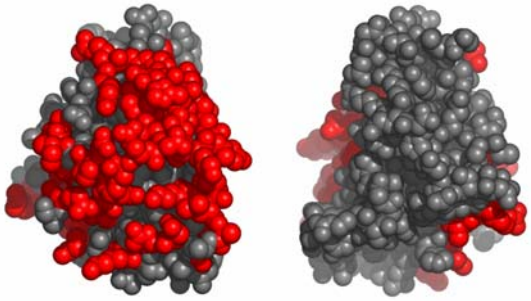
DISCUSSION

Trends Based on Solvent Polarity

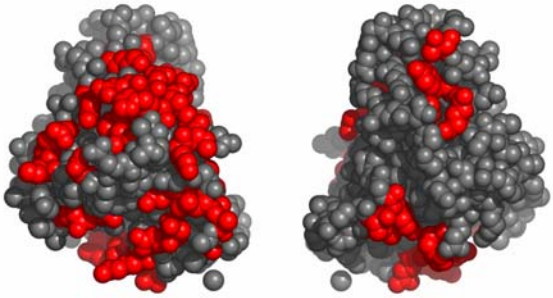
As has already been noted, the known Ras effector binding region (residues 20-48) is hydrophilic and charge interactions dominate protein binding. Although a few hydrophobic residues do contribute to the binding affinity of some Ras protein interfaces, the largest contributions to binding affinity come from polar residues (Kiel, Serrano, et al. 2004). Although a number of hydrophobic residues are on the surface of Ras protein, only a few are found in any of the putative protein binding pockets. Therefore, we used a panel of organic solvents with a range of polarity as measured by dielectric constant and total polar surface area (TPSA). The relative order of organic solvents, from most polar to least polar, is BIR > GLY > DMF >TFE > ISO > PEN > HXN. As expected, polar solvent molecules bound to more of the hydrophilic protein surface than non-polar solvent molecules. In Figure 7, we show the surface residues of Ras which are buried by different organic solvents. The protein-organic solvent interface was calculated for each solvent soaked structure in CNS. The B-factor column of each PDB file is initially set at 0.0. Protein atoms which are buried in the interface are then given a B-factor of 1.0. The total buried surface area is given in $Å^2$. Atoms which are buried by the organic solvent are colored red and atoms which are not buried by

Figure 7. Surface residues of Ras-GMPPNP, buried by organic solvents. The Ras protein is shown in a van der waals surface representation, with atoms buried by the organic solvent colored red and atoms not buried by the organic solvent colored gray. The buried interface was calculated in CNS. Molecules in the second column were rotated 180 degrees on the y-axis, relative to the first column.

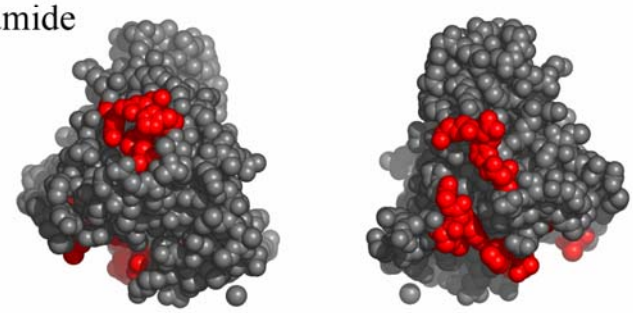
Bisfuranol



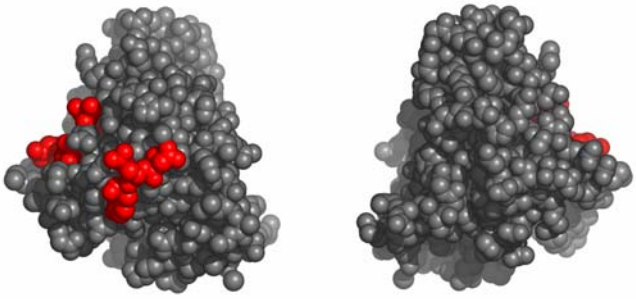
Glycerol



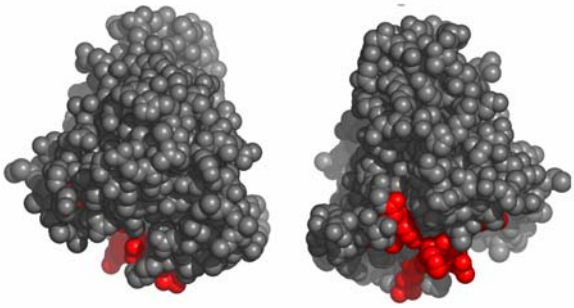
Dimethylformamide



Cyclopentanol



Hexane

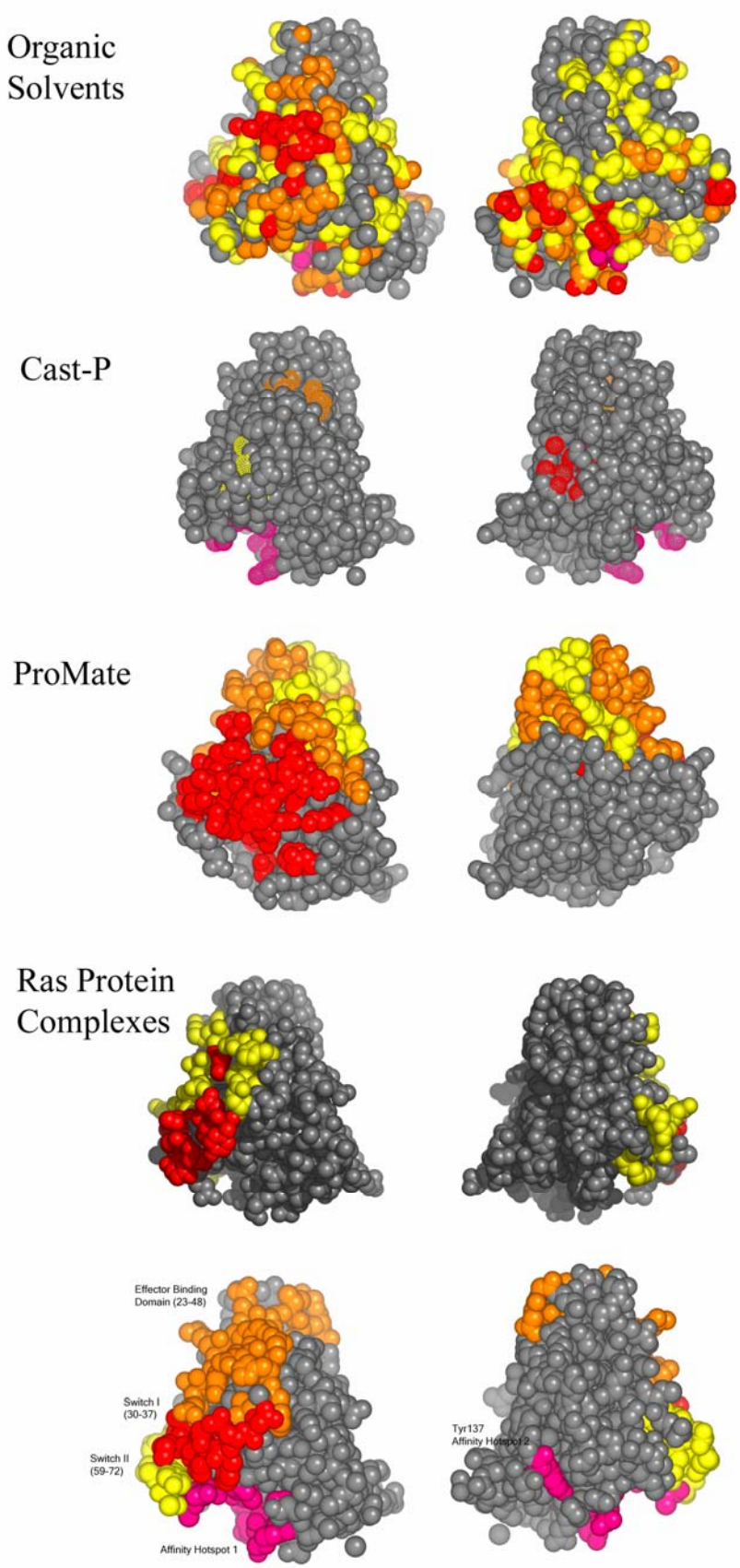


the organic solvent are colored gray in Figure 7. The total surface area of buried interface for each solvent is: 3982 $Å^2$ BIR, 2680 $Å^2$ GLY, 980 $Å^2$ DMF, 438 $Å^2$ PEN and 583 $Å^2$ HXN . Solvents are shown in order of decreasing polarity, as calculated by Total Polar Surface Area measurement and /or solvent dielectric constants. Solvents with higher polarity bury more protein surface than solvents with lower polarity. The bisfuranol solvent buries a large amount of Ras protein surface, effectively covering almost the entire Ras effector binding region (residues 20-48). Other polar solvents, like glycerol and dimethylformamide, are found in hydrophilic pockets within the effector binding surface of Ras. By contrast, hexane, a non-polar organic solvent, does not interact with the known Ras effector binding region. Hexane is only found in two hydrophobic pockets on the opposite face of the protein. One of the hydrophobic pockets, AH2, is found between helix 3 and helix 4 of Ras and is located in a region of crystal contacts with a symmetry related Ras protein. The other hydrophobic pocket, AH1, which is found between Switch II and helix 3, is not on a crystal contact and is also a solvent binding site for trifluoroethanol, dimethylformamide and hexanediol.

Comparing MSCS with Known Ras Protein Complexes

The surface of Ras that is buried by each organic solvent molecule is shown in Figure 7. In Figure 8, this information is combined in a composite van der waals representation, where atoms are colored according to the number of solvent molecules that bury each protein atom. In this composite representation, much of the Ras protein surface is buried by at least one organic solvent molecule. However, atoms that are buried by three or more organic solvents, colored red or pink, are defined as affinity hotspots and limited to four main regions; the extended Ras effector domain (Ile21,Val29,Tyr40), Switch II (Met67), AH1(Tyr96) and AH2 (Leu133). With the exception of AH1 and AH2, these are known

Figure 8. Comparison of protein binding site prediction methods. The surface buried by one organic solvent is colored yellow, two solvents colored orange, three solvents colored red and the surface buried by four or more solvents is colored hotpink. The CastP surface is colored hot pink for atoms from the mouth of the largest pocket, red for atoms from the second largest pocket, orange for the third largest and yellow for the fourth pocket identified by CastP. The ProMate surface is colored red for residues very highly predicted to participate in protein: protein interactions (b > 60), orange for highly predicted (b>40) and yellow for predicted (b > 20) to participate in protein: protein interactions. The Ras protein complex surface is colored yellow for residues buried by Ras-GAP, Ras-SOS or Raf kinase. Residues are colored red if they participate in multiple protein binding interfaces.


Organic Solvents
Cast-P
ProMate
Ras Protein Complexes
Effector Binding Domain (23-48)
Switch I (30-37)
Switch II (59-72)
Affinity Hotspot 1
Tyr137 Affinity Hotspot 2

regions of protein: protein interaction. Affinity hotspots are not located in separate, distinct regions, but are surrounded by regions of lower solvent affinity, colored yellow or orange. In general, this pattern is similar to protein binding affinity maps, where only a few residues within a larger protein binding interface make large contributions to protein binding affinity, with the surrounding residues making smaller contributions. Kiel and Herrmann have calculated protein binding affinity maps for Ras bound to Raf kinase (Kiel, Serrano, et al. 2004). While the protein binding affinity maps and the organic solvent binding maps are not directly correlated or completely overlapping, they are similar. Most of the residues Kiel and Hermann identify that make large contributions to protein binding affinity, are also identified by MSCS, see Table 1. The residues surrounding the Switch II – helix 3 solvent binding site were not tested by Kiel and Hermann and do not participate in binding the minimum binding domain of Raf kinase. However, Switch II residues do participate in a number of protein interactions with other Ras effector proteins. Specifically, Met67, which forms part of the Switch II affinity hotspot, contributes to the binding affinity of the Ras-RasGAP interaction (Scheffzek, Ahmadian, et al. 1997).

A well ordered molecule of cyclopentanol (PEN2001) interacts with the solvent exposed Tyr71. This solvent exposed conformation is also stabilized by Importin-β and by Ras-SOS, a guanine nucleotide exchange factor for Ras. In Chapter 4, we discussed how Tyr71, which oscillates between at least 2 stable conformations in solution studies of Ras-GTP, could play a critical role in determining the conformations of Arg68 and Switch II. Stabilizing a solvent exposed conformation of Tyr71 could help stabilize the non-catalytic conformation of Switch II, which forms an organic solvent affinity hotspot in our studies. The affinity hotspot we detect between Switch II and helix 3 is known to be a site of protein:

protein interaction in the Ran-GTP-Importin β complex (Vetter, Arndt, et al. 1999). Based on interface analysis, which shows that multiple proteins utilize most Ras binding modes, Corbett and Alber suggest that this unique binding mode is likely to be more common and that further structural studies of Ras superfamily complexes could discover more GTPase binding proteins that utilize this binding mode (Corbett & Alber 2001). Identification of this affinity hotspot by MSCS provides further evidence suggesting that this region could play a role in protein: protein interactions with Ras effector proteins.

Comparison of MSCS with ProMate and CastP

A number of computational methods have been developed to predict protein binding sites. In this section we compare the results of our experimental method to predict protein binding sites with the computational methods utilized by ProMate and CastP. ProMate uses a prediction algorithm that accounts for the interface properties and tendencies in known transient protein: protein complexes. ProMate ranks protein interface sites based on a number of criteria, including the presence and size of hydrophobic patches, secondary structure, amino acid type and whether or not a multiple alignment shows evolutionarily conserved positions (Neuvirth, Raz R., et al. 2004). Results of a ProMate analysis of Ras are shown in Figure 8. ProMate and MSCS correctly identify the known Ras effector binding region as very likely to participate in protein interactions. ProMate does not predict the affinity hotspot between helix 3 and helix 4, located by MSCS on the opposite face of the protein, but ProMate does predict that Tyr96, located at the base of the Switch II / helix 3 binding site, is very likely to be involved in protein: protein interactions. CastP predicts protein: ligand binding sites on the basis of binding pocket geometry (Binkowski, Naghibzadeh, et al. 2003). A pocket is defined as an empty, concave region of the protein

surface which can be accessed by solvent with a probe radius of 1.4 Å. Pockets are distinguished from shallow grooves on the protein surface, in that at least one cross section of the pocket is larger than the mouth opening, thus providing some protection from bulk solvent. Because of this protection, small molecules or protein side chains that can access these pockets do not have to compete directly with bulk solvent. The four largest pockets identified by CastP, colored pink, red , orange and yellow in Figure 8, are located at AH1, the N terminus of helix 4 and in the extended Ras effector binding region. With the exception of the helix 4 pocket, these pockets are all identified by ProMate and MSCS as likely to be involved in protein: protein interactions. The size and participating residues of the four CastP pockets are listed in Table 2. The largest ligand binding pocket corresponds to AH1, identified by MSCS. CastP nominally calculates pocket geometry from a Connolly surface, constructed by rolling a solvent molecule with a sphere of radius 1.4 Å over the protein. However, organic solvent molecules are larger than water molecules, so the Connolly surface does not necessarily represent the protein surface that is 'seen' by organic solvent molecules. CastP recognizes the AH1 site as a 'pocket' using a solvent molecule with a radius as large as 3.0 Å. Therefore, surface shape, rather than specific hydrogen bond formation could be the primary factor determining organic solvent molecule binding at this site.

One advantage of MSCS over computational methods is the ability to provide detailed chemical information about predicted protein binding sites. This information can be used not only to categorize protein binding pockets, but may be used to design specific small molecule inhibitors of specific protein: protein interactions. The conformational flexibility of the Switch II – helix 3 binding pocket that was induced by organic solvent molecule binding

Table 2. CastP calculations with crosslinked control protein with a solvent molecule probe with a 1.4 Å radius and a 3.0 Å radius. Water molecules were removed for the calculation, but the nucleotide was considered part of the protein surface.

Mouth Geometry

ID	*N_mth*	*Area_sa*	*Area_ms*
11	1	4.567	23.35
12	1	2.915	19.29
13	1	16.260	54.35
14	1	44.791	107.37
*1	2	1.267	70.22

Pocket geometry

ID	N_mth	Area_sa	Area_ms	Vol_sa	Vol_ms
11	1	22.102	125.84	5.714	81.40
12	1	20.399	101.56	4.808	78.91
13	1	45.802	106.30	25.875	128.17
14	1	61.495	88.54	60.870	164.49
*1	2	13.421	84.90	2.349	122.60

Pocket Atoms, from largest pocket to smallest (1-4:14-11):

1. 14: Ala11,Gly12,Gly60,Gln61,Glu62,Lys88,Ser89,Asp92,Gln95,Tyr96
2. 13: Arg97,Asp107,Asp108,Val109,Pro110,Met111,Tyr137,Glu162
3. 12: Leu23,Ile24,Lys42,Leu53,Arg149,Glu153,Tyr157
4. 11: Ser17,Ala18,Ile21,Val29,Glu31,Asp33

* CastP calculations with probe radius = 3.0.

N_mth = Number of mouth openings

sa = Solvent Accessible surface

ms = Molecular Surface

Area measurements are in $Å^2$

Volume measurements are in $Å^3$

is a good example of specific chemical information obtained by MSCS that could not be obtained by computational methods alone. By soaking the protein crystal in organic solvents with low dielectric constants, we simulate the local conditions that are presumed to exist at the protein - protein interface. By comparing the protein structures of solvent soaked vs. aqueous soaked protein crystals, we get a glimpse of the range of conformations that proteins can access in the environment of a protein: protein or protein: membrane interface. In cases where the solvent molecule simulates a protein side chain or main chain, conformational changes may be due to specific chemical interactions. In cases where the organic solvent molecule does not directly interact with the protein, protein conformational changes are likely due to bulk effects, providing information about the range of conformations proteins can access during protein complex formation.

Conclusions

Our results indicate that MSCS can be used to predict protein binding sites of proteins involved in the formation of polar, transient heterocomplexes. This is an important extension of the MSCS technique that had previously only been applied to analyze protein enzyme inhibitor complexes. Proteins that form transient heterocomplexes make up the bulk of intracellular signaling pathways and analyzing the protein complexes they form is a major challenge. Many of these protein complexes are not stable enough to be analyzed by crystallography or other structural methods. Therefore, definition of the protein - protein interface by alternative experimental or computational methods is critical to understanding the function of these proteins. In our experiments with Ras, we have shown that polar organic solvents map the known Ras-Raf protein interface in detail, although the pattern of organic solvents bound at this site does not reflect the tight clusters seen previously for

proteases. This is primarily due to the highly charged nature of the Switch regions. We have also identified a region between Switch II and helix 3 as a putative protein binding interface. This affinity hotspot is not in a region of crystal contact and occurs in a large binding pocket that is created by a unique Switch II conformation which is found in the Ran-GTP-Importin β complex. Solvent molecules that bind at this site introduce interesting conformational changes involving helix 3 residues. These conformational changes could provide a means of communication between helix 3 and helix 4.

