Northeast Structural Genomics Consortium (NESG)

Community Outreach PSI-2 Highlights
May 14, 2010

Collected and assembled by Greg Kornhaber, PhD
Community Outreach Coordinator, NESG

PSI-2 Community Outreach Highlights
Table of Contents

1. Structural and Functional Studies of Bacterial MiaA Orthologs

Collaborators: S. Chimnaronk, Mahidol University, Thailand, I. Tanaka, Hokkaido University, Japan M. Atta and M. Fontecave, Institut de Recherches en Technologie et Sciences pour le Vivant, France

NESG structures: 3EXA 2QGN 3D3Q

Bacterial and eukaryotic tRNAs that decode codons starting with uridine have a hydrophobically hypermodified adenosine at position 37 (A(37)) adjacent to the 3'-end of the anticodon (Fig. 1), which is essential for efficient and highly accurate protein translation by the ribosome. However, it remains unclear as to how the corresponding tRNAs are selected to be modified by alkylation at the correct position of the adenosine base (Fig. 1A). In a collaboration with the research groups of Chimnaronk (Mahidol University, Thailand), Tanaka (Hokkaido University, Japan), and Atta/Fontecave (Institut de Recherches en Technologie et Sciences pour le Vivant, France), we have determined crystal structures of three bacterial tRNA isopentenyltransferase (MiaA) in apo- and tRNA-bound forms (Fig. 1C-E), which completely render snapshots of substrate selections during the modification of RNA. A compact evolutionary inserted domain (swinging domain) in MiaA that exhibits as a highly mobile entity moves around the catalytic domain as likely to reach and trap the tRNA substrate (Fig. 1C). Thereby, MiaA clamps the anticodon stem loop of the tRNA substrate between the catalytic and swinging domains, where the two conserved elongated residues from the swinging domain push the two flanking bases out of the tRNA stem loop, thereby causing the flipping of A(37) into the reaction tunnel (Fig. 1D). The sitespecific isopentenylation of RNA is thus ensured by a characteristic pinch-and-flip mechanism and by a reaction tunnel to confine the substrate selection (Fig. 1E).

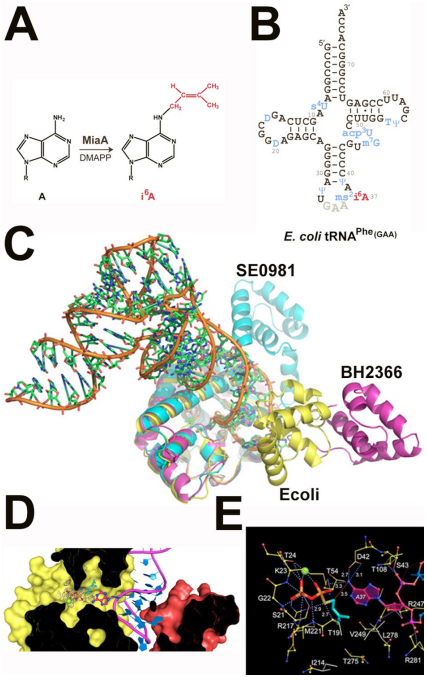
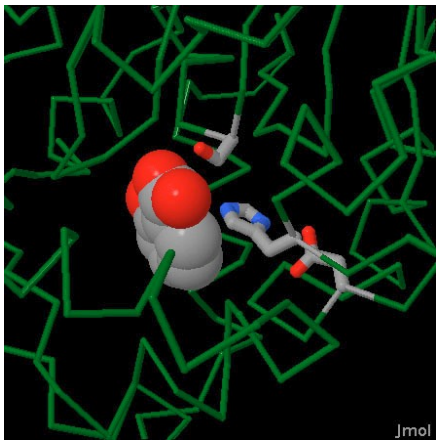


Fig. 1. (A). Reaction catalyzed by MiaA and the position of modification is highlighted in red on E. coli tRNAPhe (panel B). (C). Crystal structures of MiaA orthologs: E. coli (in yellow) in complex with tRNA, and apo Staphylococcus epidermidis SE0981 and Bacillus halodurans BH2366 (in cyan and magenta respectively). The overlaying of the catalytic domain of three MiaAs reveals that the swinging domain is highly mobile in the absence of tRNA. (D). The cross section of E. coli MiaA in complex with tRNA and IPP, showing that A37 (in red for carbon atoms) is flipped into the channel and positioned near IPP (in cyan). (C). The active site of MiaA depicting substrate IPP (in cyan for carbon atoms) and A37 (in magenta).

Chimnaronk, S.; Forouhar, F.; Sakai, J.; Yao, M.; Tron, C. M.; Atta, M.; Fontecave, M.; Hunt, J. F.; Tanaka, I. *Biochemistry* 2009, 48:5057-5065 Snapshots of dynamics in synthesizing N(6)-isopentenyladenosine at the tRNA anticodo

2. Salicylic Acid Binding Protein-2

Collaborators: Dan Klessig, Cornell University, Boyce Thompson Institute for Plant Research
NESG structures: 1Y7H 1Y7I 1XKL
Text Source: David S. Goodsell
http://kb.psi-structuralgenomics.org/KB/archives.jsp?pageshow=17



Centuries ago, both the ancient Greeks and native Americans discovered that willow bark and other plants can dull pain. By analyzing these traditional cures, scientists in the nineteenth century extracted the active molecule, salicylic acid, and developed a modified molecule with better drug properties, acetylsalicylic acid or aspirin. In our bodies, aspirin blocks an enzyme that builds one of the molecules of pain signaling. Plants, on the other hand, use salicylic acid for an entirely different type of signaling.

Defensive Action

Plants have developed a complex and multi-layered system to protect themselves from attack by bacteria and viruses. When a leaf gets infected, for instance by tobacco mosaic virus, the local cells make the ultimate sacrifice, inducing a form of programmed cell death. This helps control the spread of the virus by proactively removing all infectable cells in the neighborhood. At the same time, the plant launches a more systemic defense. It sends a signal to all of its distant parts, telling them to build defensive proteins and ready themselves for attack. These defenses are costly, and may result in stunted growth, but this is better than completely losing the battle against the attacker.

Aromatic Signals

Methyl salicylate, the methyl ester of salicylic acid, is one of the signals that spreads through plants, readying them for attack. Methyl salicylate is a familiar molecule, since it provides the distinctive taste and smell of wintergreen flavorings. It is used as a neutral messenger, which is created by cells under attack and delivered to cells throughout the plant. Then, the enzyme SABP2 (salicylic acid binding protein 2) takes methyl salicylate and cleaves off the methyl group, releasing active salicylic acid, which then stimulates the production of defensive proteins in the target cells.

SABP2 in Action

SABP2 was originally discovered based on its ability to bind to salicylic acid (hence its name), but the recent structure of the protein solved by researchers at the NESG revealed its role in cleavage of methyl salicylate and inhibition of the reaction by the product, salicylic acid. The structure, available in PDB entry 1y7i, shows that SABP2 is one of a class of alpha/beta hydrolase enzymes that cleave small esters and other molecules. The active site completely surrounds the molecule, recognizing both the distinctive aromatic ring and the acidic group. A catalytic triad reminiscent of the digestive serine proteases performs the

cleavage reaction. Based on this structure, researchers have designed analogues of salicylic acid to probe signaling methods in other plants. To explore this structure in more detail, you can click on the image below for an interactive Jmol view of the active site.

Forouhar, F.; Yang, Y.; Kumar, D.; Chen, Y.; Fridman, E.; Park, S. W.; Chiang, Y.; Acton, T. B.; Montelione, G. T.; Pichersky, E.; Klessig, D. F.; Tong, L. *Proc Natl Acad Sci U S A* 2005, 102:1773-1778 Structural and biochemical studies identify tobacco sabp2 as a methyl salicylate esterase and implicate it in plant innate immunity.

Park, S. W.; Liu, P. P.; Forouhar, F.; Vlot, A. C.; Tong, L.; Tietjen, K.; Klessig, D. F. *J Biol Chem* 2009, 284:7307-7317 Use of a synthetic salicylic acid analog to investigate the roles of methyl salicylate and its esterases in plant disease resistance.

