Detailed Compilation of Scientific Accomplishments
Dinshaw J. Patel Laboratory

Last Decade (2003 to present) Time Frame

Earlier Decades (1969 to 2003) Time Frame

The web site address of the Patel lab is http://www.mskcc.org/mskcc/html/10829.cfm. Importantly, color images of the structures of systems of interest spanning the last decade of research (topics 1 to 7) can be found by accessing the web site, thereby matching text in this detailed overview with structures presented in the web site.

Last Updated: Dec 1, 2013

1. RIBOSWITCHES AND RIBOZYMES

(2004 to present)

Contributions

The role of RNA in information transfer and catalysis highlights its dual functionalities. RNAs can adopt distinct scaffolds for ligand recognition and catalysis, exhibiting tunable specificities and enantiomeric selectivities. Metabolite-sensing non-coding RNAs, or riboswitches, specifically interact with small ligands and direct expression of the genes involved in their metabolism. Riboswitches are composed of metabolite-sensing modules, capable of ligand-induced structural changes, and adjacent downstream modules, harboring expression-controlling elements. We have focused our regulatory RNA research to metabolite-sensing domains of riboswitches, given that the resulting structural information will be critical for our understanding of their contribution to modulation of gene expression levels and metabolic homeostasis.

1.1 RIBOSWITCHES

Purine Riboswitches. A key unanswered question at the beginning of our research was how riboswitch RNAs containing helical imperfections such as internal loops zippered up to form binding pockets for small ligands despite containing only four nucleotides. A follow up question relates to the molecular principles underlying recognition of cognate ligands and discrimination against closely related analogs. Our comparative x-ray structural study of guanine and adenine riboswitch-sensing domains (Serganov et al. *Chem. Biol.* 2004; see also Batey et al. *Nature* 2004) addressed these issues by demonstrating that the three-helical junctional architecture of purine riboswitches in the bound state involved zippering up of junctional elements through maximization of stacking and hydrogen-bonding interactions mediated by base triple and tetrad formation. In addition, parallel alignment of helical stems facilitated by complementary loop-loop interactions impacted at a distance on the junctional architecture. Strikingly, the bound purines were fully encapsulated in the binding pocket by stacking and hydrogen-bonding interactions, but the selectivity for guanine versus adenine recognition was dependent solely on Watson-Crick pairing of the bound purines with a single pyrimidine (G with C, and A with T) in the riboswitch. These results raised the challenging question as to whether details of ligand recognition by riboswitches could facilitate structure-based redesign of binding pockets for recognition by ligand analogs. Comparison of structures of guanine- and 2'-deoxyguanosine-bound ribowitches (Serganov et al. *Nat. Chem. Biol.* 2011) established that re-engineering riboswitch scaffolds required consideration of selectivity features dispersed throughout the riboswitch tertiary fold and not just restricted to the binding pocket.

Phosphate-containing Metabolite Riboswitches. An intriguing question relates to how riboswitches with their negatively-charged polynucleotide RNA backbones are capable of binding metabolites containing negatively-charged phosphates. This challenge was addressed by our group by undertaking structural studies on the sensing domains of thiamine pyrophosphate (TPP) and flavin mononucleotide (FMN) riboswitches. The x-ray based structure of the TPP riboswitch was unique for two striking observations (Serganov et al**.** *Nature* 2006; see also Thore et al. *Science* 2006**).** The first highlighted the discovery that TPP bound the riboswtich in an extended conformation with its aromatic heterocyclic ring and diphosphate ends anchored in separate pockets on adjacently-positioned helices generated by the tuning fork-like overall architecture in the bound state. Strikingly, recognition of the diphosphate of TPP was mediated by a pair of Mg^{2+} cations that coordinated the phosphates of TPP, but not those of the RNA, but instead were hydrogen-bonded to the base edges of the RNA. The FMN riboswitch used the same principles for Mg^{2+}-mediated recognition of the monophosphate of FMN (Serganov et al. *Nature* 2009), but exhibited an overall butterfly-like topology, stapled together by opposingly-directed but nearly identical folded peripheral domains, with the FMN bound at a zippered up six-helical junctional site.

Amino Acid Riboswitches. Though hydrated Mg^{2+} cations with their octahedral coordination geometry have long been known to be important both for charge neutralization of closely juxtaposed phosphates in higher-order RNA folds and for their role in RNA catalysis, much less is understood regarding the role of monovalent cations in RNA folding and catalysis. Our x-ray structure of the sensing domain of the five-

helical junction lysine riboswitch in the free and bound state (Serganov et al. *Nature* 2008; see also Garst et al. *J. Biol. Chem*. 2008) established the contribution of a non-octahedrally-coordinated bridging K^+ cation in mediating a hydrogen bond recognition network between the backbone amide functionality of the bound lysine and recognition elements on the RNA. At the opposite end of the bound lysine, the ε-ammonium group used two bridging water molecules as part of its intermolecular recognition network, thereby allowing replacement of the ε-ammonium group by a bulky guanidinium functionality, which could be readily accommodated in the binding pocket following expulsion of the bound waters. Selectivity for lysine over other long chain amino acids reflected the contribution of shape-complementarity within the elongated binding pocket, as well as the fixed distance separating RNA elements involved in recognition of the amide and ε-ammonium ends of the bound lysine.