CHAPTER SIX

Two enantiomers bound to H-Ras and Hen Egg White Lysozyme Protein

ABSTRACT

It is well known that mirror image stereoisomers of small molecules can interact very differently with proteins, which are themselves chiral molecules. In this chapter we examine the interactions of (3R,3aS,6aR)-hexahydrofuro[2,3-b]furan-3-ol and (3S,3aR,6aS)-hexahydrofuro [2,3-b]furan-3-ol (referred to as R,S,R and S,R,S bisfuranol), with H-Ras-GMPPNP (Ras) and Hen egg white Lysozyme (HEWL) using the multiple solvent crystal structures method (MSCS). Protein crystals of HEWL were soaked in 50% R,S,R and 50% S,R,S bisfuranol. These enantiomers are bound at subsites A, D and F within the active site cleft of Lysozyme. Subsites A and F are stereoselective and bind only one enantiomer. Subsite D binds both enantiomers, but in different orientations. Protein crystals of Ras were soaked in 20% S,R,S – bisfuranol (BIS) and 20%, 50% and 90% R,S,R – bisfuranol (BIR). Two molecules of the S,R,S enantiomer are found at binding sites also occupied at low concentrations of R,S,R bisfuranol.

INTRODUCTION

Stereoisomers are molecules with identical atomic composition and bonding that differ in the arrangement of atoms around one or more chiral centers. Enantiomers are designated R or S depending upon the clockwise (R) or anti-clockwise (S) orientation of substituents around the chiral center, when drawn with the lowest priority substituent

(usually Hydrogen) pointing away from the viewer and going from highest to lowest priority for the remaining substituents (Gal 2002). Enantiomers of a compound have identical physicochemical properties, except for optical properties, but are often readily distinguished by proteins, that are themselves chiral molecules (Mason 1991). Chiral recognition, whereby a chiral protein/receptor interacts selectively with one enantiomer of a chiral compound, is common in biological processes and therefore is an important issue in drug development. Mirror image stereoisomers, or enantiomers, are of special interest in drug development. In fact, the Federal Drug Administration issued a policy statement for the development of stereoisomeric drugs which recommends that individual enantiomers of racemic mixtures should be characterized separately (Gal 2002).

Several crystallographic studies of proteins bound to enantiomers illustrate the structural basis of biological stereoselectivity. Examples include : Human Serum Albumin bound to R and S-warfarin and human aldose reductase bound to stereoisomers of Fidarestat (Petitpas, Bhattacharya, et al. 2001), (El Kabbani, Darmanin, et al. 2004). To examine the way in which proteins selectively recognize compounds that have subtle structural differences, we examined the binding of the enantiomers S,R,S bisfuranol and R,S,R bisfuranol to two proteins, H-Ras and hen egg white Lysozyme, using the Multiple Solvent Crystal Structures (MSCS) method.

R,S,R-bisfuranol (BIR) and S,R,S-bisfuranol (BIS) are enantiomers with three chiral centers (C3, C3a and C6a). All three chiral centers are sp^3 hybridized carbon atoms, with C3a and C6a connecting the two furan rings and C3 bonded to the hydroxyl substituent. Bisfuranol can be visualized as a cupped hand formed by the two furan rings with the hydroxyl group pointing out from the rings like a thumb. The two molecules are mirror

images of each other, with BIR being a cupped right hand and BIS the left hand, as shown in Figure 1. R,S,R-bisfuranol is a novel substituent in a potentially new class of HIV protease inhibitors (UIC-94017) and is attached to the body of the inhibitor molecule via the hydroxyl

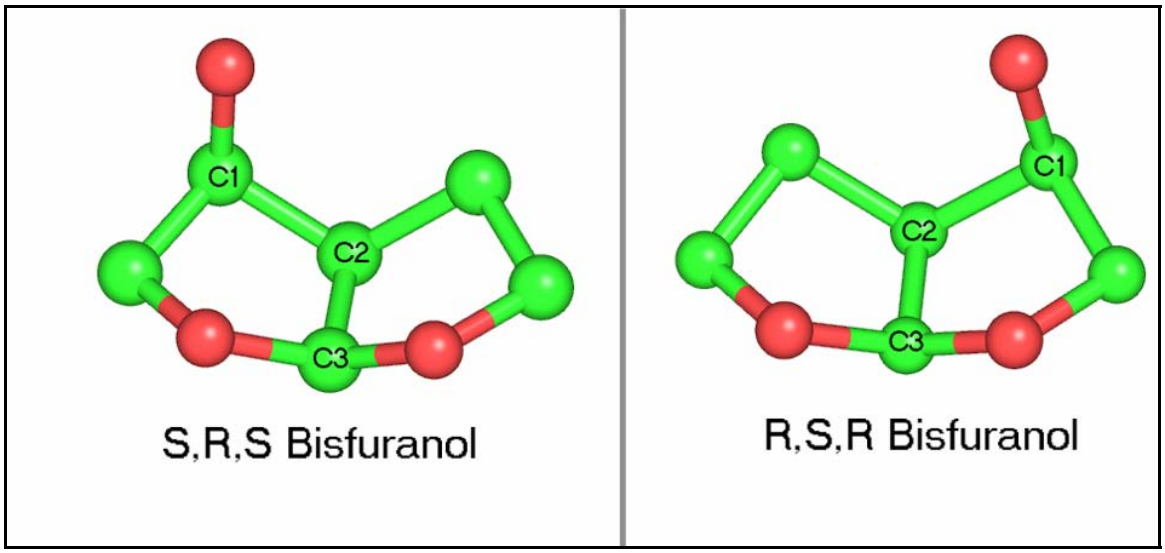


Figure 1. Enantiomers of bisfuranol. The chiral carbons labeled C1, C2 and C3 correspond to C3, C3a and C6a atoms respectively. Carbon atoms are colored green and oxygen atoms are colored red.

oxygen atom (Koh, Nakata, et al. 2003) . The bisfuran moiety increases the affinity of the inhibitor for HIV protease by making good hydrogen bonds with the backbone amide of Asp30 and the sidechain OD2 of Asp30 via the two furan ring oxygens. This is one example of stereospecificity within the active site of HIV protease. We find other examples of stereospecificity for the two enantiomers within the active site cleft of hen egg white Lysozyme (HEWL) and with the activated form of H-Ras.

Because of its central role in signal pathways that lead to cell growth and differentiation, H-Ras is an important anti-cancer drug target and is a good example of a protein that functions mainly via protein: protein interactions (Cox & Der 2002). In this chapter we discuss the results of soaking Ras protein crystals in 20% S,R,S bisfuranol and 20 % and 50 % R,S,R bisfuranol. We compare the results of the low concentration R,S,R-bisfuranol soaks with the results of high concentration (90%) R,S,R bisfuranol soaks discussed in Chapter 4.

Hen egg white Lysozyme (HEWL) was the first enzyme to have its structure determined by x-ray crystallography and is often utilized as a model system to address basic structural questions due to its availability and ease of crystallization (Blake, Koenig, et al. 1965). HEWL catalyzes the hydrolysis of oligosaccharides that make up the peptidoglycan cell wall of Gram-positive bacteria. HEWL has a large substrate binding cleft which can accommodate 6 total molecules of alternating 2-acetoamido-2-deoxy-glucopyranoside (NAG) and 2-acetoamido-2-deoxy-3-O-lactyl-glucopyranoside (NAM) in six subsites, commonly referred to as subsites A-F, in which the catalytic site is located between subsite C and D (Von Dreele 2001), (Pincus, Zimmerman, et al. 1977). HEWL is an attractive

candidate for MSCS because it has also been used to study the effects of a variety of solvents on protein structure, including; acetonitrile, glycerol, methyl-pentanediol (MPD), and dimethylsulfoxide (DMSO) (Wang, Zhu, et al. 1998), (Weiss, Palm, et al. 2000), (Mande & Sobhia 2000). In this chapter, we discuss the results obtained on soaking crosslinked HEWL crystals with 50% BIR and 50% BIS.

We included lysozyme in the enantiomer study because we wanted to determine if the enantiomer specific effects we saw with Ras were general to protein crystals. Ras crystals tolerate high concentrations of the R,S,R isomer, but show visible cracks at concentrations over 20% of the S,R,S isomer. Because protein crystals have a similar maximum concentration limit for many organic solvents we've tested, it seemed possible that the enantiomer specific concentration effects could be similar as well. However, tetragonal Lysozyme crystals tolerate both isomers equally well, with a maximum achievable concentration of 50% for each isomer. Apparently, the enantiomer specific effects seen with Ras are not general to protein crystals.

RESULTS

Ras soaked in S,R,S and R,S,R bisfuranol

Protein crystals of Ras-GMPPNP with the symmetry of the space group R32 were grown as previously discussed, see Experimental Procedures. Crosslinking, solvent soaking, data collection and refinement procedures are described in Experimental Procedures. The highest concentration of R,S,R – bisfuranol (BIR) that diffracted well was 90%. The highest concentration of S,R,S - bisfuranol (BIS) that diffracted well was 20%. Soaking the crystals in solvent concentrations greater than 20% BIS resulted in visible cracks within the crystal

and loss of diffraction. In order to obtain a BIR data set that was comparable to the BIS data set, we soaked the Ras protein crystals at solvent concentrations of 20% and 50 %, following the procedure described in Chapter 5. These data sets allow us to examine the stereospecific binding of these two enantiomers. All data collection was done at the SER-CAT beam line and relevant data collection and refinement statistics are presented in Table 2.3. Validation of the solvent binding sites was done following the solvent validation protocol, see Experimental Procedures.

BIR binding sites can be ranked based on concentration dependence. In Chapter 5 we discussed the results of the 90% R,S,R - bisfuranol soak. This high concentration soak had 19 bound solvent molecules. In a similar set of experiments by English *et al*., crystals of Thermolysin were soaked in a range of isopropanol concentrations (English, Done, *et al.* 1999). English *et al*. found that the solvent binding sites could be experimentally ranked. Of the 12 sites occupied at high isopropanol concentration, only two were occupied at low concentration. Similarly, of the 19 bisfuranol binding sites found at high concentration, only four were found at the lower concentrations of 20% and 50 %.

Non selective solvent binding sites

Of the four sites identified at low R,S,R – bisfuranol solvent concentrations (1203, 1207, 1213 and 1220), two sites (1203 and 1213) were also occupied by the S,R,S bisfuranol molecule at 20%. At both non-selective sites, the conformations of low BIR, high BIR and low BIS are all different, showing that these sites can accommodate multiple solvent conformations, see Figures 2 and 3. English *et. al.* also observed multiple isopropanol conformations at several solvent binding sites in Thermolysin and have suggested that the ability to accommodate multiple ligand conformations is a distinguishing characteristic of

ligand binding sites (English, Groom, et al. 2001). At both of the shared sites, the aliphatic portion of the solvent molecule is oriented to make good hydrophobic interactions with a non-polar residue (Phe28/1203,Tyr32/1213), while the polar portion of the solvent molecule (O1 and O2) is oriented to make a polar interaction with a nearby charged residue. All of the low concentration binding sites have nearby positively charged residues (Arg or Lys) which are positioned to make polar interactions with O1 and O2 of the bisfuranol molecule (Lys147/1203,Arg149/1207,Lys117/1213,Arg123/1220) .

Stereo selective solvent binding sites

At a similar solvent concentration, two additional binding sites for BIR were found (1207 and 1220) that were not occupied by BIS molecules, indicating that these sites are stereo-specific. At 1207, both low and high concentrations of BIR adopt the same conformation. Unlike the non selective solvent binding sites, specific hydrogen bond formation plays an important role in orienting BIR 1207, see Figure 2. The O1 and O2 atoms of BIR 1207 are oriented to make hydrogen bonds with the backbone amide of Phe28 and the side chain amide of Gln22. Because the hydroxyl group is solvent exposed in both stereoisomers, the selectivity at these sites must be guided by subtle structural differences

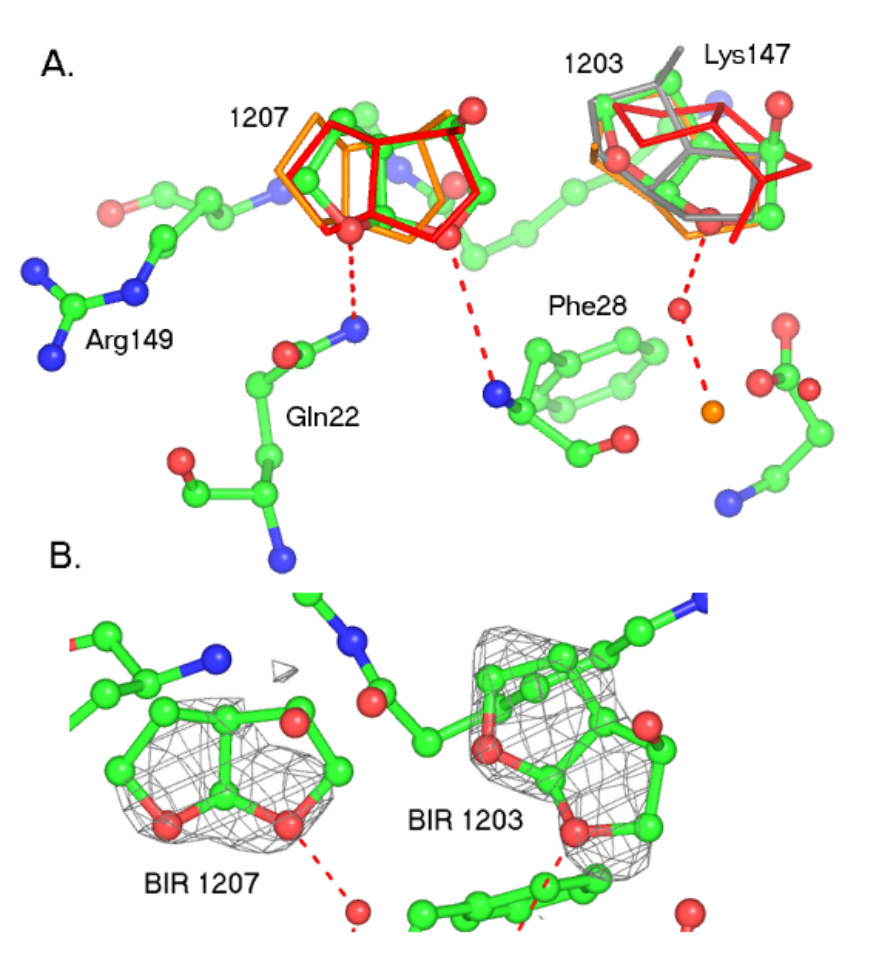


Figure 2 A. Non-selective and stereo-selective Ras solvent binding sites, 1203 and 1207. 20% R,S,R-bisfuranol in green, 20% S,R,S bisfuranol in grey, 50% BIR in orange and 90% BIR in red. B. Electron density for 2Fo-Fc maps contoured at 1.0 Sigma is shown for 20% BIR 1203 and 1207.

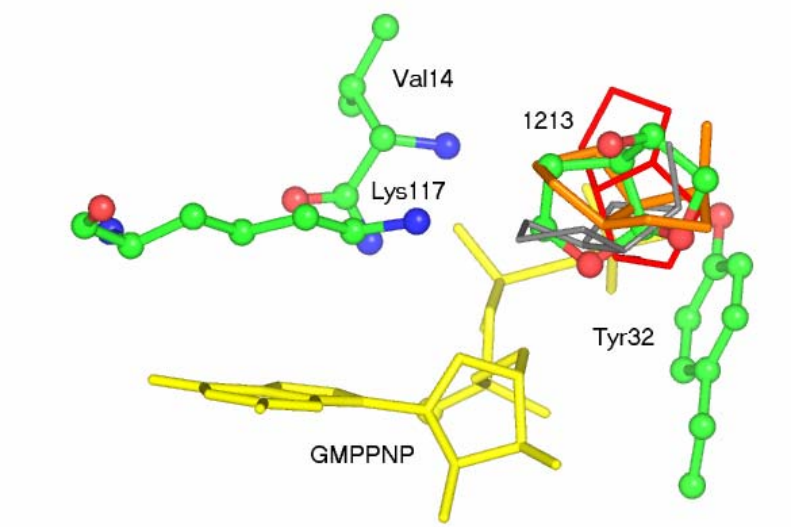


Figure 3. Non-stereo selective Ras solvent binding site 1213. The solvent binding site is sandwiched between Switch I residue Tyr32 and the P-loop (11-14). 20% R,S,R-bisfuranol in green, 20% S,R,S bisfuranol in grey, 50% BIR in orange and 90% BIR in red.

between the two enantiomers. At the other unique site 1220, no BIS molecule is found, but the conformations of low and high BIR molecules are all different. Both of the stereo-specific sites are surrounded by solvent molecules that are bound at high solvent concentrations.

Lysozyme Bisfuranol Binding Sites

Data sets for the bisfuranol soaks of tetragonal HEWL crystals were provided by Dr. Paul Swartz. All data collection was done at 100K at the SER-CAT beamline. The highest concentration of either solvent that gave diffraction quality crystals of Lysozyme was 50%. Initial phases were obtained using a protein only model of HEWL (PDB code: 3LYO), for molecular replacement. Ions, water molecules and solvent molecules were added to the model during multiple rounds of manual rebuilding in O. The relevant data collection and refinement statistics are presented in Table 2.4. All superpositions of organic solvent molecules with substrate were made using the structure of Lysozyme bound to 6 molecules of NAG (PDB code: 1SFG), with the subsites designated based on the location of the six substrate molecules (Von Dreele 2001). However, because Lysozyme binds substrate in several distinct binding modes, these subsite designations do not necessarily agree with subsite designations cited using other Lysozyme / substrate structures or models. At least one of the enantiomers binds in each of the subsites A, D and F within the active site cleft of Lysozyme, see Figure 4. Subsites A and F are stereo-selective and bind only one enantiomer. Subsite D binds both enantiomers, but in different orientations.