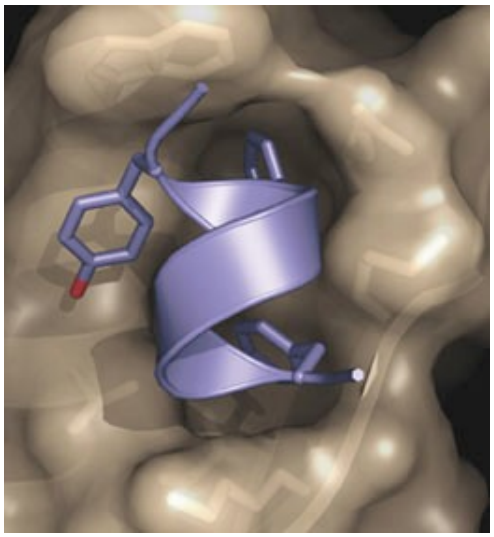
Park, S. W.; Kaimoyo, E.; Kumar, D.; Mosher, S.; Klessig, D. F. *Science* 2007, 318:113-116 Methyl salicylate is a critical mobile signal for plant systemic acquired resistance.

Loake, G.; Grant, M. *Curr Opin Plant Biol* 2007, 10:466-472 Salicylic acid in plant defence--the players and protagonists.

Durrant, W. E.; Dong, X. *Annu Rev Phytopathol* 2004, 42:185-209 Systemic acquired resistance.

3. Influenza A virus NS1 Effector Domain : Human CPSF30 Complex – Antiviral Evasion

Collaborators: Robert Krug, University of Texas at Austin
NESG structures: 2RHK
Text Source: Maria Hodges
http://kb.psi-structuralgenomics.org/update/2008/10/full/fa_psisgkb.2008.10.html



Influenza A viruses cause both seasonal flu outbreaks and the severe pandemics that result in high death rates. The H5N1 viruses, often referred to as bird flu, are type A influenza viruses, as was the virus responsible for the flu pandemic in 1918 that caused approximately 50 million deaths.

Fig. 1. The hydrophobic pocket on the surface of the influenza A protein NS1A (brown) interacts with a fragment of the human CPSF30 protein (blue). This structure provides insights into the mechanism by which influenza A suppresses the production of interferon-β. (PDB 2RHK)

In infected cells, double-stranded RNA produced by the virus usually triggers the production of interferon-β, a cytokine with antiproliferative and antiviral activities. But influenza A avoids this response by producing the virulence factor NS1A, which inhibits the production of interferon-β.

The amino-terminal domain of NS1A binds to double-stranded RNA (Yin et al., 2006), but previous studies have shown no role for this fragment in inhibiting interferon-β messenger RNA production. The remainder of the protein, residues 85–215, termed the effector domain, binds to the 30 kDa host protein cleavage and polyadenylation specificity factor (CPSF30). This factor is required for all 3′ end processing of pre-mRNAs, including interferon-β mRNA.

A collaboration between the Northeast Structural Genomics Consortium (NESG) and the laboratories of Robert Krug at the University of Texas at Austin and Kalyan Das and Eddy Arnold at Rutgers University resulted in the 1.95 Å resolution X-ray crystallographic structure of NS1A of human influenza A/Udorn/72 (Ud) virus in a complex with CPSF30. The complex was formed by the second and third zinc finger motif (F2F3) domain of CPSF30 and the effector domain of NS1A. The F2F3 domain was chosen because it had been previously shown to bind to Ud NS1A and because expression of F2F3 in virus-infected cells leads to inhibition of Ud virus replication and increased interferon-β production.

The complex formed is tetrameric, with two F2F3 molecules wrapped around two NS1A molecules, whose effector domains interact in a head-to-head conformation. The structure revealed a binding pocket for CPSF30 that is

composed mainly of hydrophobic residues on NS1A.

The authors mutated three individual residues within the pocket: Gly184, Trp187 and Gln121. Substitutions of Gly184 and Trp187 with arginine abolished binding, as did mutation of Gln121 to alanine.

The researchers then performed an in-depth analysis of the Gly184Arg mutation but first they confirmed using circular dichroism and NMR that the substitution did not disrupt the overall fold.

Then they generated a mutant virus that produced an NS1A protein with the Gly184Arg mutation and examined its effect on the processing of interferon-β pre-mRNA. Substantial levels of unprocessed interferon-β mRNA were found in cells infected with the wild-type virus, whereas roughly one fifth of the unprocessed mRNA accumulated in cells infected with the mutated virus. In contrast, they found that the amount of mature interferon-β mRNA was five times that found in wild-type-infected cells. They concluded that processing of interferon-β pre-mRNA is much more efficient in cells infected with the Gly184Arg virus, thereby explaining why replication of the Gly184Arg virus was attenuated.

The amino acids in the CPSF30-binding pocket are highly conserved among human influenza A viruses, including H5N1 viruses isolated from humans and the pandemic 1918 strain. Thus, this site looks like a good target for antiviral drug development.

The structure also showed that the interaction site between NS1A and CPSF30 stretches beyond the pocket. Two amino acids outside of the pocket, Phe103 and Met106, are highly conserved and help to stabilize the complex through hydrophobic interactions. These residues are conserved in more than 99% of NS1A proteins, but a few of the proteins have other residues at this site.

Influenza A viruses H5N1, HK97 (the 1997 bird flu) and PR8 encode NS1A proteins with different amino acids at positions 103 and 106. In HK97, binding is stabilized in infected cells by the interaction of the viral polymerase with the NS1A:CPSF30 complex. All H5N1 viruses isolated since 2003 have the consensus Phe103 and Met106.

PR8 uses a different approach from HK97. At position 103, it has a hydrophilic residue that almost eliminates binding; instead it works by suppressing the activation of the transcription factor IRF-3 by an as yet unknown mechanism that does not involve the NS1A protein. Only five out of approximately 2,800 influenza A viruses isolated from humans have a hydrophilic residue at position 103, and the last such virus appeared in 1976. All influenza A viruses currently circulating in humans have the consensus Phe103 and Met106, demonstrating the importance of the CPSF30 binding site for the replication of influenza A viruses in humans.

Consequently, the structure of the NS1A:CPSF30 complex identifies the CPSF30- binding pocket on NS1A as a potential target site for the development of small-molecule antiviral drugs.

Das, K.; Ma, L. C.; Xiao, R.; Radvansky, B.; Aramini, J.; Zhao, L.; Marklund, J.; Kuo, R. L.; Twu, K. Y.; Arnold, E.; Krug, R. M.; Montelione, G. T. *Proc Natl Acad Sci USA* 2008, 105:13093-13098 Structural basis for suppression of a host antiviral response by influenza A virus.

Yin, C.; Khan, J. A.; Swapna, G. V.; Ertekin, A.; Krug, R. M.; Tong, L.; Montelione, G. T. *J Biol Chem* 2007, 282:20584-20592 Conserved surface features form the double-stranded RNA binding site of non-structural protein 1 (NS1) from influenza A and B viruses.

4. Influenza A virus NS1 dsRNA-binding Domain

Collaborators: Robert Krug, University of Texas at Austin
NESG structures: 2RHK

Influenza viruses cause a highly contagious respiratory disease in humans, and are responsible for periodic wide-spread epidemics, or pandemics, with high mortality rates. The most devastating pandemic occurred in 1918, resulting in approximately 30 million deaths worldwide. The avian H5N1 influenza A viruses, currently circulating in birds, are candidates for causing such a pandemic if they acquire efficient human-to-human transmission. "Swine flu" is caused by a less virulent influenza A H1N1 virus. Influenza A non-structural protein 1 (NS1A) is a multi-functional dimeric protein that participates in both protein-RNA and protein-protein interactions. NS1A plays a key role in countering host cell antiviral defenses and in viral virulence. The 73-residue N-terminal domain of NS1A [NS1A(1-73)] forms a symmetric homodimer with a unique six-helical chain fold and binds double-stranded (ds) A-form RNA (Chien *et al.*, 1997; Chien *et al.*, 2004; Liu *et al.*, 1997). As a Community Outreach activity of NESG project, we have characterized dsRNA-binding epitope of NS1A(1-73) using NMR methods (Chien *et al.*, 2004; Yin *et al.*, 2007). We also determined the 2.1 Å X-ray crystal structure of the corresponding dsRNA-binding domain, NS1B(15-93), from human influenza B virus (Yin *et al.*, 2007). Comparison of surface features of NS1A(1-73) and NS1B(15-93) reveal highly conserved "tracks" of basic and hydrophilic residues, complementary to the polyphosphate backbone conformation of A-form dsRNA. Chemical shift perturbation data, together with docking calculations, provided a model of the NS1:dsRNA complex, shown in Fig.1 (Yin *et al.*, 2007). As predicted by this model, mutation of NS1A surface residue Arg-38 to Ala abrogates dsRNA binding *in vitro*, and a recombinant influenza A virus expressing the [R38A]-NS1A mutant protein is highly attenuated. These structural data reveal key features of NS1:dsRNA interactions, and provide valuable information for the development of antiviral drugs targeting influenza A viruses.

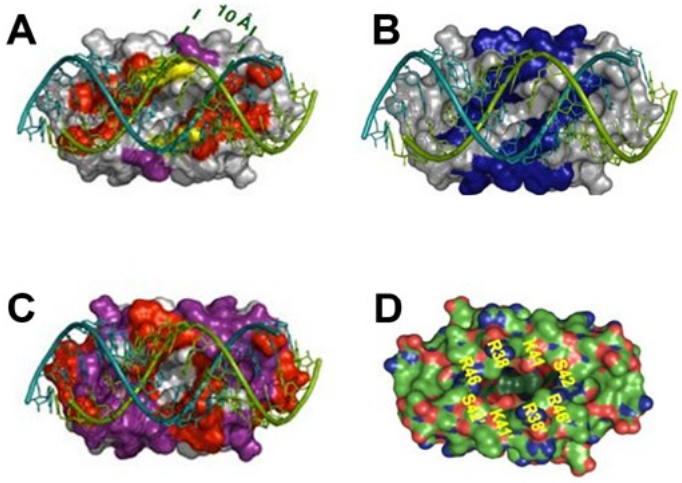


Fig. 1. Model of the NS1A(1-73):dsRNA complex. (A) Conserved residues correspond to dsRNA binding interface. Key RNA binding residues Arg_{38} (yellow) and Lys_{41} (violet) are highlighted, along with other residues of the conserved tracks (red) (B) Basic Arg and Lys residues line the RNA binding "track". Basic Lys and Arg residues are indicated (blue) (C) dsRNA induced NMR chemical shift perturbation in NS1A (1-73). $\Delta\delta_{avg} < 50$Hz (gray), 50Hz $\leq \Delta\delta_{avg} <$ 75Hz (violet), $\Delta\delta_{avg} \geq 75$ Hz (red). (D) Deep cavity of NS1 RNA binding site is a target for antiviral drug design. Heavy atoms are shown color coded by atom type: C, green; N, blue; O, red.

Chien, C. Y.; Tejero, R.; Huang, Y.; Zimmerman, D. E.; Rios, C. B.; Krug, R. M.; Montelione, G. T. Nat Struct Biol 1997, 4:891-895 A novel RNA-binding motif in influenza A virus non-structural protein 1.

Chien, C. Y.; Xu, Y.; Xiao, R.; Aramini, J. M.; Sahasrabudhe, P. V.; Krug, R. M.; Montelione, G. T. Biochemistry 2004, 43:1950-1962 Biophysical characterization of the complex between double-stranded RNA and the N-terminal domain of the NS1 protein from influenza A virus: evidence for a novel RNA-binding mode.

Liu, J.; Lynch, P. A.; Chien, C. Y.; Montelione, G. T.; Krug, R. M.; Berman, H. M. Nat Struct Biol 1997, 4:896-899 Crystal structure of the unique RNA-binding domain of the influenza virus NS1 protein.

Yin, C.; Khan, J. A.; Swapna, G. V.; Ertekin, A.; Krug, R. M.; Tong, L.; Montelione, G. T. *J Biol Chem* 2007, 282:20584-20592 Conserved surface features form the double-stranded RNA binding site of non-structural protein 1 (NS1) from influenza A and B viruses.

5. Polyamine Synthesis Enzymes Structural and Functional Studies of the Bacillus subtilis PaiA

Collaborators: Carl Porter, Roswell Park Cancer Institute
NESG structures: 1RTY 1SQS 1TM0 2HD3 2NV4 2OYS 1Z6P

PaiA has been implicated in the negative control of sporulation in B. subtilis, as well as production of degradative enzymes (Honjo et al., 1990 J Bacteriol 172, 1783). The NESG Consortium has determined the crystal structure of PaiA in complex with CoA at 1.9 Å resolution. The structure reveals that PaiA is a member of the N-acetyltransferase superfamily of enzymes. Unexpectedly, we observed the binding of an oxidized CoA dimer in the active site of PaiA (Fig. 1A). This unanticipated structural information suggests the substrates of the enzyme could be linear, positively charged compounds (Fig. 1B). In a collaboration with Dr. Porter group at the Roswell Park Cancer Institute, Buffalo, we subsequently demonstrated that PaiA possesses N1-acetyltransferase activity toward polyamine substrates including spermidine and spermine. The Porter group also found that conditional overexpression of PaiA in bacteria results in increased acetylation of endogenous spermidine pools.

Our structural and biochemical analyses indicate that PaiA is a novel N-acetyltransferase capable of acetylating both spermidine and spermine. In this way, the pai operon may function in regulating intracellular polyamine concentrations and/or binding capabilities. In addition to preventing toxicity due to polyamine excess, this function may also serve to regulate expression of certain bacterial gene products such as those involved in sporulation.

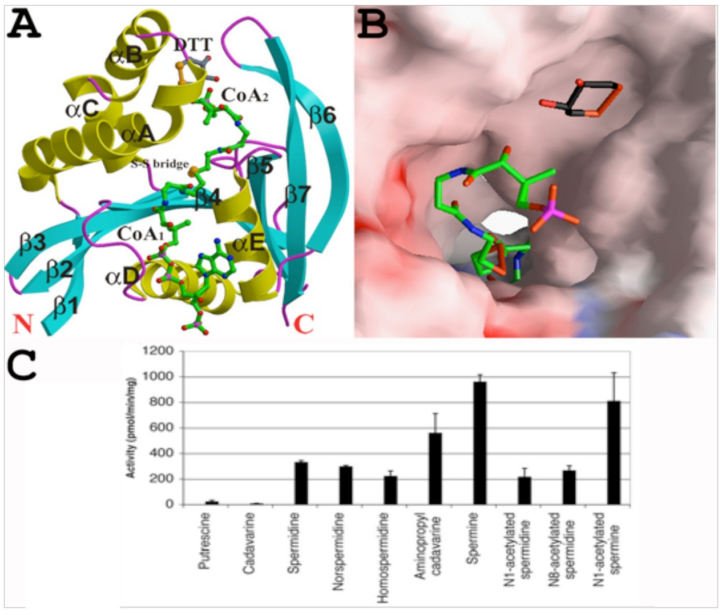


Fig. 1. (A). Structure of PaiA in complex with oxidized CoA dimer (in green for carbon atoms) and DTT (in gray). (B). Molecular surface of the PaiA molecule, showing that the active site of the enzyme is located at the center of the tunnel and viewed from the CoA2 molecule. This view of PaiA also suggests that negatively-charged groups of CoA2 and DTT are incompatible with the predominantly neutral and negatively charged surface of PaiA (C). The catalytic activity of PaiA towards a collection of polyamine and diamine substrates.

Forouhar, F.; Lee, I. S.; Vujcic, J.; Vujcic, S.; Shen, J.; Vorobiev, S. M.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Porter, C. W.; Tong, L. *J Biol Chem* 2005, 280:40328-40336 Structural and functional evidence for bacillus subtilis PaiA as a novel N1-spermidine/spermine acetyltransferase.

Forouhar, F.; Kuzin, A.; Seetharaman, J.; Lee, I.; Zhou, W.; Abashidze, M.; Chen, Y.; Yong, W.; Janjua, H.; Fang, Y.; Wang, D.; Cunningham, K.; Xiao, R.; Acton, T. B.; Pichersky, E.; Klessig, D. F.; Porter, C. W.; Montelione, G. T.; Tong, L. *J Struct Funct Genomics* 2007, 8:37-44 Functional insights from structural genomics.