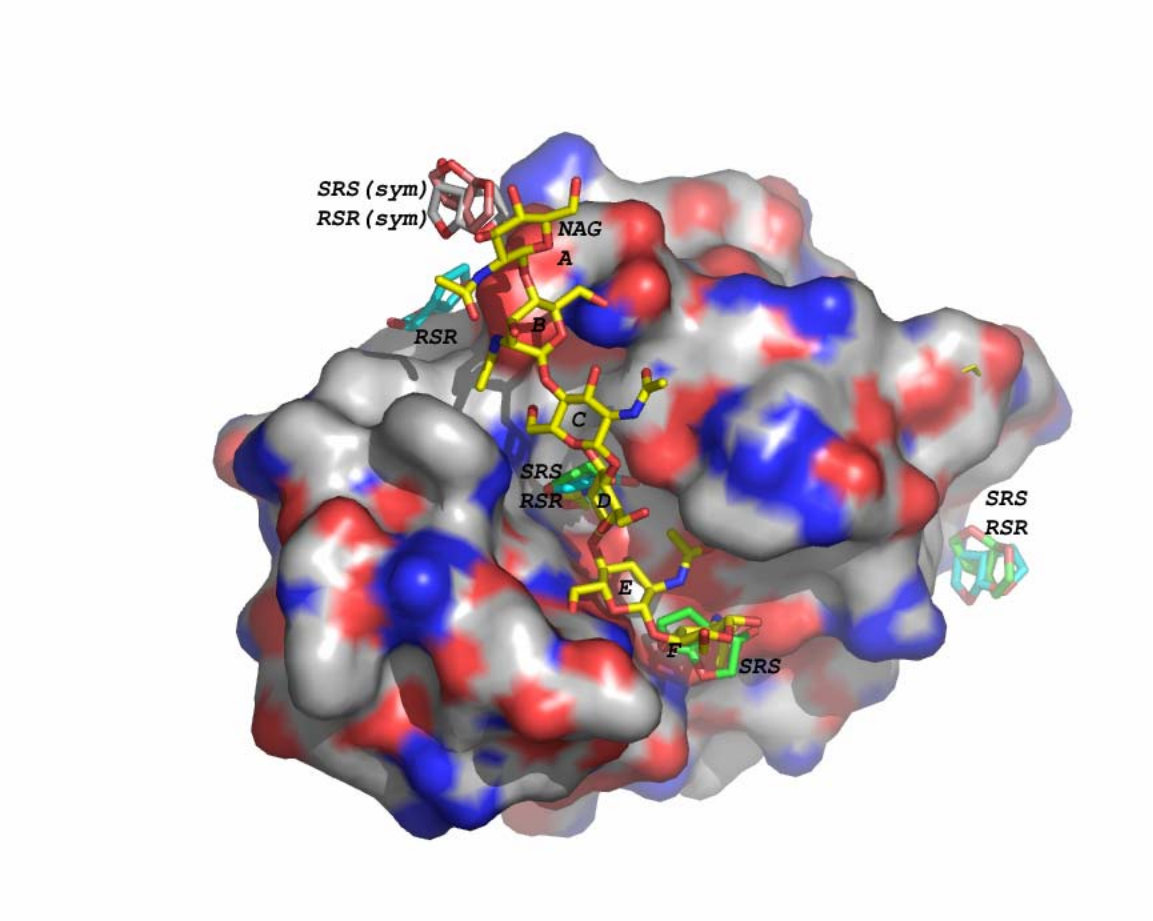


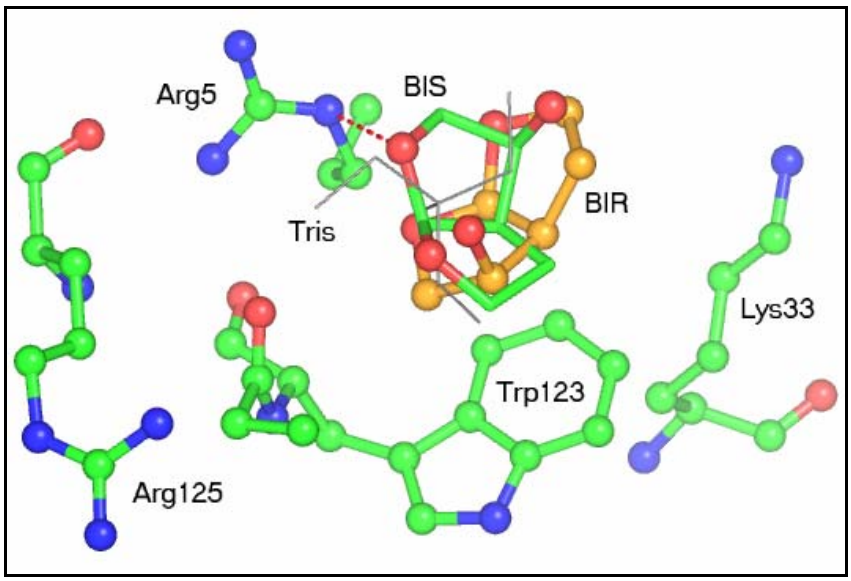
Figure 4. Lysozyme molecular surface with basic residues colored blue and acidic residues colored red, with NAG substrate in yellow and BIR and/or BIS solvent molecules bound at subsites A, D and F colored cyan and green.

Subsite A

Two solvent binding sites are located near subsite A. The first site is located at the top of the binding cleft, 4.5 Å from NAG A. This site is not occupied by the BIS solvent, although there is no obvious steric or hydrogen bonding reason for the preference of the BIR isomer. Lys97 (6.7 Å), Arg21 (7.3 Å) and Arg73 (7.3 Å) all contribute to the overall electropositive character of this binding site. This binding site is further defined by hydrophobic interactions with Trp63, which affects binding mode selectivity, and Leu75. Interestingly, Lys97 from the BIS structure has an alternate conformation that encroaches into the binding site, therefore solvent molecules that bind here displace a flexible protein side chain.

The second solvent binding site is on a shallow pocket distal to the sugar binding site. Symmetry generated solvent molecules BIR401sym and BIS402sym are within 3 Angstroms of Subsite A, however the symmetry position was not used for refinement because the closest protein: solvent interactions occur in the shallow pocket outside of the sugar binding site. The orientation of both solvent molecules is enforced by stacking interactions with Trp123 (4.8 Å) and positive electrostatic contributions from Arg5 (3.6 Å). Lys33 (7 Å) and Arg125 (7 Å) also contribute to the electropositive nature of this binding site, see Figure 5. The weaker electron density and higher B-factors for both solvent molecules reflect the flexible nature of this binding site. Interestingly, HEWL crystals grown in 70% MPD and 50mM Tris have a molecule of Tris that superimposes with the solvent molecules at this site (PDB code: 1DPW) (Weiss, Palm, et al. 2000). Since Tris and bisfuranol are both electonegative molecules and polar interactions between protein and

A.



B.

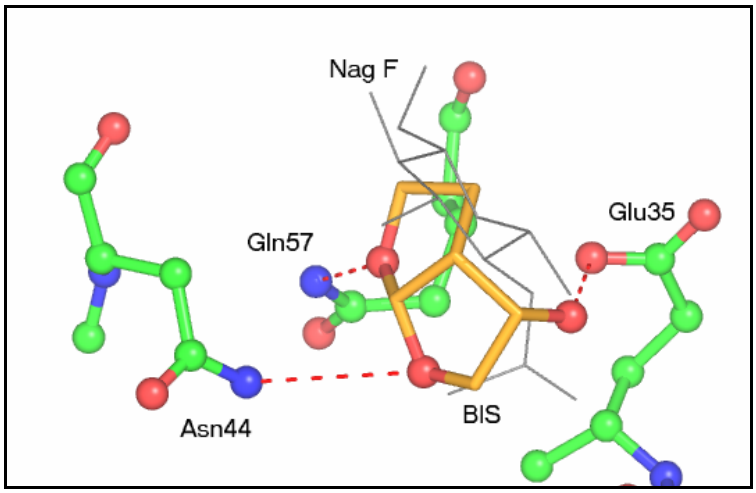


Figure 5. A. Non-stereo selective site, superimposes with a molecule of Tris. Three positively charged side chains create an electropositive solvent binding site. B. Stereo-selective subsite F. Protein-solvent interactions include 3 hydrogen bonds, shown as dashed lines.

ligand become more favorable in a low dielectric solvent, it is not unexpected to find both molecules at this electropositive binding site.

Subsite D

This binding site is found in the middle of the catalytic sugar binding groove near subsite D , see Figure 5. This site is also bound by a HEWL inhibitor, tri-N-acetylchitotriose (Cheetham, Artymiuk, et al. 1992) . The binding site is found in a hydrophobic pocket formed by Ile98, Ile58 and Trp63 at the base of the pocket. The walls of the pocket are formed by Trp62, and Trp108. These residues play a key role in catalysis, especially Trp62, as mutation of this residue to Phe or Tyr effects binding site specificity (Kumagai, Maenaka, et al. 1993). Interestingly, both BIS and BIR molecules are found in different orientations, with very good electron density, and low B-factors. The BIR solvent is rotated 180 degrees on the y axis, relative to the BIS solvent molecule. This orientation allows the hydroxyl group of both solvents to make a good hydrogen bond with the carbonyl oxygen of Ala107. The bisfuran oxygens also make good hydrogen bonds with the backbone amide of Asn59 and the side chain nitrogen of Trp63. This binding site is an interesting example of a site which accomodates both isomers, but in a stereospecific manner.

Subsite F

This site is located at the bottom of the sugar binding cleft, near subsite F, see Figure 6. This binding site is stereo-selective, binding only the S,R,S-bisfuranol enantiomer. A second molecule of Tris from the MPD soaked structure superimposes at this site as well. The interactions between the protein and solvent are dominated by hydrogen bonds at this site. The bisfuran oxygens form hydrogen bonds with the side chain nitrogens of Asn44 and

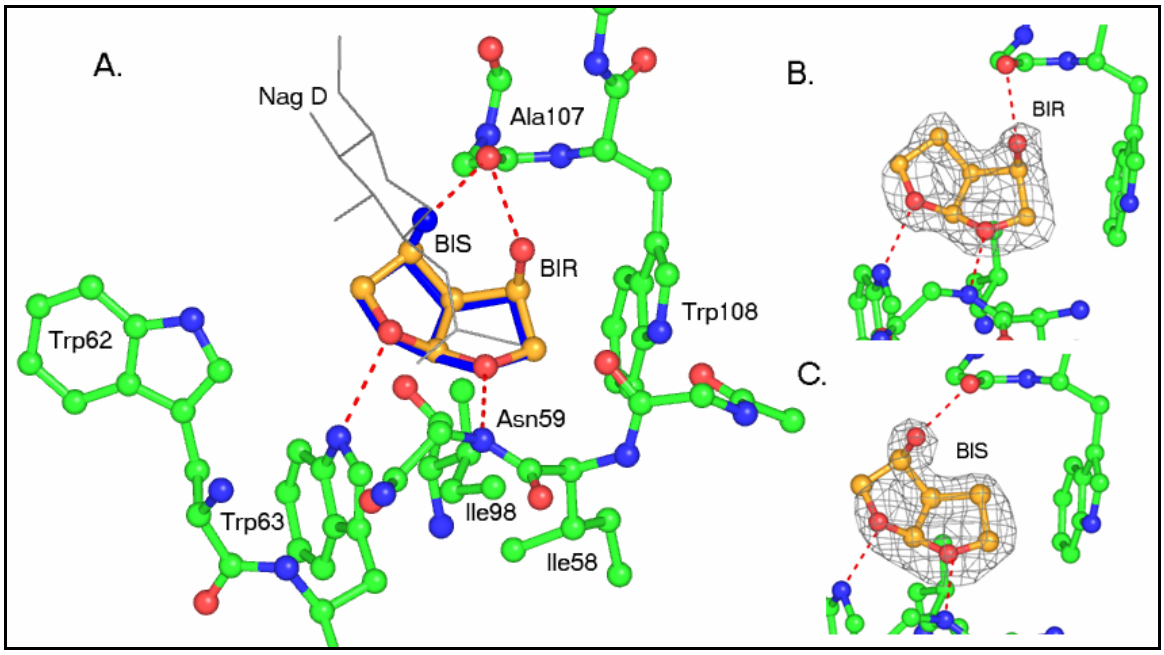


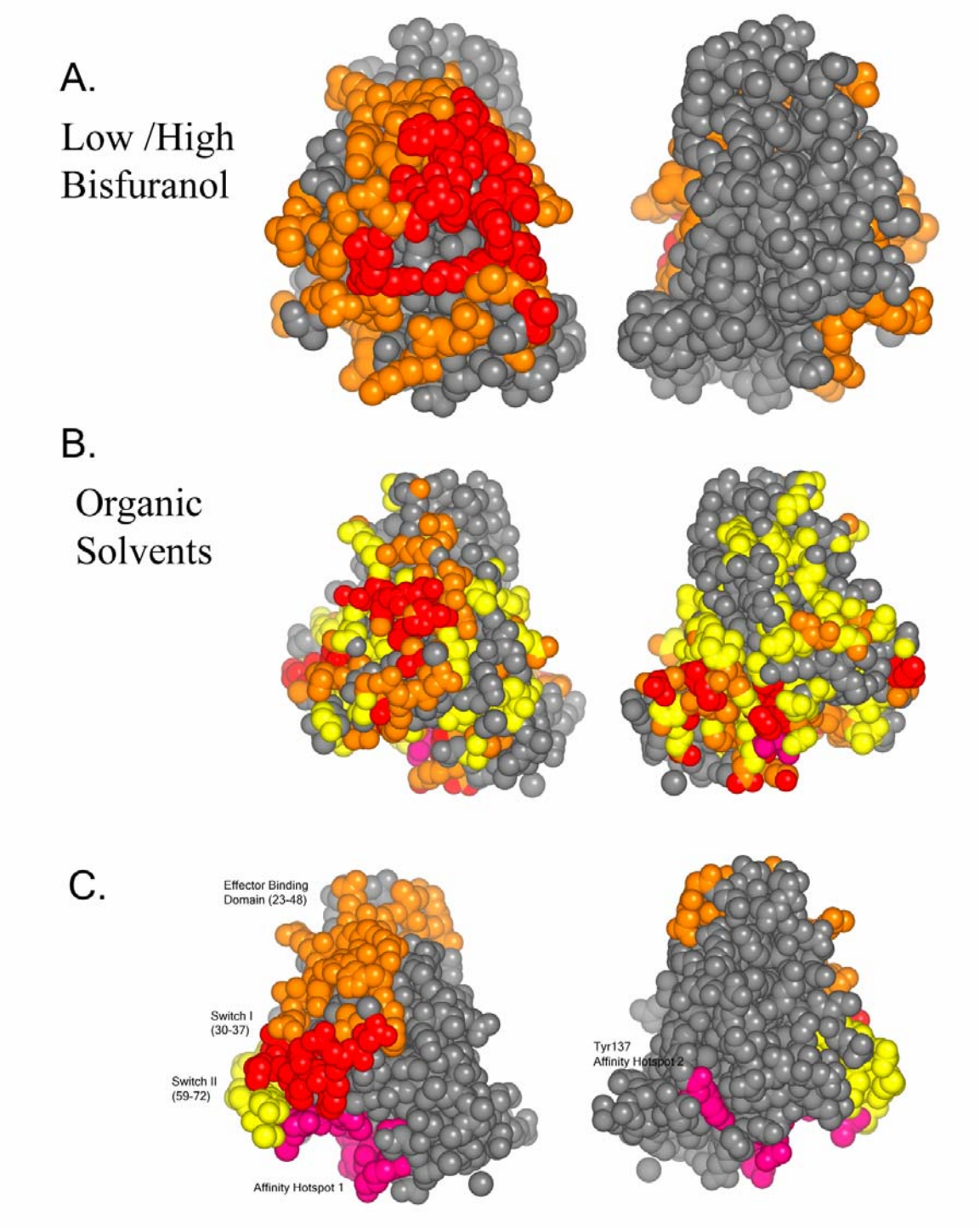
Figure 6. A. Non stereo-selective subsite D. Substrate Nag D is shown, superimposed on the solvent binding site in gray. Both solvent molecules make three hydrogen bonds, shown as dashed lines. B, C. Electron density for 2Fo-Fc maps contoured at 1.0 Sigma is shown for both solvent molecules.

Gln57 and the bisfuranol hydroxyl forms a hydrogen bond with the side chain oxygens of Glu35.

DISCUSSION

In Figure 7, we calculate the protein solvent interface for the four solvent binding sites in Ras and compare this surface with the surface buried by BIR at high concentration and the surface buried by all of the organic solvents, discussed in Chapter 5. The low concentration solvent binding sites (colored red) form a contiguous surface patch that is surrounded on all sides by high concentration solvent binding sites (colored orange). Several of the binding sites occupied at low bisfuranol concentrations are also occupied by multiple organic solvent molecules (colored red). For example, a glycerol molecule from the 70% glycerol soak can be superimposed at BIR 1207. The organization of high and low concentration solvent binding sites on the Ras protein surface is similar in principle to the protein binding affinity maps calculated by Clackson and Wells, in which a small number of residues within the larger protein-protein interface constitute a functional binding epitope responsible for the majority of the protein binding affinity (Clackson & Wells 1995). The region of Ras that contains the functional protein binding epitope mapped by Kiel and Hermann is also mapped in the concentration dependent solvent mapping experiments (Kiel, Serrano, et al. 2004). The functional binding epitope for the Ras-Raf and Ras-RalGDS interaction includes residues E31, D33 and D38. These electronegative residues are not buried by the electronegative bisfuranol molecules, however the surrounding residues are buried. The total buried surface area of the low concentration bisfuranol soak is 877.1 $Å^2$. Buried residues include: Gly13, Gln22, Asn26, His27, Phe28, Asp30, Tyr32, Lys117, Cys118, Asp119, Ala121, Lys147, Thr148 and Arg149. BIR 1220 buries protein residues

Figure 7. A. The Ras protein surface buried by R,S,R bisfuranol solvents soaked at 90% is colored orange, the surface buried by R,S,R bisfuranol solvents soaked at 20% is colored red. B. Also shown is the buried surface calculated from all organic solvents, from Chapter 5. C. A legend with specific regions of Ras labeled for comparison. The protein surface buried by solvent molecules was calculated in CNS.


A.
Low /High
Bisfuranol
B.
Organic
Solvents
C.
Effector Binding
Domain (23-48)
Switch I
(30-37)
Switch II
(59-72)
Affinity Hotspot 1
Tyr137
Affinity Hotspot 2

117-121 and 147-149, which are found on two proximal loops that form a positively charged pocket near the N-terminal portion of the C-terminal alpha helix. Solvent molecules BIR 1203, 1213 and 1207 bury protein residues 13 and 22-32, which are within the known Ras effector binding domain. Unlike the high concentration BIR binding sites discussed in Chapter five, in which hydrogen bonding between the protein and solvent occurs at many of the binding sites, the low concentration BIR binding sites do not include many potential hydrogen bonds. Instead, the presence of nearby positively charged residues is the common denominator at these solvent binding sites.

In Chapter 5, we ranked the high concentration bisfuranol sites by B-factor and electron density fits, see Table 5.1. Our hypothesis was that bisfuranol molecules with lower B-factors and better electron density fits were bound at higher affinity solvent binding sites that should also bind bisfuranol molecules at lower solvent concentration. The low concentration bisfuranol data shows that this hypothesis is incorrect. Only one of the four low concentration solvent binding sites (1203) had a low B-factor and high electron density fit at higher bisfuranol concentrations. Overall, there is no correlation between well ordered bisfuranol molecules and concentration dependence. Therefore, solvent effects likely play a role in formation of the high concentration bisfuranol solvent binding sites. Most bisfuranol molecules interact with at least one additional bisfuranol molecule in the high concentration data set. Because the bisfuranol solvent displaces not only water molecules in the immediate vicinity of the solvent, but surrounding water molecules as well, this creates a local environment in the high concentration bisfuranol soak that is different from the local environment at low concentrations of bisfuranol. Even at 50% BIR, 117 water molecules

were located in electron density maps. This is more than twice the number of water molecules (46) found in the 90% BIR structure, at comparable resolution.

A similar concentration dependence study could not be done with HEWL, because crystals lost diffraction ability at concentrations of solvent greater than 50%. Interestingly, a similar number of bisfuranol binding sites (4) were found for both low concentration studies of Lysozyme and Ras protein. All of the low concentration solvent binding sites are found within or near the active site sugar binding cleft of Lysozyme and the known effector protein binding domain of Ras. In Lysozyme, both solvent molecules bound at subsite D with low B-factors and well-resolved electron density, which is the same subsite targeted by Lysozyme inhibitor molecules. In Ras, the non-selective solvent binding sites bury protein residues that are also well-buried in Ras:protein complexes. It is tempting to speculate that small molecule inhibitors designed to target these specific sites could inhibit specific Ras: protein interactions.

Conclusions

We initiated the concentration dependence studies because we were intrigued by the concentration dependent effects the two bisfuranol enantiomers had on Ras crystals. Because protein crystals have a similar maximum concentration limit for many organic solvents we've tested, we extended this study to Lysozyme to determine whether this was a more general effect. Ras crystals tolerate high concentrations of the R,S,R isomer, but show visible cracks at concentrations over 20% of the S,R,S isomer. Tetragonal Lysozyme crystals tolerate both isomers equally well, with a maximum achievable concentration of 50% for each isomer. Apparently, the enantiomer specific effects seen with Ras are not general to protein crystals.

It will be interesting to see if other nucleotide binding proteins have similar stereo-specific effects that are seen with Ras.

We identified stereo-specific and non-stereospecific binding sites for R,S,R and S,R,S bisfuranol in both the hen egg white Lysozyme and Ras proteins. Analysis of these binding sites shows several common trends. Most of the binding sites identified at 50% solvent concentration or less are found at known Ras protein binding sites or Lysozyme oligosaccharide binding sites. All of the binding sites include nearby electropositive residues (Arg, Lys or His). The stereo-specific binding sites usually include specific protein-solvent hydrogen bonds, whereas the non-stereospecific binding sites accommodate multiple solvent conformations and do not necessarily include specific protein-solvent hydrogen bonds.

Perhaps the most interesting solvent binding site identified was found near Lysozyme subsite D. This binding site buries important catalytic residues and superimposes with a known Lysozyme inhibitor, tri-N-acetylchitotriose. Although these experiments do not allow a direct calculation of binding affinity, both stereoisomers are well ordered, with relatively low B factors (22-28), and each oxygen atom is positioned to make good hydrogen bonds with the protein. Furthermore, these organic solvent binding sites were identified at solvent concentrations that do not show significant bulk solvent effects. Our model indicates that either stereoisomer could be used as a good initial molecule for inhibitor design projects.

APPENDIX

CNS Topology and Parameter Files for residue types R,S,R bisfuranol (BIR) and S,R,S bisfuranol (BIS)

```
Remarks Created by XPLO2D V. 031127/3.2.1 at Tue May 18 17:54:03 2004
Remarks R,S,R bisfuranol, CNS topology file
 set echo=false end

 MASS CW1    14.02700 ! assuming C -> 12.01100 + 1.008 * 2 (Hs)
 MASS CW2    13.01900 ! assuming C -> 12.01100 + 1.008 * 1 (Hs)
 MASS CW3    13.01900 ! assuming C -> 12.01100 + 1.008 * 1 (Hs)
 MASS CW4    14.02700 ! assuming C -> 12.01100 + 1.008 * 2 (Hs)
 MASS CW5    14.02700 ! assuming C -> 12.01100 + 1.008 * 2 (Hs)
 MASS CW6    13.01900 ! assuming C -> 12.01100 + 1.008 * 1 (Hs)
 MASS OW7    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS OW8    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS OW9    17.00700 ! assuming O -> 15.99900 + 1.008 * 1 (Hs)

 autogenerate angles=true end

RESIdue BIR

GROUp
 ATOM  C1   TYPE CW1   CHARge  0.0  END ! Nr of Hs =  2
 ATOM  C2   TYPE CW2   CHARge  0.0  END ! Nr of Hs =  1
 ATOM  C3   TYPE CW3   CHARge  0.0  END ! Nr of Hs =  1
 ATOM  C4   TYPE CW4   CHARge  0.0  END ! Nr of Hs =  2
 ATOM  C5   TYPE CW5   CHARge  0.0  END ! Nr of Hs =  2
 ATOM  C6   TYPE CW6   CHARge  0.0  END ! Nr of Hs =  1
 ATOM  O1   TYPE OW7   CHARge  0.0  END ! Nr of Hs =  0
 ATOM  O2   TYPE OW8   CHARge  0.0  END ! Nr of Hs =  0
 ATOM  O3   TYPE OW9   CHARge  0.0  END ! Nr of Hs =  1

 BOND  C1   C2       BOND  C1   C4       BOND  C2   C3
 BOND  C2 C6
 BOND  C3   O1       BOND  C3   O2       BOND  C4   O1
 BOND  C5   C6
 BOND  C5   O2       BOND  C6   O3

  DIHEdral  C1   C2   C3   O1  ! flat ? (0 degrees = cis)     5.94
  DIHEdral  C6   C2   C3   O2  ! flat ? (0 degrees = cis)     8.51

 IMPRoper  C2   C1   C3   C6  ! chirality or flatness improper    36.59
 IMPRoper  C3   C2   O1   O2  ! chirality or flatness improper    36.41
 IMPRoper  C6   C2   C5   O3  ! chirality or flatness improper    36.25

 ACCEptor  O1   C3
 ACCEptor  O2   C3
 ACCEptor  O3   C6
```

```
END ( RESIdue BIR )



Remarks Created by XPLO2D V. 031127/3.2.1 at Tue May 18 17:54:03 2004
Remarks Auto-generated by XPLO2D
Remarks CNS Parameters for residue type BIR
set echo=false end
 BOND CW1  CW2    1000.0  1.557 ! Nobs =    1
 BOND CW1  CW4    1000.0  1.511 ! Nobs =    1
 BOND CW2  CW3    1000.0  1.525 ! Nobs =    1
 BOND CW2  CW6    1000.0  1.585 ! Nobs =    1
 BOND CW3  OW7    1000.0  1.444 ! Nobs =    1
 BOND CW3  OW8    1000.0  1.435 ! Nobs =    1
 BOND CW4  OW7    1000.0  1.419 ! Nobs =    1
 BOND CW5  CW6    1000.0  1.524 ! Nobs =    1
 BOND CW5  OW8    1000.0  1.417 ! Nobs =    1
 BOND CW6  OW9    1000.0  1.442 ! Nobs =    1

 ANGLe CW2  CW1  CW4     500.0    98.60 ! Nobs =    1
 ANGLe CW1  CW2  CW3     500.0   103.92 ! Nobs =    1
 ANGLe CW1  CW2  CW6     500.0   115.04 ! Nobs =    1
 ANGLe CW3  CW2  CW6     500.0   103.40 ! Nobs =    1
 ANGLe CW2  CW3  OW7     500.0   106.29 ! Nobs =    1
 ANGLe CW2  CW3  OW8     500.0   106.40 ! Nobs =    1
 ANGLe OW7  CW3  OW8     500.0   110.60 ! Nobs =    1
 ANGLe CW1  CW4  OW7     500.0   104.23 ! Nobs =    1
 ANGLe CW6  CW5  OW8     500.0   103.68 ! Nobs =    1
 ANGLe CW2  CW6  CW5     500.0   101.52 ! Nobs =    1
 ANGLe CW2  CW6  OW9     500.0   108.04 ! Nobs =    1
 ANGLe CW5  CW6  OW9     500.0   108.80 ! Nobs =    1
 ANGLe CW3  OW7  CW4     500.0   105.73 ! Nobs =    1
 ANGLe CW3  OW8  CW5     500.0   106.47 ! Nobs =    1


 DIHEdral CW1  CW2  CW3  OW7     750.0 0     0.00 ! Nobs =    1 ... Value
=     5.94
 DIHEdral CW6  CW2  CW3  OW8     750.0 0     0.00 ! Nobs =    1 ... Value
=     8.51


 IMPRoper CW2  CW1  CW3  CW6     750.0 0   35.000 ! Nobs =    1 ... Value
=   36.585
 IMPRoper CW3  CW2  OW7  OW8     750.0 0   35.000 ! Nobs =    1 ... Value
=   36.408
 IMPRoper CW6  CW2  CW5  OW9     750.0 0   35.000 ! Nobs =    1 ... Value
=   36.253

 NONBonded CW1  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW2  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW3  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW4  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW5  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
```

```
 NONBonded CW6  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded OW7  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded OW8  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded OW9  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen



Remarks Created by XPLO2D V. 031127/3.2.1 at Tue May 18 17:46:32 2004
Remarks Auto-generated by XPLO2D
Remarks CNS Topology file for residue type BIS
 MASS CW1    13.01900 ! assuming C -> 12.01100 + 1.008 * 1 (Hs)
 MASS CW2    13.01900 ! assuming C -> 12.01100 + 1.008 * 1 (Hs)
 MASS CW3    13.01900 ! assuming C -> 12.01100 + 1.008 * 1 (Hs)
 MASS CW4    14.02700 ! assuming C -> 12.01100 + 1.008 * 2 (Hs)
 MASS CW5    14.02700 ! assuming C -> 12.01100 + 1.008 * 2 (Hs)
 MASS CW6    14.02700 ! assuming C -> 12.01100 + 1.008 * 2 (Hs)
 MASS OW7    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS OW8    15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS OW9    17.00700 ! assuming O -> 15.99900 + 1.008 * 1 (Hs)

RESIdue BIS

GROUp
 ATOM  C1   TYPE CW1   CHARge  0.0  END ! Nr of Hs =  1
 ATOM  C2   TYPE CW2   CHARge  0.0  END ! Nr of Hs =  1
 ATOM  C3   TYPE CW3   CHARge  0.0  END ! Nr of Hs =  1
 ATOM  C4   TYPE CW4   CHARge  0.0  END ! Nr of Hs =  2
 ATOM  C5   TYPE CW5   CHARge  0.0  END ! Nr of Hs =  2
 ATOM  C6   TYPE CW6   CHARge  0.0  END ! Nr of Hs =  2
 ATOM  O1   TYPE OW7   CHARge  0.0  END ! Nr of Hs =  0
 ATOM  O2   TYPE OW8   CHARge  0.0  END ! Nr of Hs =  0
 ATOM  O3   TYPE OW9   CHARge  0.0  END ! Nr of Hs =  1

 BOND  C1   C2       BOND  C1   C4       BOND  C1   O3
 BOND  C2   C3
 BOND  C2   C6       BOND  C3   O1       BOND  C3   O2
 BOND  C4   O1
 BOND  C5   C6       BOND  C5   O2

  DIHEdral  C1   C2   C3   O1  ! flat ? (0 degrees = cis)    -8.66
  DIHEdral  C6   C2   C3   O2  ! flat ? (0 degrees = cis)    -4.17


 IMPRoper  C1   C2   C4   O3  ! chirality or flatness improper   -36.33
 IMPRoper  C2   C1   C3   C6  ! chirality or flatness improper    34.12
 IMPRoper  C3   C2   O1   O2  ! chirality or flatness improper    35.71

 ACCEptor  O1   C3
 ACCEptor  O2   C3
 ACCEptor  O3   C1

END ( RESIdue BIS )
```

```
Remarks Created by XPLO2D V. 031127/3.2.1 at Tue May 18 17:46:32 2004
Remarks Auto-generated by XPLO2D
Remarks CNS Parameter file for residue type BIS

 BOND CW1  CW2    1000.0  1.569 ! Nobs =    1
 BOND CW1  CW4    1000.0  1.502 ! Nobs =    1
 BOND CW1  OW9    1000.0  1.441 ! Nobs =    1
 BOND CW2  CW3    1000.0  1.521 ! Nobs =    1
 BOND CW2  CW6    1000.0  1.574 ! Nobs =    1
 BOND CW3  OW7    1000.0  1.431 ! Nobs =    1
 BOND CW3  OW8    1000.0  1.435 ! Nobs =    1
 BOND CW4  OW7    1000.0  1.411 ! Nobs =    1
 BOND CW5  CW6    1000.0  1.559 ! Nobs =    1
 BOND CW5  OW8    1000.0  1.421 ! Nobs =    1

 ANGLe CW2  CW1  CW4     500.0   100.60 ! Nobs =    1
 ANGLe CW2  CW1  OW9     500.0   112.09 ! Nobs =    1
 ANGLe CW4  CW1  OW9     500.0   104.99 ! Nobs =    1
 ANGLe CW1  CW2  CW3     500.0   103.63 ! Nobs =    1
 ANGLe CW1  CW2  CW6     500.0   117.20 ! Nobs =    1
 ANGLe CW3  CW2  CW6     500.0   105.18 ! Nobs =    1
 ANGLe CW2  CW3  OW7     500.0   105.89 ! Nobs =    1
 ANGLe CW2  CW3  OW8     500.0   106.64 ! Nobs =    1
 ANGLe OW7  CW3  OW8     500.0   111.70 ! Nobs =    1
 ANGLe CW1  CW4  OW7     500.0   104.13 ! Nobs =    1
 ANGLe CW6  CW5  OW8     500.0   104.01 ! Nobs =    1
 ANGLe CW2  CW6  CW5     500.0    99.88 ! Nobs =    1
 ANGLe CW3  OW7  CW4     500.0   104.79 ! Nobs =    1
 ANGLe CW3  OW8  CW5     500.0   109.33 ! Nobs =    1

DIHEdral CW1  CW2  CW3  OW7     750.0 0     0.00 ! Nobs =    1 ... Value =
-8.66
DIHEdral CW6  CW2  CW3  OW8     750.0 0     0.00 ! Nobs =    1 ... Value =
-4.17

IMPRoper CW1  CW2  CW4  OW9     750.0 0  -35.000 ! Nobs =    1 ... Value =
-36.326
 IMPRoper CW2  CW1  CW3  CW6     750.0 0   35.000 ! Nobs =    1 ... Value
=   34.120
 IMPRoper CW3  CW2  OW7  OW8     750.0 0   35.000 ! Nobs =    1 ... Value
=   35.710

 NONBonded CW1  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW2  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW3  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW4  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW5  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW6  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded OW7  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded OW8  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded OW9  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
```

Chapter 7: Structural mechanism of oxidative regulation of the phosphatase Cdc25B via an intramolecular disulfide bond

Greg Buhrman[1], Benjamin Parker[2], Jungsan Sohn[2], Johannes Rudolph[2] and Carla Mattos[1]

[1]Department of Molecular and Structural Biochemistry
128 Polk Hall – CB 7622
North Carolina State University
Raleigh, NC 27695

and

[2]Department of Biochemistry
LSRC, C125 – Mail stop 3813
Duke University Medical Center
Durham, NC 27710

[Biochemistry in press]

Abstract

Cdc25B phosphatase, an important regulator of the cell cycle, forms an intramolecular disulfide bond in response to oxidation leading to reversible inactivation of phosphatase activity. We have obtained a crystallographic time course revealing the structural rearrangements that occur in the P-loop as the enzyme goes from its apo state, through the sulfenic (Cys-SO^-) intermediate to the stable disulfide. We have also obtained the structures of the irreversibly oxidized sulfinic (Cys-SO_2^-) and sulfonic (Cys-SO_3^-) Cdc25B. The active site P-loop is found in three conformations. In the apo enzyme the P-loop is in the active conformation. In the sulfenic intermediate the P-loop partially obstructs the active site cysteine, poised to undergo the conformational changes that accompany disulfide bond formation. In the disulfide form the P-loop is closed over the active site cysteine, resulting in an enzyme that is unable to bind substrate. The structural changes that occur in the sulfenic intermediate of Cdc25B are distinctly different from those seen in protein tyrosine phosphatase 1B where a 5-membered sulfenyl amide ring is generated as the stable end product. This work elucidates the mechanism by which chemistry and structure are coupled in the regulation of Cdc25B by reactive oxygen species.

Introduction

The Cdc25 phosphatases (Cdc25A, Cdc25B, and Cdc25C in humans) function as essential regulators of cell cycle control during normal eukaryotic cell division and as mediators of the checkpoint response in cells with DNA damage (Strausfeld et al., 1991). These dual-specificity phosphatases (DSPs) belong to a subclass of the protein tyrosine phosphatases (PTPs) and dephosphorylate phospho-Thr14 and phospho-Tyr15 on the

Cdk/cyclin complexes (e.g. Cdk2-pTpY/CycA). Dephosphorylation of the Cdk/cyclin complexes leads to activation of these regulatory kinases, phosphorylation of their numerous cellular targets, and cell cycle progression (Morgan, 1997). The important role of the Cdc25A and Cdc25B homologues in cell cycle regulation is emphasized in the many studies that show their increased expression in a wide variety of cancers (Kristjansdottir and Rudolph, 2004). Regulation of the Cdc25 phosphatases is therefore of critical importance in controlling cell proliferation. The activity of the Cdc25 phosphatases is controlled by factors such as intracellular localization, phosphorylation by Cdk/cyclins, degradation, and oxidation by reactive oxygen species (ROS).