6. Structural and Functional Studies of Murine Gamma Herpes Virus-68 ORF52

Collaborators: Ren Sun, University of California Los Angeles
Hongyu Deng, Institute of Biophysics, Chinese Academy of Sciences, Beijing
NESG structures: 2OA5

Murine gammaherpesvirus-68 (MHV-68) ORF52 is an abundant, capsid-associated tegument protein. It is well conserved among the gammaherpesviruses, but does not appear to be present in the alpha- and betaherpesviruses. ORF52 is essential for the tegumentation and egress of infectious MHV-68 particles in the cytoplasm. In a collaboration with Dr. Ren Sun at the University of California, Los Angeles and Dr. Hongyu Deng at the Institute of Biophysics, Chinese Academy of Sciences in Beijing, we have carried out structural and functional studies of this important protein. The first results of our studies were published in 2007 (Benach et al., 2007).

As a Community Outreach activity of the NESG, we have determined the crystal structure of ORF52 at 2.1 Å resolution. The structure reveals a dimeric association of this protein, except that the N-terminal α-helix does not obey the symmetry of the dimer (Fig. 1). This helix contains many highly conserved residues, and we hypothesize that it is more likely involved in interactions with other components of the tegument or nucleocapsid of the virus. We have confirmed the self-association of ORF52 by co-immunoprecipitation experiments, and deletion of this N-terminal α-helix abolished the function of ORF52 (Fig. 1). Dr. Deng's laboratory is currently pursuing additional experiments to further elucidate the function of this N-terminal helix as well as the ORF52 protein overall.

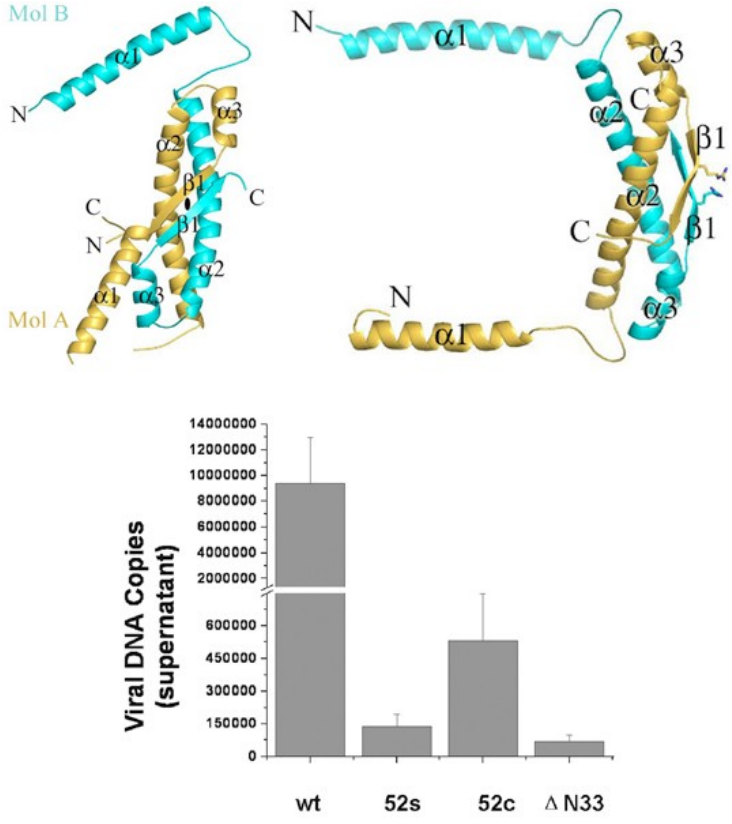


Fig. 1. Structure and function of MHV-68

ORF52. The structure of ORF52 dimer viewed down the two-fold axis (Top left) and from the side (Top right). The Nterminal helix (α1) does not obey the symmetry of the dimer in the crystal (Top left). A model for this helix that obeys the symmetry is shown in the panel at the Top right. Deletion of this helix (ΔN33) abolishes the function of ORF52 (Bottom panel), and the resulting virus is no longer infectious.

Benach, J.; Wang, L.; Chen, Y.; Ho, C. K.; Lee, S.; Seetharaman, J.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Deng, H.; Sun, R.; Tong, L. *J Biol Chem* 2007, 282:31534-31541 Structural and functional studies of the abundant tegument protein ORF52 from murine gammaherpesvirus 68.

7. Structural and Functional Studies of Tryptophan 2,3-Dioxygenase

Collaborators: Stephen Chapman, University of Edinburgh
NESG structures: 2NW7 2NWB 2NW8 2NW9

Tryptophan 2,3-dioxygenase (TDO) and indoleamine 2,3-dioxygenase (IDO) catalyze the oxidative cleavage of the L-tryptophan (L-Trp) pyrrole ring, the first and rate-limiting step in L-Trp catabolism through the kynurenine pathway (Fig. 1A). Both enzymes constitute an important, yet poorly understood, family of heme-containing enzymes. TDO is a homotetrameric enzyme and is highly specific for L-Trp and related derivatives such as 6-fluoro-Trp as the substrate. In comparison, IDO is monomeric, and shows activity toward a larger collection of substrates, including L-Trp, D-Trp, serotonin, and tryptamine. In a collaboration with Chapman group at the University of Edinburgh, United Kingdom, we succeeded in conducting extensive structural and biochemical studies of the Xanthomonas campestris TDO, including the structure at 1.6-A resolution of the catalytically active, ferrous form of TDO in a binary complex with the substrate L-Trp (Fig. 1B&C). The carboxylate (Fig. 1D) and ammonium moieties (Fig.1E) of tryptophan are recognized by electrostatic and hydrogen-bonding interactions with the enzyme and a propionate group of the heme, thus defining the L-stereospecificity (Fig. 1B). A second, possibly allosteric, L-Trp-binding site is present at the tetramer interface (Fig. 1C). The sixth coordination site of the heme-iron is vacant, providing a dioxygenbinding site that would also involve interactions with the ammonium moiety of L-Trp and the amide nitrogen of a glycine residue. The indole ring is positioned correctly for oxygenation at the C2 and C3 atoms. The active site is fully formed only in the binary complex, and biochemical experiments confirm this induced-fit behavior of the enzyme. The active site is completely devoid of water during catalysis, which is supported by our electrochemical studies showing significant stabilization of the enzyme upon substrate binding.

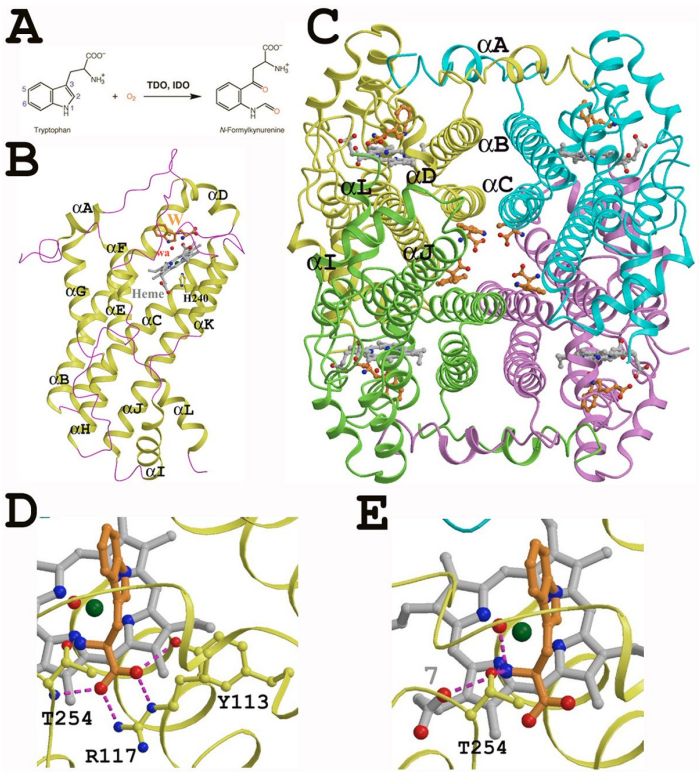


Fig. 1. (A). Reaction catalyzed by TDO & IDO. (B). Structure of TDO in complex with heme (in gray for carbon atoms) and its natural substrate L-Trp (in orange). (C). The tetramer is the biological unit of TDO. The four protomers are shown in yellow, cyan, green and magenta. Four tryptophan molecules (in orange) at the interface of the TDO tetramer possibly have an allosteric role in TDO activity. (D). Recognition of carboxyl group of LTrp by the backbone amide of T254, and invariant residues Y113 and R117. The ferrous ion of the heme and a water molecule in the active site are shown in dark green and red solid spheres. (E). Recognition of ammonium ion of L-Trp by invariant residue T254 and propionate group of heme.