It has become increasingly apparent that ROS play a critical role in controlling the activity of a number of PTPs (Claiborne et al., 1999; Finkel, 2000) including the Cdc25s (Savitsky and Finkel, 2002; Sohn and Rudolph, 2003), and *in vivo* experiments have shown that they are essential in regulating mitogenic processes (Devadas et al., 2002; Finkel, 2000; Meng et al., 2002; Rhee, 1999). The regulation of the PTPs by ROS is mediated by the one shared feature of this class of enzymes, the P-loop containing the active site sequence (CX_5R) (Jackson and Denu, 2001; Tonks and Neel, 2001). The catalytic cysteine in the P-loop in these enzymes sits in a unique dipole environment created by an α-helix, the amides of the five X residues and the conserved arginine (Peters et al., 1998). This conformation generates an unusually low pK_a (5.6 – 6.3) for the active site cysteine (Cys473 in Cdc25B) (Chen et al., 2000; Zhang and Dixon, 1993), making it an optimal nucleophile in the catalytic reaction and also leaving it highly susceptible to oxidation by ROS. Given the instability of the first oxidation of cysteine to sulfenic acid, direct oxidation as a form of reversible regulation is mechanistically feasible only if further oxidation to the sulfinic or sulfonic acids

is prevented (Claiborne et al., 1999). This is accomplished in the phosphatase PTP1B by formation of an unusual sulfenyl-amide bond between the side chain sulfur atom of the active site Cys215 and the backbone amide of the adjacent Ser216 (Salmeen et al., 2003; van Montfort et al., 2003). In other families of phosphatases, such as the Cdc25s, there is a nearby cysteine (Cys426 in Cdc25B), the so-called back door cysteine, that forms a disulfide bond with the active site cysteine under mildly oxidative conditions (Chiarugi et al., 2001; Lee et al., 2002; Savitsky and Finkel, 2002). We have recently demonstrated the transient formation of a sulfenic acid followed by stable formation of a reversible intramolecular disulfide for Cdc25B using kinetic measurements and mass spectrometry (Sohn and Rudolph, 2003). Herein we show the structural changes that occur in the P-loop during this process. We have performed a crystallographic time course capturing four different oxidation states of Cys473, providing the first view of how the conformational flexibility of the P-loop leads to the reversible deactivation of a phosphatase as the result of the disulfide bond with the back door cysteine.

Methods

Expression, purification and crystallization of the catalytic domain of Cdc25B

The catalytic domain of Cdc25B (residues 377 to 566, herein referred to as Cdc25B) was expressed and purified as previously described (Chen et al., 2000). The protein was crystallized by modification of a published procedure (Reynolds et al., 1999). Purified Cdc25B was concentrated to 8.5 mg/mL in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, and 1 mM DTT. Crystals were grown using the vapor diffusion method in the presence of a reservoir solution containing 100 mM Tris (pH 7.25), 1.75 M ammonium

sulfate and 1 mM DTT. The initial crystallization drops contained 2 μL of protein solution and 2 μL of reservoir solution. Crystals grew in about 2 weeks. Cdc25B crystallizes with the symmetry of the orthorhombic space group $P2_12_12_1$ and with one molecule in the asymmetric unit. The unit cell parameters for each of the five new crystal structures are shown in Table I.

Data collection and model refinement

Hydrogen peroxide was used as the ROS in our experiments. Prior to soaking in H_2O_2, the crystallization mother liquor was replaced with a buffer solution containing 100 mM Tris-HCl (pH 7.25), 100 mM NaCl, 25% polyethylene glycol 3350 and 10% glycerol, by extensive buffer exchange. This step was critical to remove sulfate from the active site of Cdc25B and to provide suitable cryo-protection for data collection at 100K.

The soaks in the presence of H_2O_2 were done by exchanging the buffer used to obtain the apo form with one which differed only in the addition of 50 μM hydrogen peroxide. Crystals were flash frozen in liquid nitrogen after soaking for 20, 30, 70 and 80 minutes as well as for 6 hours, and stored for synchrotron data collection. The crystals soaked in the presence of H_2O_2 for 70 minutes and 6 hours did not diffract well. Data sets were collected at 100K at the SER-CAT synchrotron beamline, APS (Argonne, IL) for crystals flash frozen after 20, 30 and 80 minute soaks in the solution containing H_2O_2. Initial $2F_o$-F_c and F_o-F_c electron density maps calculated using the published sulfate-bound structure of Cdc25B (Reynolds et al., 1999) (PDB code 1QB0) as a starting model (with sulfate and water molecules removed) established that our data sets corresponded to the sulfenic, sulfinic and sulfonic forms of the protein respectively.

Having determined that the sulfenic form is present after soaking in presence of H_2O_2 for 20 minutes, and that within the next 10 minute interval there is irreversible progression to the sulfinic form, we were successful in obtaining the disulfide state by removing the crystals from hydrogen peroxide after a 20 minute soak. This was done by buffer exchange back to the conditions used to obtain the apo protein. The crystals were flash-frozen after soaking in this H_2O_2-free buffer for 5, 30, 60, 90 and 120 minutes. X-ray diffraction data were collected for each of these time points at the SER-CAT synchrotron beamline, APS (Argonne, IL) at 100K. The intramolecular disulfide bond could be seen most clearly using a data set collected after the 90 minute back-soak. The starting model for refinement against this data set was again the protein model in the sulfate-bound structure (PDB 1QB0). The wave length of the synchrotron X-rays used for data collection was 1 Å in all cases. Data collection and refinement statistics are shown in Table I.

The program Crystallography and NMR System (CNS) (Brunger et al., 1998) was used for all reciprocal space refinement, with randomly selected 10% of the unique reflections reserved for the calculation of R-free (Brunger, 1997). The program O (Jones et al., 1991) was used for manual rebuilding of the models with visualization of F_o-F_c, and $2F_o$-F_c electron density maps. A similar protocol was used to refine each of the five models. The initial round of refinement consisted of simulated annealing followed by energy minimization and B-factor refinement with CNS. The P-loop was excluded from this initial round of refinement. The protein model was then checked using O and manually adjusted. Subsequent cycles using CNS involved only positional and B-factor refinement. The P-loop and water molecules for each model were built using F_o-F_c electron density maps contoured at the 3σ level. CNS topology and parameter files for the oxidized Cys residues were

generated with HICCUP (Kleywegt and Jones, 1998) and initial structure for Cys-SO^-, Cys-SO_2^- and Cys-SO_3^- were taken from the respective structures of PTP1B (van Montfort et al., 2003) and moved into the appropriate electron density in the corresponding Cdc25B structures. The structures were checked using MolProbity (Lovell et al., 2003).

Substrate binding experiments

Cdc25B (1 μM each of untreated WT, H_2O_2-treated WT, or C473S mutant) was incubated with 1 μM of His_6-Cdk2-pTpY/CycA and nickel-bead depletion experiments were performed as previously described (Rudolph, 2002). Oxidation of Cdc25B was optimized for generation of the disulfide species (Sohn and Rudolph, 2003). To show that Cdc25B oxidized to the disulfide does not protect against Cdc25B (WT) dephosphorylation of protein substrate, Cdc25B (untreated WT, H_2O_2-treated WT, or C473S mutant) was incubated with stoichiometric amounts of Cdk2-pTpY/CycA prepared with ATPγ-^{32}P as previously described (Rudolph et al., 2001). Excess (5-fold) Cdc25B (WT) was used to dephosphorylate unbound protein substrate within 20 seconds.

Results

Crystals of the catalytic domain of Cdc25B (residues 377 to 566, herein referred to as Cdc25B) have the symmetry of space group $P2_12_12_1$. They are grown in the presence of high concentrations of ammonium sulfate and the previously published structure of Cdc25B (PDB code 1QB0) has a sulfate ion bound in the active site (Reynolds et al., 1999). The published coordinates were used to phase the five structures presented here: the apo form and four oxidation states. The overall protein structure remained unchanged in all cases, as indicated

by a pairwise Cα root mean square deviation in the range of 0.10 – 0.27 Å for the structures with the P-loop residues excluded (473-478). Significant conformational changes were observed only for the P-loop, which is anchored by His472 making H-bonds to the carbonyl of Cys473 and to a nearby water molecule, and by Arg479 through a salt bridge to Glu431.

In order to capture the oxidized forms of the phosphatase we first obtained the apo form of the protein by removing the sulfate ion. This was accomplished by transferring the crystals to buffer containing a mixture of polyethene glycol (PEG) and glycerol. The crystal structures of Cdc25B in four oxidation states were obtained by incubating the sulfate-free crystals with 50 μM H_2O_2 for various time intervals. The sulfenic (Cys-SO^-) form was captured by flash-freezing crystals in liquid nitrogen after a 20 min soak in the presence of hydrogen peroxide. Interestingly, accumulation of the sulfenic acid appears to be more favored in the crystals than in solution as this oxidation state is not observed at a significant level using ESI- or MALDI-MS detection of intact protein or peptidic fragments (Sohn and Rudolph, 2003). Even though the disulfide form, with a crosslink between Cys473 and the backdoor Cys426, is the predominant form under mildly oxidizing conditions in solution, we were unable to capture the disulfide in the crystals without removing H_2O_2. Thus, the disulfide bond was trapped by treating the crystals with 50 μM H_2O_2 for 20 minutes (to generate the sulfenic form) followed by incubation for 90 minutes in fresh buffer without H_2O_2 prior to flash freezing. To obtain the irreversibly oxidized sulfinic (Cys-SO_2^-) and sulfonic (Cys-SO_3^-) forms of Cdc25B the crystals were incubated for 30 minutes and 80 minutes respectively in buffer containing 50 μM H_2O_2 prior to flash freezing for data collection. Although the rates by which the various oxidized forms are achieved seem to differ in the crystal from those obtained in solution (Sohn and Rudolph, 2003), our

crystallographic studies are in accord with the solution results in that the sulfenic form is transient in nature and leads either to the intramolecular disulfide bond or to higher oxidation states upon exposure to hydrogen peroxide.

The apo form of Cdc25B

Our crystal structure of the apo enzyme shows no bound sulfate and instead presents three water molecules in the active site (Figure 1). Each of the three water molecules is 3.0 Å from the other two, and their oxygen atoms form an isosceles triangle at the center of the active site P-loop. The first water molecule H-bonds to backbone N atoms of Glu474 and Phe475 and with a terminal N atom of the Arg479 side chain that extends from the end of the P-loop back toward Cys473, completing the active site ring of amides. The second water molecule interacts with the main chain N atoms of Ser477 and Glu478. The third completes the circle of interactions by H-bonding to the backbone N atom of Arg479 and to Nε of the Arg479 side chain. These active site water molecules are located above the reduced Cys473 and mimic the interactions that three of the four sulfate oxygen atoms make with the P-loop residues in the previously published structure (Reynolds et al., 1999) (Figure 1). The fourth oxygen atom of the sulfate faces solvent and there is no crystallographic water molecule in its place in the apo form. It appears that the sulfate ion is an excellent mimic of bound substrate, where three of the phosphate oxygen atoms would be held in place for nucleophilic attack by Cys473, with the rest of the substrate binding away from the P-loop. Interestingly, the conformation of the P-loop in the apo form is unchanged relative to the sulfate-bound structure, suggesting that the apo form of the enzyme is in a conformation ready to bind the phosphorylated substrate (Figure 1). The incoming phosphate group of the substrate must

displace the three well-ordered water molecules, and the release of these molecules would be expected to contribute entropically to the binding event.

Chemistry and structure are coupled in the sulfenic intermediate

The sulfenic, sulfinic and sulfonic forms of Cdc25B all have the same P-loop conformation, but this structure is very different from that of the apo form where Cys473 is reduced. In all three cases there would be steric clashes between the oxygenated cysteine and the amide nitrogen atoms of Phe475 and Ser476 if the P-loop were to remain in the conformation found in the apo form. To avoid this unfavorable interaction, residues 475 through 477 of the P-loop move over the oxygenated Cys473, filling part of the entrance to the active site. The Cα atoms of Ser476 and Ser477 move by 3.8 Å and 2.3 Å, respectively. This dislocates the hydroxyl oxygen atom of Ser476 by 7.9 Å and its carbonyl oxygen by 4.5 Å. These atoms are replaced by water molecules in the structure of the sulfenic form of the phosphatase (Figure 2b). The hydroxyl group of Ser477 moves into the active site and Phe475 is rotated away from the pocket by about 100 degrees on the Chi1 dihedral angle. In this conformation the backbone amides of Ser476 and Glu478 are no longer directed inward toward the active site, leaving the amides of Phe475, Ser477, Arg479 and Gly480 available to interact with the oxidized cysteine. The oxygenated forms of Cys473 reach into the pocket differently than would be expected for substrate. As a result the oxidized cysteine interacts with the amide of Gly480, rather than with the amide of Glu474. In the sulfenic form the oxygen atom is facing the end part of the P-loop and H-bonds to the backbone amides of Arg479 and Gly480 (Figure 2b). This leaves the N-terminal portion of the P-loop without an anchor to the negatively charged sulfenic acid, resulting in higher disorder in this region.

Residues 475-478 have higher thermal factors than other residues in the P-loop, there is a break in the electron density between the backbone Cα and C atoms of Ser477, and there is no electron density for the side chain atoms of Ser476. As discussed below, the crystal structure suggests that the conformation of the P-loop in the sulfenic intermediate, coupled to an increase in disorder, facilitates the formation of the disulfide bond and the resulting conformational change in the active site (Figure 3). In contrast, in the sulfinic and sulfonic forms, both ends of the P-loop are well anchored to the oxygenated active site cysteine. The sulfinic acid interacts with the amides of Phe475, Ser477 and Gly480 (Figure 2c) whereas the sulfonic acid makes all four possible H-bonds (Figure 2d). Not surprisingly, there is an increase in the quality of the electron density for the P-loop, particularly residues 476 and 477, in the higher oxygenated forms of Cdc25B.

The disulfide bond sequesters the active site cysteine and prevents binding to substrate

The P-loop in the disulfide form of Cdc25B collapses toward Arg479 very much like a lid closing over the active site cysteine (Figure 3). In order to form the disulfide bond, Cys473 moves toward the back door Cys426, pulling the initial part of the P-loop with it and changing the direction of the main chain. As a result, the P-loop residues undergo a dramatic flip with Cα displacements relative to the apo form of 3.5 Å, 4.2 Å, 6.3 Å, and 6.9 Å for Glu474, Phe475, Ser476, and Ser477, respectively. This rearrangement leads to the displacement of functional groups on the side chains of Glu474, Ser476, and Ser477 by as much as 9.8 Å. The unusual conformation of the P-loop in the apo form, with all the amides pointing inward and carbonyl groups directed outward, is converted to one in which some of the backbone carbonyl groups are turned inward. The carbonyl oxygen atom of Ser476 now

occupies the center of the distorted P-loop and hydrogen bonds to the amide N atom of Arg479 and to the Nε atom of its side chain, replacing another of the three water molecules found at the center of the amide ring in the apo form. This conformation of the P-loop is supported by clear electron density for the disulfide bond between Cys473 and Cys426, for all of residues Cys473, Glu474, Glu478 and Arg479, as well as for the main chain atoms of Phe475, Ser476 and Ser477. There is a break in the electron density between the N and Cα atoms of Phe475, partial electron density for the Phe475 side chain and no electron density for the side chain Cβ and Oγ atoms of either Ser476 or Ser477. The temperature factors for Phe475, Ser476 and Ser477 are quite high, suggesting a fair amount of local motion in this area of the structure. Thus, in the disulfide conformation, the two ends of the P-loop are firmly anchored by surrounding residues in the protein, with the three residues at the top of the lid over the active site cysteine appearing somewhat more mobile. The main chain atoms of Glu474 and Phe475 in this structure occupy the space where the sulfate ion is found in the sulfate-bound Cdc25B structure, making some of the same interactions with Arg479. This is also the position occupied by the cysteine-phosphate intermediate (Pannifer et al., 1998) and bound substrate (Wang et al., 2000) in other phosphatase structures. Our crystal structure predicts that the disulfide state of Cdc25B is unable to bind substrate without a significant conformational rearrangement that can only occur by breaking the disulfide bond.

It is indeed the case that the disulfide form of Cdc25B does not interact with its substrate Cdk2-pTpY/CycA. In nickel-bead pull-down experiments with stoichiometric ratios of the disulfide form of Cdc25B and His-tagged Cdk2 in the Cdk2-pTpY/CycA complex, we were unable to observe a high affinity complex, as indicated by the fact that the phosphatase is not pulled out of solution by the substrate attached to the Ni^{+} agarose beads (Figure 4). In

contrast, when the same experiment is repeated in the presence of the catalytically incompetent C473S mutant, the complex does form as indicated by co-depletion of the mutant Cdc25B with substrate in the nickel pull-down experiment. Our conclusion that the disulfide form of Cdc25B does not bind substrate was further supported by the fact that pre-incubation of a 1:1 ratio of Cdk2-pTpY/CycA and the C473S mutant of Cdc25B protects the substrate against subsequent dephosphorylation by addition of wild-type Cdc25B, whereas incubation of substrate with the disulfide form does not provide any protection. Taken together, these results establish that C473S is a substrate trapping mutant of Cdc25B and that the disulfide form does not bind the substrate.

Discussion

Comparison with previously published structures of Cdc25

The crystal structure of the apo form of Cdc25A revealed the unique fold of this group of phosphatases and determined structural homology to rhodanese, rather than to other groups of phosphatases (Fauman and Saper, 1996). The salts included in the crystallization solution for Cdc25A were sodium chloride and sodium citrate, neither of which bind in the active site. Thus, the apo form of Cdc25A was obtained. However, no crystallographic water molecules are found in the active site of Cdc25A that are equivalent to the three water molecules present in Cdc25B (Figure 1). The previously described differences in conformations of the P-loop in Cdc25A and Cdc25B (Reynolds et al., 1999) lead to the different hydration pattern in the active site. Each of the three crystallographic water molecules that appear in the active site of the apo Cdc25B forms a hydrogen bond with a protein atom that is in a different position in Cdc25A. Arg436 and Glu431 in Cdc25A are

found in unusual conformations and the carbonyl group of Ser434 is pointing toward the empty phosphate binding site (Reynolds et al., 1999). These residues are the equivalent of Arg479, Glu474 and Ser477 that H-bond respectively with the first, second and third water molecules described above in the results section (Figure 1). It appears that the deviation of the Cdc25A P-loop from the canonical conformation found in Cdc25B and other PTPases/DSPses disrupts the hydrogen bonding network that would position the active site water molecules observed in our apo structure, as well as the sulfate ion found in the previously determined structure of Cdc25B (Reynolds et al., 1999).

In the process of solving the crystal structure of Cdc25A, several data sets were collected in the presence of heavy atoms for phasing. Electron density maps calculated with a data set taken from crystals grown in 3mM sodium vanadate clearly showed a disulfide bond between the active site Cys430 and the back door Cys384 (Fauman et al., 1998). No bound vanadate was observed in this structure and the authors did not discuss the conformation of the P-loop. Furthermore, since a model was not pursued using this data set, we cannot compare the structure of the P-loop in the disulfide form of Cdc25A with that of Cdc25B.

Electron density for the disulfide bond between Cys473 and Cys426 in Cdc25B has also been previously reported for a room temperature crystal structure of Cdc25B (Reynolds et al., 1999). In this case the P-loop was not well ordered, but there was some electron density in the active site which the authors interpreted to be a mixture of low occupancy sulfate ion and solvent molecules. In light of our structure of the disulfide form solved under cryogenic conditions, an alternative possibility is that the electron density in the active site represented partial occupancy of the P-loop conformation that accompanies the intramolecular disulfide bond (Figure 3). As mentioned in the results section above, there is

no electron density for the side chains of Ser476 and Ser477 and the B-factors in this region are high in our structure of Cdc25B containing the intramolecular disulfide bond. It would not be surprising if there is greater disorder in this region at room temperature, impairing interpretation of the electron density map. This interpretation would be made even more difficult if some fraction of the phosphatase molecules in the crystal had a sulfate in the active site, as Reynolds et al (1999) reported to be the case in their structure.

The mechanism of reversible inactivation in Cdc25B

Solution studies have shown that the sulfenic acid is an obligate intermediate in the formation of the intramolecular disulfide bond during reversible regulation of Cdc25B by ROS, and that in the absence of the back door Cys426 the unstable sulfenic acid rapidly oxidizes further to an irreversibly oxygenated form (Sohn and Rudolph, 2003). Here we complement the kinetic experiments with a view of the structural rearrangements that occur in each oxidation step. The manner in which the Cys-SO^- interacts with the P-loop is of critical importance to facilitate the conformational changes necessary to form the intramolecular disulfide bond in Cdc25B. In this intermediate, the section of the P-loop adjacent to the active site cysteine is released from H-bonding interactions observed in the apo form, as it moves toward the closed conformation to avoid steric contacts with the oxygenated cysteine. These changes most likely facilitate the motions that result in the approach of Cys473 toward the backdoor Cys426. Concurrently, the H-bonding interactions between the O atom of the Cys-SO^- and the amide groups on the far side of the P-loop increase the electrophilic character of the oxygenated sulfur atom, thereby making it more prone to nucleophilic attack by the backdoor cysteine. Once the disulfide bond is made, the

active site becomes inaccessible and the protein remains inactive until the disulfide bond is reduced. The superimposed structures of the P-loop in the apo, sulfenic and disulfide forms show the increasing obstruction of the active site cysteine as it reacts to form the disulfide bond (Figure 5). Interestingly, while the P-loops in the sulfinic and sulfonic forms have the intermediate conformation observed for the sulfenic Cdc25B, they do not display the same level of conformational freedom near Cys473. It is significant that the sulfenic form of Cdc25B is a structural intermediate in the conformational rearrangements of the P-loop as well as a chemical intermediate in the oxidation reaction. This phosphatase has evolved to exquisitely couple oxidation chemistry and the structural rearrangements that lead to its reversible inactivation in the presence of ROS.

Comparison with PTP1B

Two important groups of topologically distinct phoshatases (Fauman et al., 1998) consist of those that have the back door cysteine and those that do not. Cdc25B belongs to the former group and is a good representative of the dual specificity phosphatases, while PTP1B belongs to the latter group and has been the canonical example for tyrosine phosphatases. The oxidation states of PTP1B, including the sulfenyl amide structure that forms during reversible inactivation, shed light on the process through which one group of phosphatases is reversibly regulated by ROS (Salmeen et al., 2003; van Montfort et al., 2003). Here we complement the picture by providing the structural mechanism of reversible inactivation in Cdc25B, a dual specificity phosphatase that forms the intramolecular disulfide bond. Although inactivation in both groups results in the collapse of the P-loop over the active site Cys residue (Figure 6), the mechanisms by which this happens are intrinsically

different. Both groups of enzymes use the sulfenic acid as an intermediate to the more stable reversibly oxidized form. While in both cases the sulfenic acid interacts with the far side of the P-loop, the P-loop in PTP1B, in contrast to Cdc25B, remains unchanged relative to its structure in the apo enzyme (van Montfort et al., 2003). This is consistent with a mechanism where Cys215 reacts locally with the amide group of its neighbor Ser216 and is possible due to the presence of the Oγ atom of Ser222 which interacts with Cys-SO^- deep in the active site pocket. Cdc25B has a glycine at the analogous position (Gly480). In the absence of the serine side chain, the sulfenic intermediate interacts with the amide group of Gly480 in Cdc25B, forcing the oxygenated cysteine to be located closer to the surface and causing the P-loop to move in the direction of Cys426 to avoid steric hindrance. The conformational differences in the P-loop observed in the sulfenic intermediate during reversible inactivation of Cdc25B versus PTP1B are mechanistically significant and reflect the specific details of each system.

Interestingly another structure that shows a collapsed P-loop is that of the PTP1B substrate trapping mutant in which the active site Cys215 is changed to a Ser (Scapin et al., 2001), analogous to the C473S mutant of Cdc25B discussed above. The C215S PTP1B structure suggests that simply removing the negative charge in the active site is sufficient to cause the collapse of the unusual amide ring structure of the P-loop. Yet, this mutant has been established to bind substrate (Zhang et al., 2000) in spite of the obstructed active site cysteine observed in the crystal structure. This apparent paradox can be explained by the fact that in the Cys to Ser mutant there is no covalent bond restraining the P-loop to an inactive conformation. As a result, the substrate binding conformation of the P-loop is dynamically accessible and is selected in the presence of a binding partner. The disulfide bond in Cdc25B provides a covalent linkage that holds the P-loop over the active site cysteine such that the

amide ring conformation becomes inaccessible. The covalent bond that sequesters the active site Cys215 in the sulfenyl-amide form of PTP1B is expected to have the same effect as the disulfide bond in Cdc25B, disallowing a substrate binding conformation. It appears that in both groups of phosphatases, ROS is involved in regulation not only by chemically inactivating the phosphatase but also by stabilizing a structure that cannot bind substrate. In the case of Cdc25B and probably of related phosphatases containing the backdoor cysteine, the disulfide bond is buried by the collapsed P-loop, which protects the active site cysteine from either further oxidation or nonspecific reactivation by reducing agents present in the cell.

Solution studies have shown that the disulfide form of Cdc25B is rapidly reduced by thioredoxin but not by glutathione (GSH) (Sohn and Rudolph, 2003). In contrast, PTP1B (Denu and Tanner, 1998), which forms a sulfenyl-amide in the oxidized form, and the LMW phosphatase (Chiarugi et al., 2001), which forms an intramolecular disulfide bond between two cysteine residues in the active site P-loop, are readily reduced by GSH. The sulfenyl-amide structure in PTP1B maintains the active site cysteine in a solvent-exposed conformation that is easily accessed by GSH. A similar situation would be expected for the disulfide bond in the LMW phosphatases. In contrast, the disulfide in Cdc25B is buried at the bottom of the distorted active site, from which GSH is excluded. There is a surface pocket adjacent to the active site lined by many residues, including the back-door Cys426 at the base and residues 428, 444-446, 477-479, 482, 483, 544, 547, 549 and 550 forming the walls (Figure 7). This site was identified as the largest pocket in the disulfide state of Cdc25B using CastP, a program that identifies pockets in proteins (Liang et al., 1998). A solvent accessible surface calculation yields a surface area of 161 A^2 and a volume of 101 A^3

for this pocket, whereas the corresponding values obtained using a molecular surface calculation are 320 A^2 and 429 A^3 respectively. We hypothesize that thioredoxin binds specifically to this pocket to catalyze reduction of the intramolecular disulfide in the Cdc25 dual specificity phosphatases. This may represent a general reduction mechanism utilized by phosphatases that form a disulfide with a vicinal cysteine located outside the active site P-loop, as is also known to occur in PTEN and KAP (Lee et al., 2002; Song et al., 2001). Importantly, this structure-function relationship may differentiate between the two main cellular reducing mechanisms (GSH and thioredoxin) found to reactivate PTPs.

Concluding remarks

Cdc25B and PTP1B, topologically different proteins, are regulated by ROS through distinct mechanisms that achieve the same results: sequestration of the nucleophilic cysteine and obstruction of the active site. The two groups of phosphatases provide a remarkable example of convergent evolution where similar end results are attained by different processes that reflect the separate structural origins and diverse functional roles. The two distinct mechanisms used in the regulation of phosphatases influence the oxidation/reduction strategies employed by these proteins, result in different responses to cellular oxidation levels, and provide specificity to the cellular reductants used in reactivation. It will be interesting to see how these paradigms of structural changes coupled to chemistry in the reversible inactivation of phosphatases extend to other members of the family and allow for the differential signaling in response to the common ROS second messengers.

Coordinates

The coordinates and structure factors for the apo, disulfide, sulfenic, sulfinic and sulfonic forms of Cdc25B have been deposited in the Protein Data Bank (accession codes 1YMK, 1YS0, 1YML, 1YM9 and 1YMD respectively).

Acknowledgments

Data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advance Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. We are grateful to the SER-CAT beamline staff for their support.

References

Brunger, A.T. (1997) Free R Value: Cross-Validation in Crystallography. *Methods Enzymol*, **277**, 366-396.

Brunger, A.T., Adams, P.D., Clore, G.M., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallographic & NMR system (CNS): A new software system for macromolecular structure determination. *Acta Crystallographica D*, **54**, 905-921.

Chen, W., Wilborn, M. and Rudolph, J. (2000) Dual-specific Cdc25B phosphatase: in search of the catalytic acid. *Biochemistry*, **39**, 10781-10789.

Chiarugi, P., Fiaschi, T., Taddei, M.L., Talini, D., Giannoni, E., Raugei, G. and Ramponi, G. (2001) Two vicinal cysteines confer a peculiar redox regulation to low molecular

weight protein tyrosine phosphatase in response to platelet-derived growth factor receptor stimulation. *J Biol Chem*, **276**, 33478-33487.

Claiborne, A., Yeh, J.I., Mallett, T.C., Luba, J., Crane, E.J., 3rd, Charrier, V. and Parsonage, D. (1999) Protein-sulfenic acids: diverse roles for an unlikely player in enzyme catalysis and redox regulation. *Biochemistry*, **38**, 15407-15416.

Denu, J.M. and Tanner, K.G. (1998) Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation. *Biochemistry*, **37**, 5633-5642.

Devadas, S., Zaritskaya, L., Rhee, S.G., Oberley, L. and Williams, M.S. (2002) Discrete generation of superoxide and hydrogen peroxide by T cell receptor stimulation: selective regulation of mitogen-activated protein kinase activation and fas ligand expression. *J Exp Med*, **195**, 59-70.

Fauman, E.B., Cogswell, J.P., Lovejoy, B., Rocque, W.J., Holmes, W., Montana, V.G., Piwnica-Worms, H., Rink, M.J. and Saper, M.A. (1998) Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell*, **93**, 617-625.

Fauman, E.B. and Saper, M.A. (1996) Structure and function of the protein tyrosine phosphatases. *Trends Biochem Sci*, **21**, 413-417.

Finkel, T. (2000) Redox-dependent signal transduction. *FEBS Lett*, **476**, 52-54.

Jackson, M.D. and Denu, J.M. (2001) Molecular reactions of protein phosphatases--insights from structure and chemistry. *Chem Rev*, **101**, 2313-2340.

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica A*, **47**, 110-119.

Kleywegt, G.J. and Jones, T.A. (1994) A super position. *ESF/CCP4 Newsletter*, **31**, 9-14.

Kleywegt, G.J. and Jones, T.A. (1998) Databases in protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **54**, 1119-1131.

Kristjansdottir, K. and Rudolph, J. (2004) Cdc25 phosphatases and cancer. *Chem Biol*, **11**, 1043-1051.

Lee, S.R., Yang, K.S., Kwon, J., Lee, C., Jeong, W. and Rhee, S.G. (2002) Reversible inactivation of the tumor suppressor PTEN by H2O2. *J Biol Chem*, **277**, 20336-20342.

Liang, J., Edelsbrunner, H. and Woodward, C. (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci*, **7**, 1884-1897.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S. and Richardson, D.C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, **50**, 437-450.

Meng, T.C., Fukada, T. and Tonks, N.K. (2002) Reversible oxidation and inactivation of protein tyrosine phosphatases in vivo. *Mol Cell*, **9**, 387-399.

Morgan, D.O. (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol*, **13**, 261-291.

Pannifer, A.D., Flint, A.J., Tonks, N.K. and Barford, D. (1998) Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by x-ray crystallography. *J Biol Chem*, **273**, 10454-10462.

Peters, G.H., Frimurer, T.M. and Olsen, O.H. (1998) Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases. *Biochemistry*, **37**, 5383-5393.

Reynolds, R.A., Yem, A.W., Wolfe, C.L., Deibel, M.R., Jr., Chidester, C.G. and Watenpaugh, K.D. (1999) Crystal structure of the catalytic subunit of Cdc25B required for G2/M phase transition of the cell cycle. *J Mol Biol*, **293**, 559-568.

Rhee, S.G. (1999) Redox signaling: hydrogen peroxide as intracellular messenger. *Exp Mol Med*, **31**, 53-59.

Rudolph, J. (2002) Catalytic mechanism of Cdc25. *Biochemistry*, **41**, 14613-14623.

Rudolph, J., Epstein, D.M., Parker, L. and Eckstein, J. (2001) Specificity of natural and artificial substrates for human Cdc25A. *Anal Biochem*, **289**, 43-51.

Salmeen, A., Andersen, J.N., Myers, M.P., Meng, T.C., Hinks, J.A., Tonks, N.K. and Barford, D. (2003) Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate. *Nature*, **423**, 769-773.

Savitsky, P.A. and Finkel, T. (2002) Redox regulation of Cdc25C. *J Biol Chem*, **277**, 20535-20540.

Scapin, G., Patel, S., Patel, V., Kennedy, B. and Asante-Appiah, E. (2001) The structure of apo protein-tyrosine phosphatase 1B C215S mutant: more than just an S --> O change. *Protein Sci*, **10**, 1596-1605.

Sohn, J. and Rudolph, J. (2003) Catalytic and chemical competence of regulation of cdc25 phosphatase by oxidation/reduction. *Biochemistry*, **42**, 10060-10070.

Song, H., Hanlon, N., Brown, N.R., Noble, M.E., Johnson, L.N. and Barford, D. (2001) Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2. *Mol Cell*, **7**, 615-626.

Strausfeld, U., Labbe, J.C., Fesquet, D., Cavadore, J.C., Picard, A., Sadhu, K., Russell, P. and Doree, M. (1991) Dephosphorylation and activation of a p34cdc2/cyclin B complex in vitro by human CDC25 protein. *Nature*, **351**, 242-245.

Tonks, N.K. and Neel, B.G. (2001) Combinatorial control of the specificity of protein tyrosine phosphatases. *Curr Opin Cell Biol*, **13**, 182-195.

van Montfort, R.L., Congreve, M., Tisi, D., Carr, R. and Jhoti, H. (2003) Oxidation state of the active-site cysteine in protein tyrosine phosphatase 1B. *Nature*, **423**, 773-777.

Wang, S., Tabernero, L., Zhang, M., Harms, E., Van Etten, R.L. and Stauffacher, C.V. (2000) Crystal structures of a low-molecular weight protein tyrosine phosphatase from Saccharomyces cerevisiae and its complex with the substrate p-nitrophenyl phosphate. *Biochemistry*, **39**, 1903-1914.

Zhang, Y.L., Yao, Z.J., Sarmiento, M., Wu, L., Burke, T.R., Jr. and Zhang, Z.Y. (2000) Thermodynamic study of ligand binding to protein-tyrosine phosphatase 1B and its substrate-trapping mutants. *J Biol Chem*, **275**, 34205-34212.

Zhang, Z.Y. and Dixon, J.E. (1993) Active site labeling of the Yersinia protein tyrosine phosphatase: the determination of the pKa of the active site cysteine and the function of the conserved histidine 402. *Biochemistry*, **32**, 9340-9345.

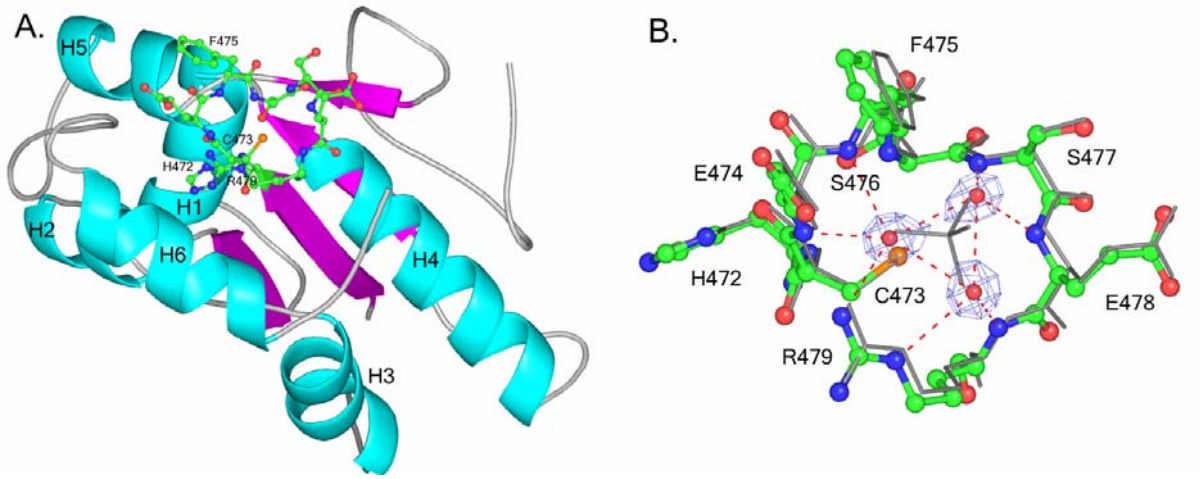


Figure 1. Active site P-loop containing residues 472-479 of apo Cdc25B. The P-loop is shown in sticks with the following color code: carbon, green; oxygen, red; nitrogen, blue; sulfur, orange. Three active site water molecules are represented by red spheres, with electron density taken from a $2F_o$-F_c electron density map contoured at the 1.8 σ level. The sulfate-bound structure of the P-loop (PDB code 1QB0) is shown in gray after superposition on the apo structure. Hydrogen bonds are shown as red dashed lines with distances in Angstroms. All least squares superpositions of structures shown in this and other Figures were calculated using LSQMAN (Kleywegt and Jones, 1994). All Figures in this article, with the exception of Figure 3 were generated using the PyMOL Molecular Graphics System (DeLano Sientific, San Carlos, CA).