Forouhar, F.; Anderson, J. L.; Mowat, C. G.; Vorobiev, S. M.; Hussain, A.; Abashidze, M.; Bruckmann, C.; Thackray, S. J.; Seetharaman, J.; Tucker, T.; Xiao, R.; Ma, L. C.; Zhao, L.; Acton, T. B.; Montelione, G. T.; Chapman, S. K.; Tong, L. *Proc Natl Acad Sci U S A* 2007, 104:473-478 Molecular insights into substrate recognition and catalysis by tryptophan 2,3-dioxygenase.

8. The Mammallian Muscle Tropomyosin Junction

Collaborators: Sarah Hitchcock-DeGregori, and Norma Greenfield, Robert Wood Johnson Medical School, UMDNJ

NESG structures: 2G9J 1MV4 1IHQ

Tropomyosin is a coiled-coil protein that binds along the length of actin filaments in eukaryotic cells. Tropomyosin cooperatively regulates actin's interaction with myosin and in striated muscle mediates calcium-dependent regulation of contraction by troponin. Critical to these activities is tropomyosin's N and C-terminal "overlap" complex. Working in collaboration with Dr. Hitchcock-DeGregori and Dr. Greenfield, the NESG was able to determine the solution NMR structure of an overlap complex of model peptides of Tropomyosin (Greenfield *et al.*, 2006). In the complex (Fig.1 A&B), the chains of the C-terminal coiled-coil (CTD) spread apart so as to allow a 90° insertion of the N-terminal coiled-coil (NTD), relative to the plane of the CTD. Furthermore, the tertiary structures (but not the helical content) of both the NTD and CTD change upon complex formation (Fig. 1C). This is consistent with the observation that there are only small chemical shift differences between the resonances of the free and bound molecules. A consequence of the geometry is that the orientation of postulated periodic actin binding sites on the coiled-coil surface is retained from one molecule to the next along the actin filament. During this study NMR relaxation data indicated a high degree of flexibility at the junction, which may function to optimize actin filament binding and allow mobility of tropomyosin on the filament surface as it switches between regulatory states.

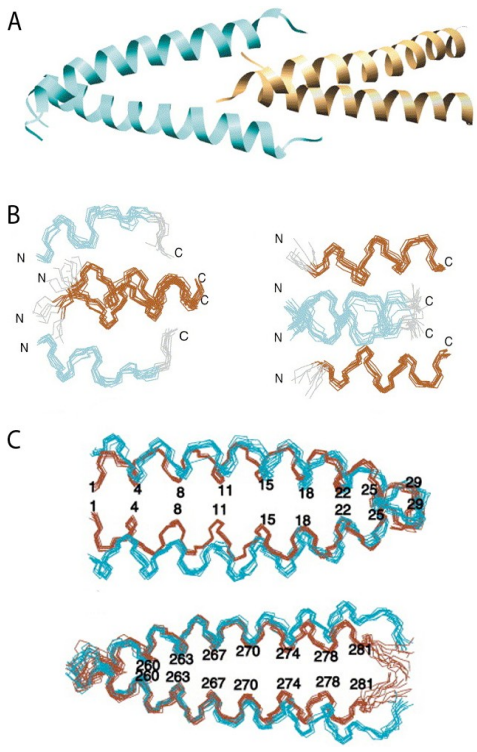


Fig. 1. (A) Ribbon drawing of a representative NMR structure of the entire peptide complex of the NTD (terra cotta) with the CTD (cyan). (B) Orthogonal views of the backbone bonds of the 11 overlapping residues in the NTD–CTD complex. (C) The conformation of NTD (cyan) in the complex with CTD compared to the conformation of unbound NTD (terra cotta). (D) The conformation of the CTD (cyan) bound to NTD compared to unbound CTD (terra cotta).

Greenfield, N. J.; Huang, Y. J.; Swapna, G. V.; Bhattacharya, A.; Rapp, B.; Singh, A.; Montelione, G. T.; Hitchcock-DeGregori, S. E. *J Mol Biol* 2006, 364:80-96 Solution NMR structure of the junction between tropomyosin molecules: implications for actin binding and regulation.

9. CS-Rosetta Project

Collaborators: David Baker, University of Washington, WA; Ad Bax, National Institutes of Health

Rosetta is one of the most successful *de novo* structure prediction programs for obtaining atomic level 3D structures of small proteins (Bradley *et al.*, 2005). Working together with Dr. David Baker and Dr. Ad Bax, the NESG found that the structural information contained in experimentally determined NMR chemical shifts greatly improved the structural accuracy of Rosetta (Shen *et al.*, 2008).

Working with Dr. David Baker, the NESG then developed CS-DP-Rosetta (Raman *et al.*, 2010a), a variant of CS-Rosetta which uses unassigned NOESY peak lists to direct the CS-Rosetta trajectory (Raman *et al.*, 2010a). This method is dependent on the use of the DP score (Huang *et al.*, 2005), a statistical measure of how well a query structure "fits" experimentally derived NOESY data. CS-DP-Rosetta is a fully automated NOESY data analysis process, providing accurate structures for targets up to 150 residues, including many for which CS-Rosetta fails. More recently we demonstrated that the inclusion of residual dipolar coupling information (CS-RDC-Rosetta) and sparse backbone H^N-H^N NOE-derived constraints on perdeuterated proteins enables an even more robust and general approach to guiding Rosetta to the native structure of ever-larger proteins (25-50 kDa) (Raman *et al.*, 2010b).

> The primary obstacle to Rosetta structure prediction from amino acid sequence information alone is conformational sampling; native structures almost always have lower energies than non-native conformations, but they are very seldom sampled in unbiased trajectories. Incorporating NMR chemical-shift information in the selection of the fragments used in the exploration phase (CS–Rosetta) provides a robust approach to determining accurate structures of small (<100-residue) proteins using only backbone and $^{13}C\beta$ chemical-shift data. For larger (>12-kD) proteins, the performance of CS-Rosetta is very target-dependent: Structures sufficiently close to the native structure for the energy to drop substantially may be generated rarely or not at all. Incorporation of RDCs dramatically improved convergence on the [native] structure … for proteins up to 120 residues and, in favorable cases, for larger proteins (Raman *et al.*, 2010b).

It should be noted that none of these methods require sidechain resonance assignments. The process of obtaining and assigning sidechain resonances is very time consuming and error prone for larger proteins (>20 kDa). Hence there is a natural marriage between sparse constraint data that can be obtained for larger perdeuterated proteins and the Rosetta approach. The rapid progress in this collaboration over the last two years demonstrates the powerful synergy between the NESG NMR technology platform and the Rosetta methods developed in the Baker laboratory.

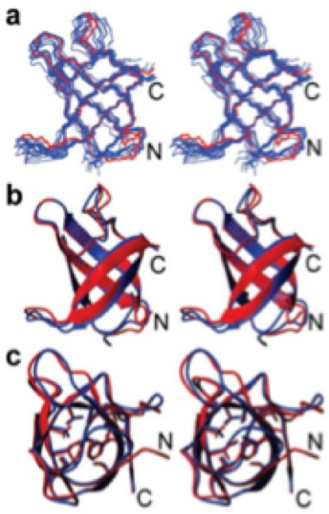


Fig. 1. Fully automated resonance assignments on a perdeuterated protein followed by CS-Rosetta. Stereoview of the superimposition of the CS-Rosetta structure for [2H,13C,15N]- enriched CspA (blue) with the 2.0 Å X-ray crystal structure of CspA (red) (pdb ID: 1mjc). (a) Backbone line representations of the 10 lowest energy conformers obtained from CS-Rosetta structure compared with the X-ray crystal structure. (b) Ribbon diagram of the lowest energy conformer versus the X-ray crystal structure. (c) The packing of the core hydrophobic residues (Y. Tang & G. Montelione, in preparation).

Bradley, P.; Misura, K. M.; Baker, D. *Science* 2005, 309:1868-1871 Toward high-resolution de novo structure prediction for small proteins.

Huang, Y. J.; Powers, R.; Montelione, G. T. *J Am Chem Soc* 2005, 127:1665-1674 Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics.

Raman, S.; Huang, Y. J.; Mao, B.; Rossi, P.; Aramini, J. M.; Liu, G.; Montelione, G. T.; Baker, D. *J Am Chem Soc* 2010a, 132:202-207 Accurate automated protein NMR structure determination using unassigned NOESY data.

Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D. *Science* 2010b, 327:1014-1018 NMR structure determination for larger proteins using backbone-only data.

Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc Natl Acad Sci U S A* 2008, 105:4685-4690 Consistent blind protein structure generation from NMR chemical shift data.

10. CASD-NMR Project

Collaborators: Numerous, Global Effort (see Rosato citation)
NESG structures:

The Critical Assessment of Automated Structure Determination of Proteins from NMR data (CASD-NMR) is an open and rolling collaborative experiment involving developers of software tools / protocols for the automated calculation of protein structures from NMR data (Rosato *et al.*, 2009) . The goal of CASD-NMR is to help advance the relevant methodology to the level of quality and reliability required for direct PDB deposition of unsupervised results. Participants include David Baker (CS-Rosetta), Peter Guentert (CYANA), Torsten Herrman (UNIO), Michael Nilges (ARIA), Michele Vendruscolo (Cheshire) and Gaetano Montelione (AutoStructure) of the NESG. The gateway to CASD-NMR is managed by e-NMR (http://www.enmr.eu/CASD-NMR).

The NESG is a key participant in CASD-NMR by (i) contributing to the project design, (ii) providing a large number of 'blind' NMR data sets (approximately one data set per month), and (iii) by developing and testing improved methods for the automated analysis of NMR assignments and structures.

The new methodologies that the NESG is testing during the CASD-NMR project include consensus approaches that combine several different protocols for automated structure determination and the use of DP scores (Huang *et al.*, 2005) for assessing 3D models against unassigned NOESY peak lists. DP scores are also being used to filter structures generated from the most unambiguously-assigned NOESY peaks, structures which are then used to guide the AutoStructure NOESY peak assignment process (Huang *et al.*, 2006). DP scores are currently being used to guide the chemical shift based protein structure prediction program CS-Rosetta and CS-RDC-Rosetta, a subsequent version incorporating bond vector constraint information (Raman *et al.*, 2009a; Raman *et al.*, 2009b).

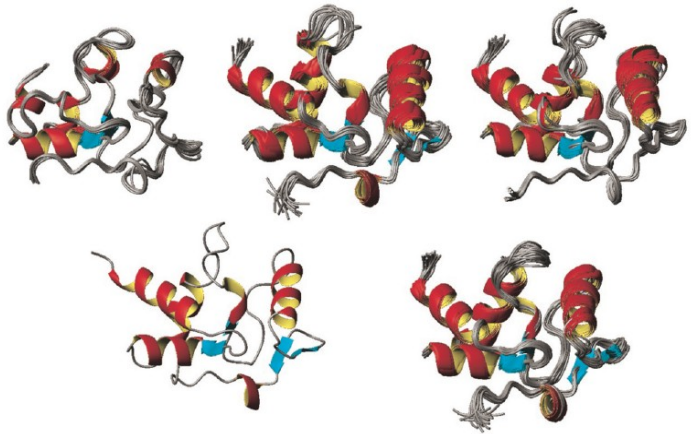


Fig 1. The results of fully automated calculations by various programs for one of the blind test data sets of the 2009 Florence workshop are compared to the reference structure (bottom right) determined by Aramini et al. (PDB ID 2kif).

Huang, Y. J.; Tejero, R.; Powers, R.; Montelione, G. T. Proteins 2006, 62:587-603 A topology-constrained distance network algorithm for protein structure determination from NOESY data.

Raman, S.; Huang, Y. J.; Mao, B.; Rossi, P.; Aramini, J. M.; Liu, G.; Montelione, G. T.; Baker, D. *J. Am. Chem. Society* 2009, In Press.:Accurate automated

protein NMR structure determinatino using unassigned NOESY data.

Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Prestegard, J. H.; Montelione, G. T.; Baker, D. 2009, Submitted:NMR structure determinatin for larger proteins using backbone-only data.

Rosato, A.; Bagaria, A.; Baker, D.; Bardiaux, B.; Cavalli, A.; Doreleijers, J. F.; Giachetti, A.; Guerry, P.; Guntert, P.; Herrmann, T.; Huang, Y. J.; Jonker, H. R.; Mao, B.; Malliavin, T. E.; Montelione, G. T.; Nilges, M.; Raman, S.; van der Schot, G.; Vranken, W. F.; Vuister, G. W.; Bonvin, A. M. *Nat Methods* 2009, 6:625-626 CASD-NMR: critical assessment of automated structure determination by NMR.

11. Quantum Mechanics-Derived $^{13}C\alpha$ Chemical Shifts for Protein NMR Structure Determination, Refinement and Validation

Collaborators: Harold Sheraga, Cornell University

A burgeoning repository of protein chemical shift data has become widely available to the scientific community in recent years (www.bmrb.wisc.edu). This database is quite valuable, since it has been long recognized that the observed chemical shift of a nucleus is highly sensitive to its chemical environment, and can be used directly in structure refinement protocols. In particular, the $^{13}C\alpha$ chemical shift for a given amino acid is almost completely determined by its backbone and side chain torsional angles, and is not correlated with neighboring residues in the sequence. Scheraga and co-workers have demonstrated that for a given conformation, 13Cα chemical shifts for each amino acid in a protein can be computed at the density functional level of theory (DFT) by treating each as a terminally blocked tripeptide (Ac-GXG-NMe). Although computationally intensive, the approach offers a number of advantages including i) $^{13}C\alpha$ chemical shifts can be calculated with high accuracy, ii) there is no need for *a priori* knowledge of the oligomerization state of the protein, and iii) no knowledge-based information is required. They developed a new, purely physics-based structure validation metric called the conformationally averaged root-mean-square-deviation, ca-rmsd (Vila *et al*., 2007, J. Biomol. NMR, *38*, 221), which reflects the agreement between experimental and computed $^{13}C\alpha$ chemical shifts for all residues in an ensemble of protein structures (Eq. 1),

$$ca-rmsd^{\alpha}=\left[(1/N)\sum_{\mu=1}^{N}(\Delta_{\mu}^{\alpha})^2\right]^{1/2} \quad (1)$$

where $1 \leq \mu \geq N$, with N being the number of observed 13Cα chemical shifts, and $\Delta_{\mu}^{\alpha} = (^{13}C_{obs,\mu} - \langle ^{13}C_{comp}\rangle_{\mu}$.

For a single conformer (i.e., an X-ray structure) ca-rmsd is equivalent to the $^{13}C\alpha$ rmsd (Eq.2).

$$ca-rmsd^{\alpha}=rmsd^{\alpha}=\left[(1/N)\sum_{\mu=1}^{N}(^{13}C^{\alpha}_{observed,\mu}-{}^{13}C^{\alpha}_{computed,\mu})^2\right]^{1/2} \quad (2)$$

In the NESG project, this physics-based approach has been applied to validate the NMR and X-ray structures of the YnzC protein from Bacillus subtilis, where it was demonstrated that ensembles of structures derived from NMR data are a better representation of the $^{13}C\alpha$ chemical shifts than the crystal structure (Vila et al., 2008).