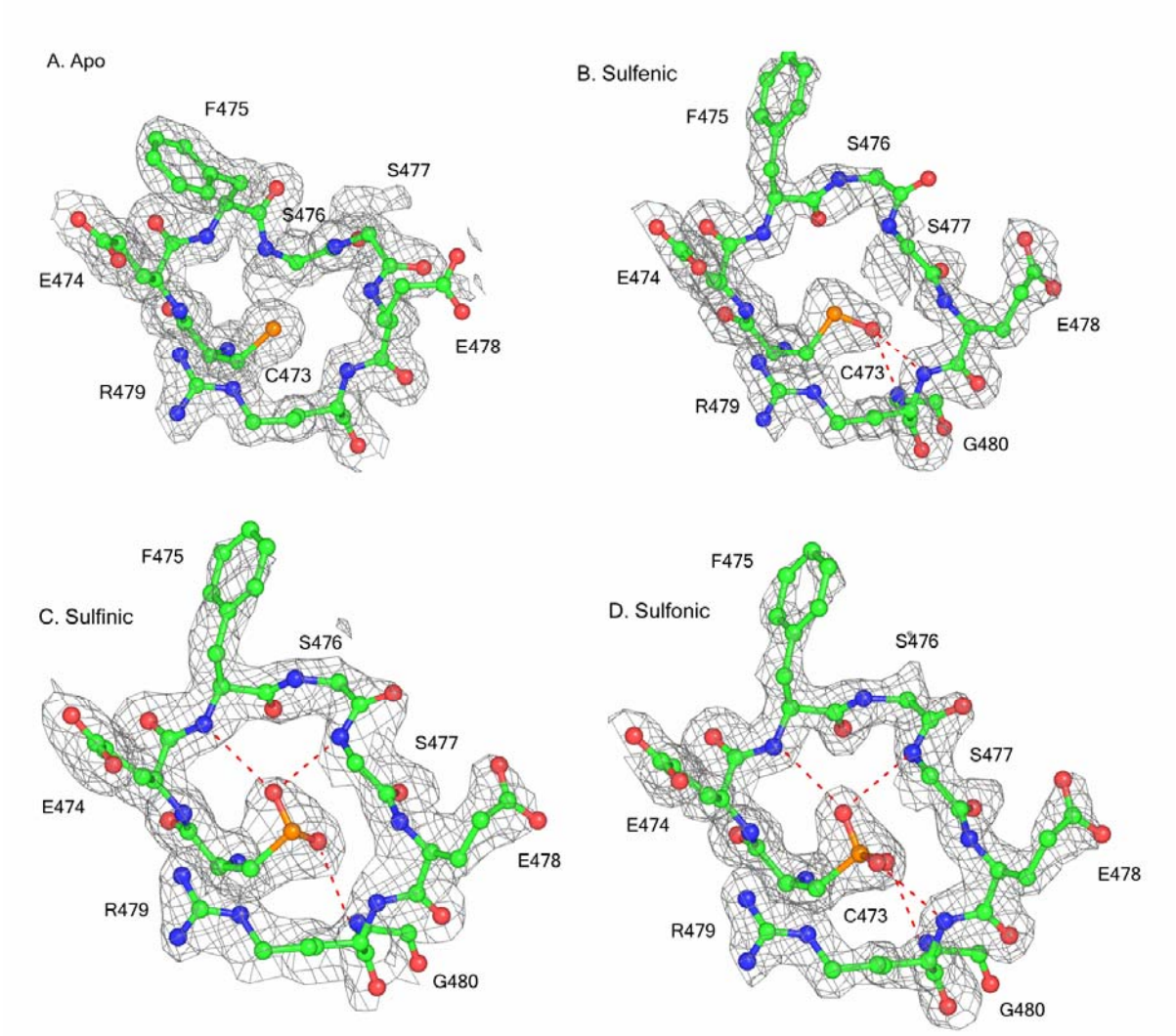


Figure 2. Active site P-loop containing residues 472-480 of the oxidized forms of Cdc25B. a) Apo Cdc25B. b) Sulfenic Cdc25B. c) Sulfinic Cdc25B. d) Sulfonic Cdc25B. The P-loop is shown with the color code and H-bonding representation as in Figure 1. The electron density shown in gray in each panel is from $2F_o$-F_c electron density maps contoured at the 1.0 σ level.

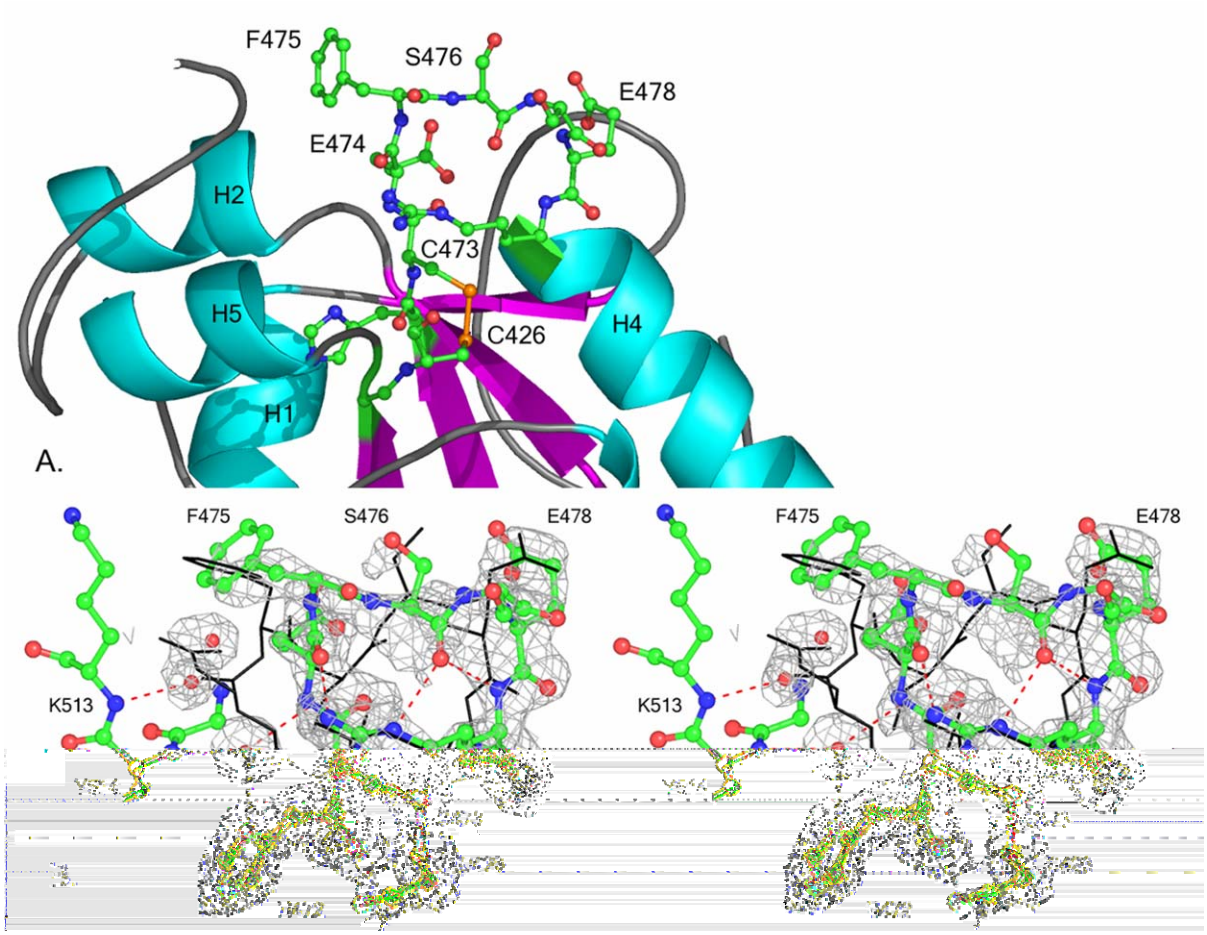


Figure 3. A The disulfide form of the Cdc25B active site is shown in ball and stick representation, with the surrounding protein shown in cartoon representation. B. The active site of the disulfide form of Cdc25B is shown in wall-eyed stereo, with the apo structure of Cdc25B shown in black, for comparison.

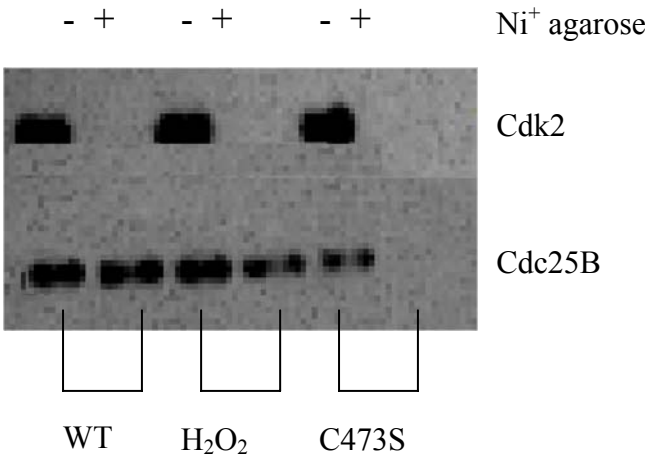


Figure 4. Cdc25B binding experiments with the substrate Cdk2-pTpY/CycA. a) SDS-PAGE showing the results for the nickel bead pull-down experiments. Lanes 1, 2 represent the stoichiometric amounts of wild type Cdc25B and its substrate in the absence and presence of Ni^+ agarose beads respectively. In lane 2 there is no Cdk2, which is His-tagged and therefore removed from solution as it binds to the beads. The Cdc25B is not pulled down with the beads, since it releases the substrate after dephosphorylation. Lanes 3, 4 show the analogous results for Cdc25B under conditions where the intramolecular disulfide bond is formed. Given that the disulfide form has no phosphatase activity we conclude from lane 4 that it does not bind the substrate, as it is still present in solution after the Ni^+ agarose pull-down. Lanes 5, 6 show the results obtained using the C473S mutant of Cdc25B. Lane 6 shows that all of the mutant Cdc25B is depleted from solution. It is pulled down with the Ni^+-bound substrate, since the dephosphorylation reaction does not occur.

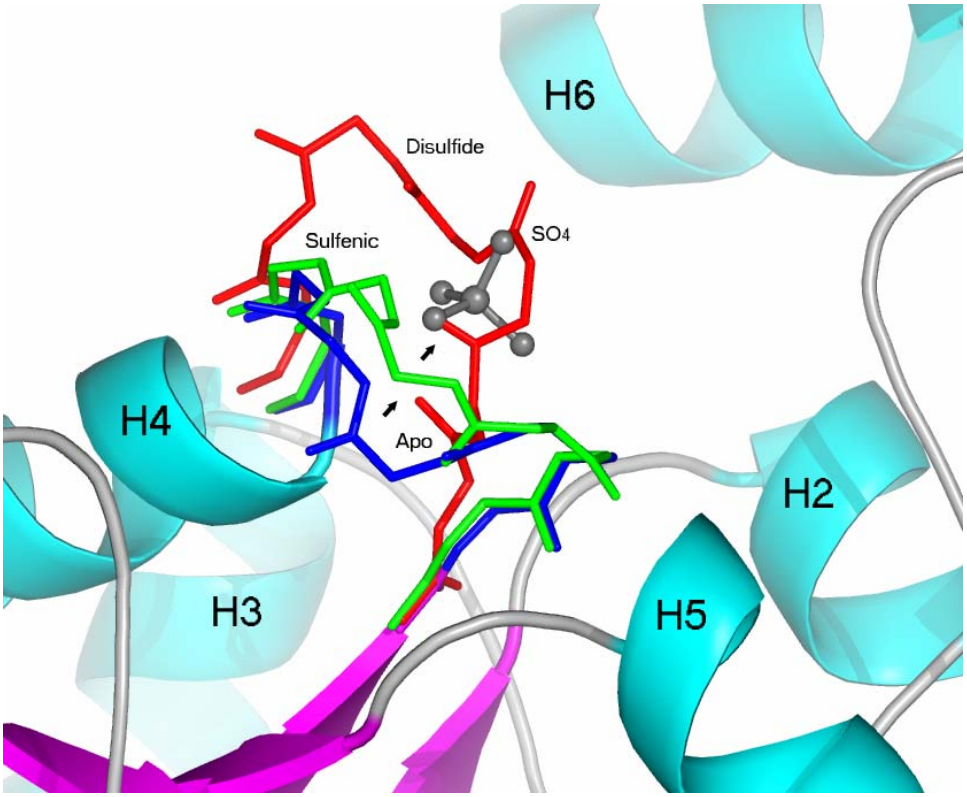


Figure 5. Active site P-loop containing residues 472-479 of Cdc25B as it undergoes changes from the reduced to the disulfide form. Only the main chain atoms are shown. The apo form is shown in blue, the sulfenic intermediate is in green and the disulfide form is in red. The sulfate taken from the sulfate-bound Cdc25B model (PDB code 1QB0) is in black and indicates the location where substrate would be expected to bind. Arrows indicate the direction of loop movement that occurs during structural transitions in the reaction from reduced Cys473 to the sulfenic intermediate and finally to the disulfide.

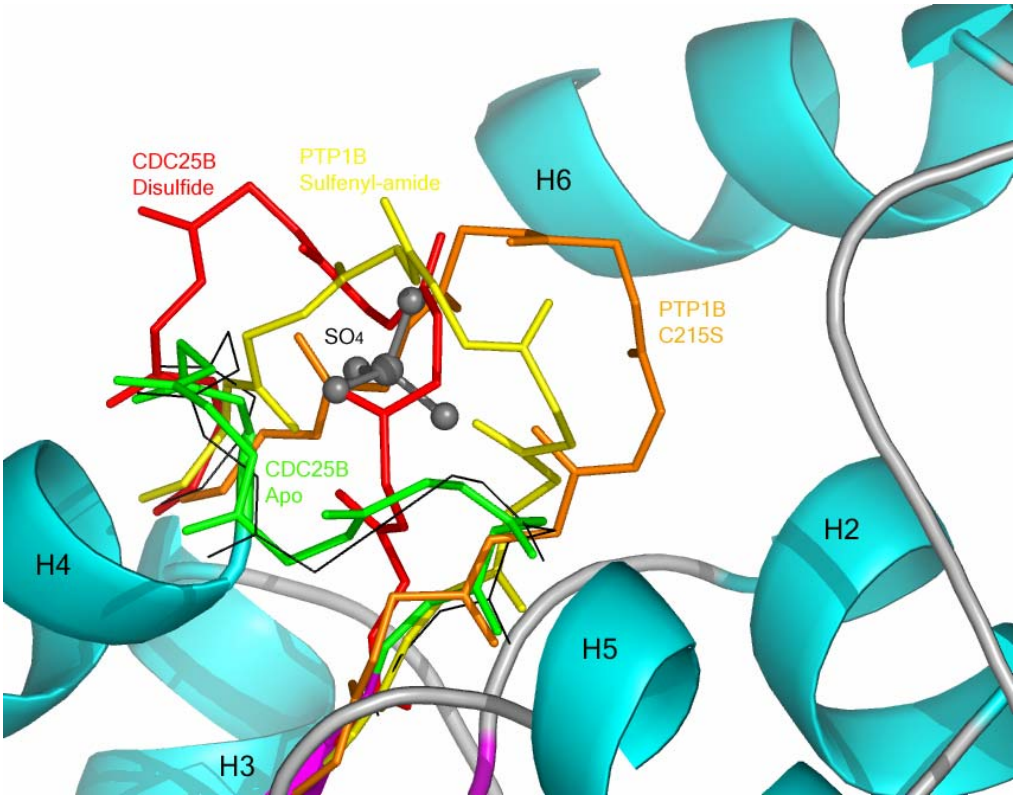


Figure 6. Comparison of the P-loop in deactivated forms of Cdc25B and PTP1B. Only the main chain atoms are included. Residues 472-479 of the disulfide Cdc25B are shown in red. Residues 214-221 of the sulfenyl-amide form of PTP1B (PDB code 1OEM) are shown in yellow and those in the C215S mutant (PDB code 1I57) are in orange. The P-loop in the active conformation of apo Cdc25B and PTP1B (PDB code 1BZC) are shown in blue and grey for comparison. The sulfate taken from the sulfate-bound Cdc25B model (PDB code 1QB0) is in black.

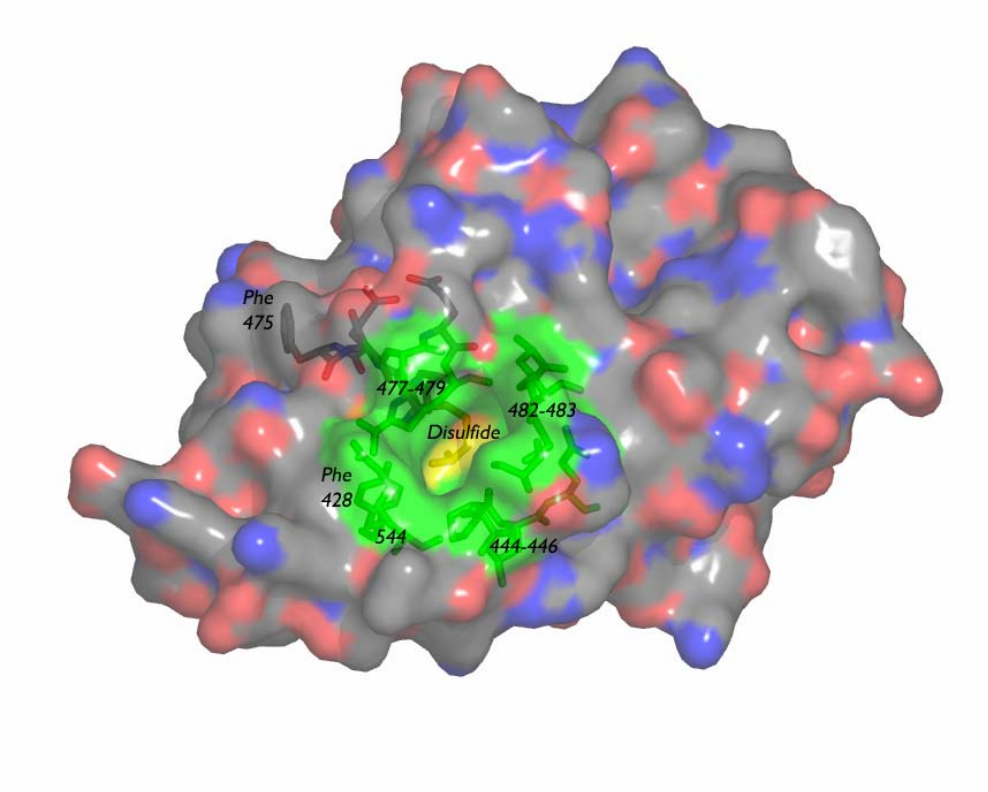


Figure 7. Molecular surface of the disulfide form of Cdc25B. The pocket adjacent to the P-loop allows access to the disulfide. Nitrogen atoms are colored blue, oxygen atoms are colored red and carbon atoms are colored grey. The surface of the pocket is in green and the disulfide bond is shown in yellow. The residues that form the walls of the pocket are drawn explicitly and so are the P-loop residues (with 477-479 and Phe475 labeled).

The active site P-loop of CDC25B adopts multiple conformations, depending on the oxidation state of the active site cysteine residues. Biochemical and genetic studies of CDC25 phosphatases often utilize active site mutants, where the active site cysteine residue is mutated to a non-reactive residue. To determine whether the P-loop is sensitive to these active site mutations, we have solved the X-ray structure of the C473S (19.5 R/20.7 R_{free}, 1.5 Å resolution) and C473D (17.3 R/19.0 R_{free}, 1.6 Å resolution) active site mutants (see Figure 1). The C473S mutation is a 'substrate trapping mutation' known to bind tightly to protein substrate, with nanomolar binding affinity. The presumed orientation of the phosphate from the substrate is clearly seen in the crystal structure. The active site includes a tightly bound sulfate molecule that acts as a mimic of the phosphate group of the protein substrate. This sulfate could not be soaked out of the active site, despite the application of a rigorous soaking procedure that had been used previously to obtain the apo form of wild-type Cdc25B (18.3 R/19.5 R_{free}, 1.7 Å resolution). In contrast, the C473D mutant, while also lacking activity, does not act as a 'substrate trap'. Instead, it binds protein substrate weakly (μM). Our structure provides a clear explanation for this observed behavior. The active site loop of the C473D mutant is distorted such that the main chain forms alternative backbone amide hydrogen bonds with the oxygens of the newly introduced Asp473 side chain. Distortion of the main chain by Asp473 results in an enzyme that does not bind sulfate, phosphate, or substrate in the active site. Thus, we propose that the observed weaker substrate binding of the C473D mutant is mediated by binding at a remote binding site.

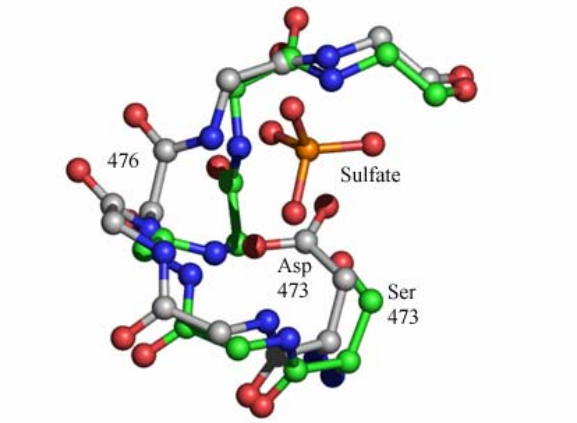


Figure 1. The active site loop of Cdc25B for the mutants C473D (white) and C473S (green). For simplicity, the side chains are omitted for all residues except 473. The sulfate molecule which is found in the C473S structure but not the C473D structure is shown in orange. Note the large main chain displacement at residue 476 in the C473D structure.

REFERENCES

1. M. R. Ahmadian et al., *Proc.Natl.Acad.Sci.U.S.A* 96, 7065-7070 (1999).
2. M. R. Ahmadian, C. Kiel, P. Stege, K. Scheffzek, *J.Mol.Biol.* 329, 699-710 (2003).
3. C. Allin, M. R. Ahmadian, A. Wittinghofer, K. Gerwert, *Proc.Natl.Acad.Sci.U.S.A* 98, 7754-7759 (2001).
4. M. Barbacid, *Annu.Rev.Biochem.* 56, 779-827 (1987).
5. T. A. Binkowski, S. Naghibzadeh, J. Liang, *Nucleic Acids Res.* 31, 3352-3355 (2003).
6. C. C. Blake et al., *Nature* 206, 757-761 (1965).
7. P. A. Boriack-Sjodin, S. M. Margarit, D. Bar-Sagi, J. Kuriyan, *Nature* 394, 337-343 (1998).
8. J. L. Bos, *Cancer Res.* 49, 4682-4689 (1989).
9. A. T. Brunger et al., *Proc.Natl.Acad.Sci.U.S.A* 87, 4849-4853 (1990).
10. A. T. Brunger, *Acta Crystallogr.D.Biol.Crystallogr.* 49, 24-36 (1993).
11. A. T. Brunger et al., *Acta Crystallogr.D.Biol.Crystallogr.* 54 (Pt 5), 905-921 (1998).
12. M. Buck, *Q.Rev.Biophys.* 31, 297-355 (1998).
13. G. Buhrman, S. de, V, C. Mattos, *Structure.(Camb.)* 11, 747-751 (2003).
14. S. L. Campbell, R. Khosravi-Far, K. L. Rossman, G. J. Clark, C. J. Der, *Oncogene* 17, 1395-1413 (1998).

15. S. L. Campbell-Burk, and J. W. Carpenter, *Methods Enzymol* 255, 3-13. (1995).

16. O. Carugo, and *D*. Bordo, *Acta Cryst. D Biol Crystallogr* 55 (Pt 2), 479-483 (1999).

17. A. Cavalli and P. Carloni, *J.Am.Chem.Soc.* 124, 3763-3768 (2002).

18. C. Charron, A. Kadri, M. C. Robert, R. Giege, B. Lorber, *Acta Crystallogr.D.Biol.Crystallogr.* 58, 2060-2065 (2002).

19. J. C. Cheetham, P. J. Artymiuk, D. C. Phillips, *J Mol Biol* 224, 613-628 (1992).

20. J. Cherfils et al., *EMBO J.* 16, 5582-5591 (1997).

21. T. Clackson and J. A. Wells, *Science* 267, 383-386 (1995).

22. J. Colicelli, *Sci.STKE.* 2004, RE13 (2004).

23. K. D. Corbett and T. Alber, *Trends Biochem.Sci.* 26, 710-716 (2001).

24. A. D. Cox and C. J. Der, *Cancer Biol.Ther.* 1, 599-606 (2002).

25. C. J. Der, T. G. Krontiris, G. M. Cooper, *Proc.Natl.Acad.Sci.U.S.A* 79, 3637-3640 (1982).

26. J. F. Diaz, B. Wroblowski, J. Schlitter, Y. Engelborghs, *Proteins* 28, 434-451 (1997).

27. X. Y. Dong, Y. Huang, Y. Sun, *J Biotechnol.* 114, 135-142 (2004).

28. A. C. English, S. H. Done, L. S. Caves, C. R. Groom, R. E. Hubbard, *Proteins* 37, 628-640 (1999).

29. A. C. English, C. R. Groom, R. E. Hubbard, *Protein Eng* 14, 47-59 (2001).

30. D. Eros et al., *Curr.Med.Chem.* 9, 1819-1829 (2002).

31. P. Ertl, B. Rohde, P. Selzer, *J Med.Chem.* 43, 3714-3717 (2000).

32. N. J. Fidyk and R. A. Cerione, *Biochemistry* 41, 15644-15653 (2002).

33. P. A. Fitzpatrick, A. C. Steinmetz, D. Ringe, A. M. Klibanov, *Proc.Natl.Acad.Sci.U.S.A* 90, 8653-8657 (1993).

34. M. E. Furth, L. J. Davis, B. Fleurdelys, E. M. Scolnick, *J.Virol.* 43, 294-304 (1982).

35. R. Gail, B. Costisella, M. R. Ahmadian, A. Wittinghofer, *Chembiochem.* 2, 570-575 (2001).

36. J. Gal, *CNS.Spectr.* 7, 8-13 (2002).

37. J. Gal, *CNS.Spectr.* 7, 45-54 (2002).

38. X. G. Gao et al., *Proc.Natl.Acad.Sci.U.S.A* 96, 10062-10067 (1999).

39. M. Geyer et al., *Biochemistry* 35, 10308-10320 (1996).

40. M. Geyer et al., *Biochemistry* 38, 11250-11260 (1999).

41. M. Geyer, C. Wilde, J. Selzer, K. Aktories, H. R. Kalbitzer, *Biochemistry* 42, 11951-11959 (2003).

42. P. Gideon et al., *Mol.Cell Biol.* 12, 2050-2056 (1992).

43. D. Gorlich et al., *J.Cell Biol.* 138, 65-80 (1997).

44. B. E. Hall, S. S. Yang, P. A. Boriack-Sjodin, J. Kuriyan, D. Bar-Sagi, *J Biol Chem.* 276, 27629-27637 (2001).

45. B. E. Hall, D. Bar-Sagi, N. Nassar, *Proc.Natl.Acad.Sci.U.S.A* 99, 12138-12142 (2002).

46. J. J. Harvey, *Nature* 204, 1104-1105 (1964).

47. C. Herrmann and N. Nassar, *Prog.Biophys.Mol.Biol.* 66, 1-41 (1996).

48. C. Herrmann, *Curr.Opin.Struct.Biol* 13, 122-129 (2003).

49. L. Huang, F. Hofer, G. S. Martin, S. H. Kim, *Nat.Struct.Biol.* 5, 422-426 (1998).

50. N. Hirota, Mizuno, K., and Goto, Y., *J Mol Biol* 275, 365-378. (1998).

51. Y. Ito et al., *Biochemistry* 36, 9109-9119 (1997).

52. J. John, I. Schlichting, E. Schiltz, P. Rosch, A. Wittinghofer, *J.Biol.Chem.* 264, 13086-13092 (1989).

53. J. John et al., *Biochemistry* 29, 6058-6065 (1990).

54. T. A. Jones, J. Y. Zou, S. W. Cowan, Kjeldgaard, *Acta Crystallogr.A* 47 (Pt 2), 110-119 (1991).

55. T. Joneson and D. Bar-Sagi, *J.Mol.Med.* 75, 587-593 (1997).

56. O. El Kabbani et al., *J Med.Chem.* 47, 4530-4537 (2004).

57. C. Kiel, L. Serrano, C. Herrmann, *J Mol Biol* 340, 1039-1058 (2004).

58. Y. Koh et al., *Antimicrob.Agents Chemother.* 47, 3123-3129 (2003).

59. P. G. Kraulis, *J. App. Cryst.* 24, 946-950. (1991).

60. U. Krengel et al., *Cell* 62, 539-548 (1990).

61. I. Kumagai, K. Maenaka, F. Sunada, S. Takeda, K. Miura, *Eur.J Biochem.* 212, 151-156 (1993).

62. S. Kumar, K. Modig, B. Halle, *Biochemistry* 42, 13708-13716 (2003).

63. R. Langen, T. Schweins, A. Warshel, *Biochemistry* 31, 8691-8696 (1992).

64. E. Liepinsh and G. Otting, *Nat.Biotechnol.* 15, 264-268 (1997).

65. K. A. Maegley, S. J. Admiraal, D. Herschlag, *Proc.Natl.Acad.Sci.U.S.A* 93, 8160-8166 (1996).

66. S. C. Mande and M. E. Sobhia, *Protein Eng* 13, 133-141 (2000).

67. S. M. Margarit et al., *Cell* 112, 685-695 (2003).

68. M. S. Marshall, *Trends Biochem.Sci.* 18, 250-254 (1993).

69. M. S. Marshall, *FASEB J.* 9, 1311-1318 (1995).

70. S. F. Mason, *Ciba Found.Symp.* 162, 3-10 (1991).

71. C. Mattos and D. Ringe, *Nat.Biotechnol.* 14, 595-599 (1996).

72. C. Mattos and D. Ringe, *Curr.Opin.Struct.Biol.* 11, 761-764 (2001).

73. L. V. Mello, D. M. van Aalten, J. B. Findlay, *Protein Eng* 10, 381-387 (1997).

74. J. Menetrey and J. Cherfils, *Proteins* 37, 465-473 (1999).

75. Y. I. Melnikova, S. G. Odintsov, Z. I. Kravchuk, S. P. Martsev, *Biochemistry (Mosc.)* 65, 1256-1265 (2000).

76. M. V. Milburn et al., *Science* 247, 939-945 (1990).

77. V. K. Mishraa et al., *Peptides* 22, 567-573 (2001).

78. R. Mittal, M. R. Ahmadian, R. S. Goody, A. Wittinghofer, *Science* 273, 115-117 (1996).

79. S. A. Moodie et al., *Oncogene* 11, 447-454 (1995).

80. N. Nassar et al., *Nat.Struct.Biol.* 3, 723-729 (1996).

81. H. Neuvirth, Raz R., Schreiber G., *J Mol Biol* 338, 181-199 (2004).

82. N. I. Nicely, J. Kosak, S. de, V, C. Mattos, *Structure.(Camb.)* 12, 2025-2036 (2004).

83. Z. Otwinowski, and Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. In Methods in Enzymology, J. Charles W. Carter, and R. B. Sweet, eds. (New York, Academic Press). (1997).

84. M. E. Pacold et al., *Cell* 103, 931-943 (2000).

85. I. Petitpas, A. A. Bhattacharya, S. Twine, M. East, S. Curry, *J Biol Chem.* 276, 22804-22809 (2001).

86. M. R. Pincus, S. S. Zimmerman, H. A. Scheraga, *Proc.Natl.Acad.Sci.U.S.A* 74, 2629-2633 (1977).

87. G. G. Prive et al., *Proc.Natl.Acad.Sci.U.S.A* 89, 3649-3653 (1992).

88. H. Resat, T. P. Straatsma, D. A. Dixon, J. H. Miller, *Proc.Natl.Acad.Sci.U.S.A* 98, 6033-6038 (2001).

89. D. Ringe, *Curr.Opin.Struct.Biol.* 5, 825-829 (1995).

90. D. Ringe and C. Mattos, *Med.Res.Rev.* 19, 321-331 (1999).

91. D. Roccatano, G. Colombo, M. Fioroni, A. E. Mark, *Proc.Natl.Acad.Sci.U.S.A* 99, 12179-12184 (2002).

92. K. Scheffzek et al., *Science* 277, 333-338 (1997).

93. A. Scherer, J. John, R. Linke, R. S. Goody,A. Wittinghofer, E.F. Pai, and K.C. Holmes, *J. Mol. Bio.* 206, 257-259. (1989).

94. I. Schlichting et al., *Nature* 345, 309-315 (1990).

95. J. L. Schmitke, L. J. Stern, A. M. Klibanov, *Biochem.Biophys.Res.Commun.* 248, 273-277 (1998).

96. J. L. Schmitke, L. J. Stern, A. M. Klibanov, *Proc.Natl.Acad.Sci.U.S.A* 95, 12918-12923 (1998).

97. T. Schweins et al., *Nat.Struct.Biol.* 2, 36-44 (1995).

98. T. Schweins, M. Geyer, H. R. Kalbitzer, A. Wittinghofer, A. Warshel, *Biochemistry* 35, 14225-14231 (1996).

99. T. Y. Shih, M. O. Weeks, H. A. Young, E. M. Scholnick, *Virology* 96, 64-79 (1979).

100. K. Shimizu et al., *Proc.Natl.Acad.Sci.U.S.A* 80, 2112-2116 (1983).

101. A. Shurki and A. Warshel, *Proteins* 55, 1-10 (2004).

102. T. A. Soares, J. H. Miller, T. P. Straatsma, *Proteins* 45, 297-312 (2001).

103. R. Sousa, *Acta Crystallogr.D.Biol Crystallogr.* 51, 271-277 (1995).

104. M. Spoerner, C. Herrmann, I. R. Vetter, H. R. Kalbitzer, A. Wittinghofer, *Proc.Natl.Acad.Sci.U.S.A* 98, 4944-4949 (2001).

105. S. R. Sprang, *Curr.Opin.Struct.Biol.* 7, 849-856 (1997).

106. D. Stokoe, S. G. Macdonald, K. Cadwallader, M. Symons, J. F. Hancock, *Science* 264, 1463-1467 (1994).

107. P. J. Stork, *Trends Biochem.Sci.* 28, 267-275 (2003).

108. M. Stumber et al., *J.Mol.Biol.* 323, 899-907 (2002).

109. T. Terada et al., *J.Mol.Biol.* 286, 219-232 (1999).

110. R. Thapar, J. G. Williams, S. L. Campbell, *J.Mol.Biol.* 343, 1391-1408 (2004).

111. L. Tong, M. V. Milburn, A. M de Vos, and S. H. Kim, *Science* 245, 244 (1989).

112. I. A. Topol, R. E. Cachau, A. V. Nemukhin, B. L. Grigorenko, S. K. Burt, *Biochim.Biophys.Acta* 1700, 125-136 (2004).

113. I. R. Vetter, A. Arndt, U. Kutay, D. Gorlich, A. Wittinghofer, *Cell* 97, 635-646 (1999).

114. I. R. Vetter, C. Nowak, T. Nishimoto, J. Kuhlmann, A. Wittinghofer, *Nature* 398, 39-46 (1999).

115. I. R. Vetter et al., *FEBS Lett.* 451, 175-180 (1999).

116. R. B. Von Dreele, *Acta Crystallogr.D.Biol Crystallogr.* 57, 1836-1842 (2001).

117. Z. Wang et al., *Biochim.Biophys.Acta* 1384, 335-344 (1998).

118. M. S. Weiss, G. J. Palm, R. Hilgenfeld, *Acta Crystallogr.D.Biol Crystallogr.* 56 (Pt 8), 952-958 (2000).

119. M. F. Wilkemeyer, C. E. Menkari, M. E. Charness, *Mol Pharmacol.* 62, 1053-1060 (2002).

120. D. G. Winkler, J. C. Johnson, J. A. Cooper, A. B. Vojtek, *J.Biol.Chem.* 272, 24402-24409 (1997).

121. A. Wittinghofer and E. F. Pai, *Trends Biochem.Sci.* 16, 382-387 (1991).

122. N. H. Yennawar, H. P. Yennawar, G. K. Farber, *Biochemistry* 33, 7326-7336 (1994).

123. J. Zeng, H. R. Treutlein, T. Simonson, *Proteins* 35, 89-100 (1999).