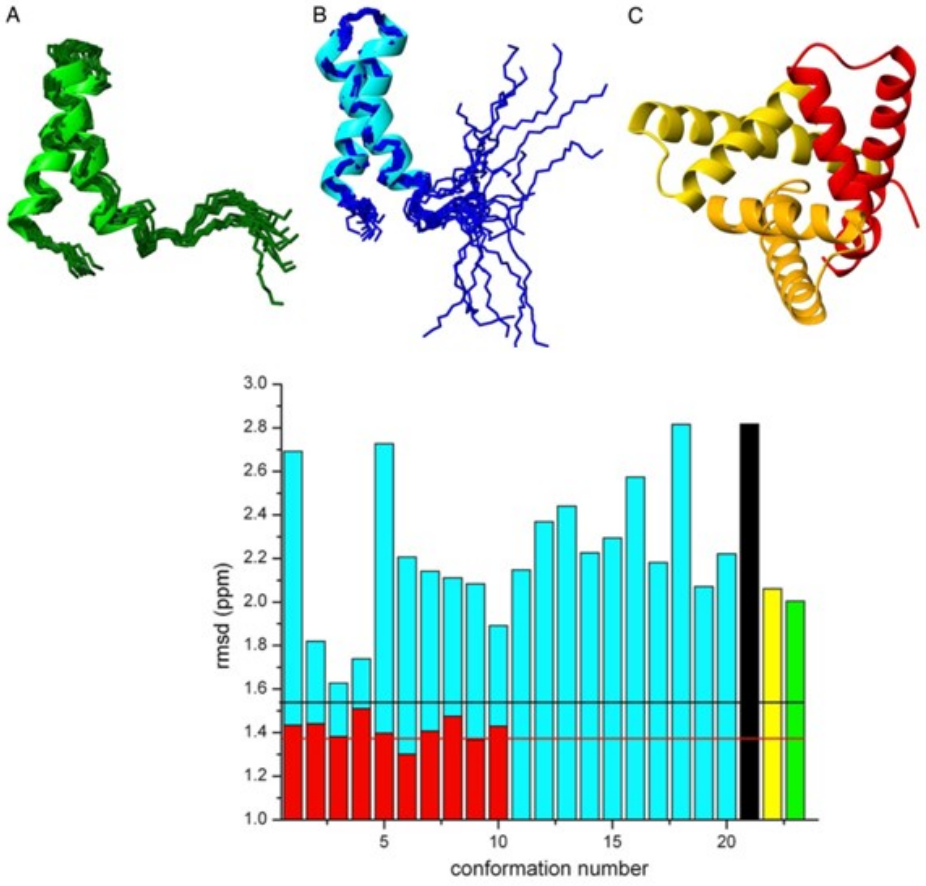


Fig.1. Validation of NMR derived and X-ray structures of B. subtilis YnzC using quantum-mechanics-derived ^{13}Cα chemical shifts.Left: (A) 10 structures calculated using NMR-derived NOE constraints and torsional angles computed from ^{13}Cα chemical shifts using DFT, (B) the final ensemble of 20 NMR structures (PDB entry, 2JVD), and (C) the 2.0 Å X-ray structure which features three monomers in the unit cell (PDB entry, 3BHP). Bottom: Plot of rmsd between computed and observed ^{13}Cα chemical shifts for each of the conformers derived from A (red), B (cyan) and C (black, yellow, and green). Horizontal lines denote the ca-rmsds for the NMR-derived structures, A (red) and B (black).

Recently, Scheraga and co-workers introduced a new quantum-mechanics-derived CheShift server (http://cheshift.com) for computing 13Cα chemical shifts of proteins based on their structure (Vila *et al*., 2009, Proc. Natl. Acad. Sci. USA). Internally, the server consults a comprehensive library of 13Cα chemical shifts computed at the DFT level of theory using a smaller basis set (for increased speed) for all 20 naturally occurring amino acids from almost 700,000 conformations generated as a function of backbone (φ,ψ,ω) and side chain (χ1,χ2) torsional angles. The server is being used in NESG NMR projects to evaluate or rank protein models (“decoys”) obtained in intermediate stages of automated protein NMR structure refinement and/or automated NOESY assignment process, and for structure validation.

References

Vila, J. A.; Villegas, M. E.; Baldoni, H. A.; Scheraga, H. A. *J Biomol NMR* 2007, 38:221-235 Predicting 13C(alpha) chemical shifts for validation of protein structures.

Vila, J. A.; Aramini, J. M.; Rossi, P.; Kuzin, A.; Su, M.; Seetharaman, J.; Xiao, R.; Tong, L.; Montelione, G. T.; Scheraga, H. A. *Proc Natl Acad Sci USA* 2008, 105:14389-14394 Quantum chemical 13Calpha chemical shift calculations for protein NMR structure determination, refinement, and validation.

Vila, J. A.; Arnautova, Y. A.; Martin, O. A.; Scheraga, H. A. *Proc Natl Acad Sci U S A* 2009, 106:16972-16977 Quantum-mechanics-derived 13Calpha chemical shift server (CheShift) for protein structure validation.

12. Structural and Functional Characterization of RimO from Thermotoga maritima, a Radical-SAM methylthiotransferase

Collaborators: Marck Fontecave, Collége de France and iRTSV-LCBM, France; Mohamed Atta, iRTSV-LCBM, France
NESG structures: 2QGQ

Post translational modifications of ribosomal proteins are important for the accuracy of the decoding machinery. A recent *in vivo* study has shown that the *rimO* gene was involved in generation of the 3-methylthio derivative of residue asp-89 in ribosomal protein S12 (Anton *et al.*, 2008, PNAS 105:1826). This reaction is formally identical to that catalyzed by MiaB on the C2 of Adenosine-37 near the anticodon of several tRNAs. We herein present spectroscopic evidences that *T. maritima* RimO, like MiaB, contains two [4Fe-4S] centers, one presumably bound to three invariant cysteines in the central Radical-SAM domain and the other to three invariant cysteines in the N-terminal UPF0004 domain. We demonstrate that *holo* RimO can specifically methylthiolate the aspartate residue of a 20-mer peptide derived from S12, yielding a mixture of mono- and *bis*-methylthio derivatives. Finally, we present the 2.0 Å crystal structure of the central Radical-SAM and the C-terminal TRAM domains in *apo* RimO. While the core of the open TIM-barrel of the Radical-SAM domain is conserved, RimO shows differences in domain organization compared to other Radical-SAM enzymes. The unusually acidic TRAM domain, likely to bind the basic S12 protein, is located at the distal edge of the Radical-SAM domain. The basic S12 protein substrate is likely to bind RimO through interactions with both the TRAM domain and the concave surface of the incomplete TIM-barrel. These biophysical results provide a foundation for understanding the mechanism of methylthioation by Radical-SAM enzymes in the MiaB/RimO family.

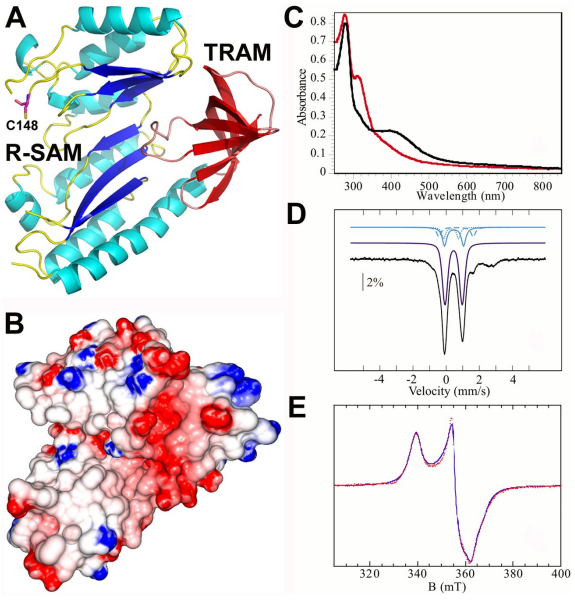


Fig. 1. (A) Cartoon diagram with the Radical-SAM (R-SAM) domain colored cyan, blue, and yellow and the TRAM domain colored red and orange. Invariant residue Cys-148, which probably ligates one [4Fe-4S] cluster in *holo* RimO, is shown in ball-and-stick representation. (B) Electrostatic surface potential of RimO, with blue and red representing acidic and basic regions, respectively. (C). UV-visible absorption spectra of oxidized (black) and reduced (red) forms of *holo* RimO (8 µM) in 50 mM NaCl, 50 mM Tris-HCl, pH 8. The absorbance at 314 nm is due to excess dithionite. (D). Mössbauer spectrum of *holo* RimO (288 µM) in the same buffer. The dark blue line represents a typical [4Fe-4S]2+ cluster, while the light blue lines represent the three distinct sites of an atypical [4Fe-4S]2+ cluster, composed of a delocalized $Fe^{II}Fe^{III}$ pair (solid line), and two distinct Fe centers, one being more ferrous in character (hatched line) than the other one (dotted line). (E) X-band

EPR spectrum of reduced *holo* RimO (144 μM) in the same buffer. The dotted red line shows the g = 2 region of the X-band EPR spectrum recorded on a dithionite reduced sample (natural ^{57}Fe abundance) at 20 K under non saturating conditions (0.63 mW microwave power at 9.65 GHz frequency). The blue line shows a theoretical simulation with two S = 1/2 species in a 51:49 ratio using a Lorentzian profile.

Arragain, S.; Garcia-Serres, R.; Blondin, G.; Douki, T.; Clemancey, M.; Latour, J. M.; Forouhar, F.; Neely, H.; Montelione, G. T.; Hunt, J. F.; Mulliez, E.; Fontecave, M.; Atta, M. *J Biol Chem* 2010, 285:5792-5801 Post-translational modification of ribosomal proteins: Structural and functional characterization of RimO from *Thermotoga maritima*, a radical S-adenosylmethionine methylthiotransferase.

13. Alg13 Adopts a Unique Topology Among Glycosyltransferases

Collaborators: Barbara Imperiali, Massachusetts Institute of Technology
NESG structures: 2JZC
Text Source: Gaetano Montelione, Rutgers University, NESG
PSI-KB Center Cut
http://kb.psi-structuralgenomics.org/update/2009/05/full/rhighlts_psisgkb.2009.21.html
Image Source: Wang et al., 2008 (see reference below).

The study presents the solution structure of Alg13, the glycosyl donor-binding domain of an important bipartite glycosyltransferase in the yeast *Saccharomyces cerevisiae*. This glycosyltransferase is unusual in that it is active only in the presence of a binding partner, Alg14. Alg13 is found to adopt a unique topology among glycosyltransferases. Rather than the conventional Rossmann fold found in all GT-B enzymes, the N-terminal half of the protein is a Rossmann-like fold with a mixed parallel and antiparallel β-sheet. The Rossmann fold of the C-terminal half of Alg13 is conserved. However, although conventional GT-B enzymes usually possess three helices at the C terminus, only two helices are present in Alg13. Titration of Alg13 with both UDP-GlcNAc, the native glycosyl donor, and a paramagnetic mimic, UDP-TEMPO, shows that the interaction of Alg13 with the sugar donor is primarily through the residues in the C-terminal half of the protein.

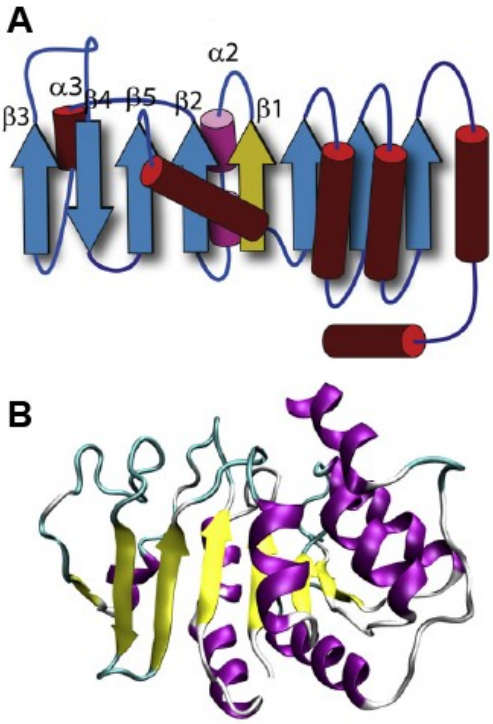


Fig. 1. Topology (A) and ribbon diagram (B) representations of experimentally determined Alg13.

Wang, X.; Weldeghiorghis, T.; Zhang, G.; Imperiali, B.; Prestegard, J. H. *Structure* 2008, 16:965-975 Solution structure of Alg13: the sugar donor subunit of a yeast N-acetylglucosamine transferase.

14. Structural Basis of O6-alkylguanine Recognition by a Bacterial Alkyltransferase-like DNA Repair Protein

Collaborators: John Tainer, Scripps Research Institute, La Jolla, California
NESG structures: 2KIF

Alkyltransferase-like proteins (ATLs) are a novel class of DNA repair proteins related to O^6-alkylguanine-DNA alkyltrasferases (AGTs). Like AGTs, ATLs tightly bind alkylated DNA, but unlike AGTs, lack alkytransferase activity. Instead ATLs shunt the damaged DNA into the nucleotide excision repair pathway. O^6 methylated guanine is highly mutagenic; during replication O^6 methylated guanine can cause a switch to thymine incorporation over cytosine.

The NESG determined the first structure of a bacterial ATL from *Vibrio parahaemolyticus* (vpAtl) (Aramini *et al.*, 2010). The structure of vpAtl was found to adopt an AGT-like fold comprised of five α-helices flanking two antiparallel β-strands (Fig.1). vpAtl is capable of tightly binding to O^6-methylguanine DNA and disrupting repair by AGTs (Fig.1B). As illustrated in Fig.1A, the structure of vpAtl is highly similar to DNA-bound *Schizosaccharomyces pombe* ATL (spAtl1) (< 2 Å r.m.s.d, C^α superposition). ATL therefore likely binds the minor groove of DNA in a sequence independent manner analogous to the DNA binding by AGT and spAtl1. Mutation of two highly conserved residues Tyr23 and Arg37 disrupt O^6-methylguanine DNA binding (Fig.1C&D). These two highly conserved residues in spAtl1 were demonstrated to directly interact with the damaged (allkylated) guanine and the orphaned cytosine. NMR relaxation data reveal conformational plasiticity in the guanine-lesion recognition cavity. This flexibility may confer broader specificity for various alkyl guanine lesions.

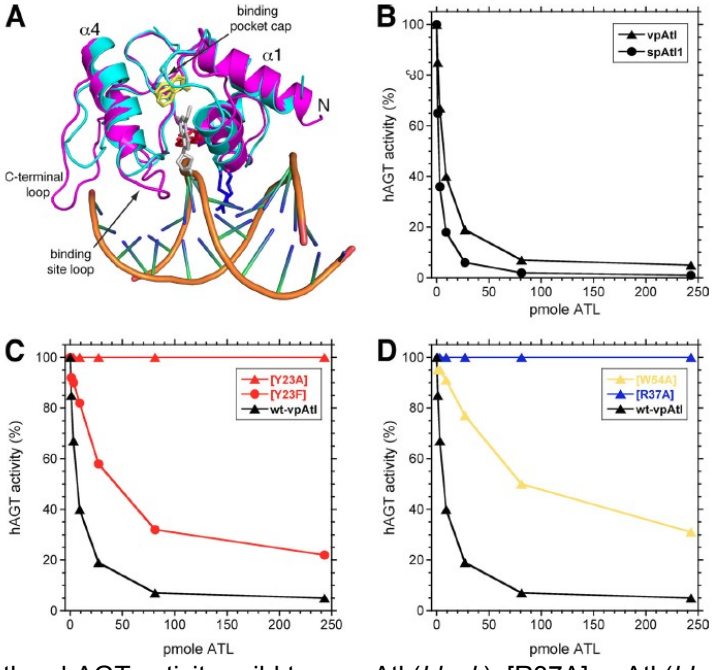


Fig.1 Conserved structure and DNAbinding properties of vpAtl. (A) overlay of the vpAtl structure (*cyan*) and the crystal structure of spAtl1 bound to *O*6-mG-DNA (PDB ID: 3GX4; *magenta*). The sidechains of Tyr23, Arg37, and Trp54 are shown in *red, blue,* and *yellow*, respectively, and the *O*6-mG in the bound spAtl1 structure is shown in *gray*. (B) percent activity of hAGT as a function of ATL concentration for vpAtl (*triangles*) and spAtl1 (*circles*). (C) effect of mutating Tyr23 in vpAtl on hAGT activity; wild type vpAtl (*black*), [Y23A]-vpAtl (*red triangles*), and [Y23F]-vpAtl (*red circles*). (D) effect of mutating Arg37 and Trp54 in vpAtl on hAGT activity; wild type vpAtl (*black*), [R37A]-vpAtl (*blue*), and [W54A]-vpAtl (*gold*).

Aramini, J. M.; Tubbs, J. L.; Kanugula, S.; Rossi, P.; Ertekin, A.; Maglaqui, M.; Hamilton, K.; Ciccosanti, C. T.; Jiang, M.; Xiao, R.; Soong, T. T.; Rost, B.; Acton,

T. B.; Everett, J. K.; Pegg, A. E.; Tainer, J. A.; Montelione, G. T. *J Biol Chem* 2010, Structural basis of O6-alkylguanine recognition by a bacterial alkyltransferase-like DNA repair protein.