Glycine Riboswitch Composed of Tandem Sensing Domains. The identification of a glycine riboswitch challenged our understanding regarding RNA's ability in generating a small enough pocket capable of recognizing this smallest of amino acids. In addition, the presence of tandem sensing domains within the glycine riboswitch raised an additional challenge, namely the elucidation of the molecular basis underlying the observed cooperativity of glycine complex formation. These challenging issues were addressed by our group following x-ray-based structural studies on the isolated second sensing domain of the glycine riboswitch in the free and lysine-bound states (Huang et al. *Mol. Cell* 2010). Notably, the solution turned out to be generation of a larger binding pocket capable of accommodating not only glycine but also two hydrated Mg^{2+} cations in the bound state, with one of the divalent cations released in the free state. Unexpectedly, two molecules of the second sensing domain of the glycine riboswitch are involved in a packing interaction in the crystal lattice, with intermolecular quarternary contacts mediated by an adenine-rich strand from one monomer forming stacked A-minor base triples with the minor groove of a duplex segment of the other monomer. Thus, glycine binding to one monomer impacts on binding to the other monomer through conformational transitions that are transmitted through inter-monomer contacts, thereby accounting for the observed cooperativity (see also Butler et al. *Chem. Biol*. 2011).

Anion Riboswitches. Recently, a riboswitch has been identified that bound fluoride anion and discriminated against larger halide anions. This striking discovery was all the more intriguing given the small size of the fluoride anion and its negative charge. Our x-ray based structure of the sensing domain of the fluoride riboswitch (Ren et al. *Nature* 2012) highlighted several new folding topologies associated with RNA-mediated molecular recognition, whereby the fluoride anion was coordinated by three surrounding hydrated Mg^{2+} cations, which in turn were coordinated by five inwardly-pointing phosphate oxygens, with these concentric opposingly-charged shells nestled within the junction of a pseudo-knot based RNA topology. Given the unusual topology with inwardly-pointing phosphate groups, tertiary interactions involving a zippered up internal loop were maintained by maximization of base stacking interactions, but at the expense of base pairing interactions.

The above examples highlight the plasticity of RNA to generate higher order architectures, including its ability to generate tailor-made pockets for ligand recognition and discrimination, as well as its capacity to form inter-domain contacts to account for cooperative ligand binding, supplemented by the role of both divalent and monovalent hydrated cations in stabilizing folds and in mediating ligand recognition. Indeed, RNA with just four nucleotides remarkably exhibits a comparable capacity for folding and recognition as proteins with twenty amino acids containing neutral, polar, aromatic and charged residues.

1.2 RIBOZYMES

RNA's ability to catalyze specific biological reactions has resulted in the identification of natural ribozymes that to date have been restricted to the making and breaking of phosphodiester bonds. However, little is known about RNA's ability to catalyze formation of other bond-forming reactions and this led to the identification of an *in vitro* selected ribozyme capable of catalyzing carbon-carbon bond formation with both significant rate enhancement and enantioselectivity. These efforts identified a 49-mer *in vitro*-selected RNA that catalyzed the carbon-carbon bond-forming Diels-Alder reaction between anthracene as the diene and maleimide as the dienophile, that yielded a chiral bicylic product. Our x-ray based structural studies on this *in vitro* selected Diels-Alder ribozyme in the absence and presence of

chiral bicyclic product established that this RNA, which contains three stems, an internal loop and a four base overhang, adopted a λ-shaped nested double pseudoknot fold with a preformed junctional hydrophobic pocket that accommodated the chiral bicyclic Diels-Alder product (Serganov et al. *Nat. Struct. Mol. Biol.* 2005; with the Andres Jaschke lab, Heidelberg University). Stereoselection of the bicyclic product was dictated by shape complementarity, while it's anchoring within the junctional pocket was governed by a combination of intermolecular stacking and a limited number of hydrogen bond interactions. Notably, common molecular recognition principles were found to govern catalysis of Dlels-Alder carbon-carbon bond-forming reactions by both RNA (*in vitro* selected aptamer) and proteins (catalytic antibodies).

Publications: Patel Lab

Huang, L., Serganov, A. & Patel, D. J. (2010). Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. ***Mol. Cell*** 40, 774-786.

Pikovskaya, O., Polonskkaya, A., Patel, D. J. & Serganov, A. (2011). Structural principles of nucleoside selectivity in a 2'-deoxyguanosine. ***Nat. Chem. Biol.*** 7, 748-755.

Ren, A., Rajashankar, K. & Patel, D. J. (2012). Fluoride ion encapsulation by Mg^{2+} and phosphates in a fluoride riboswitch. ***Nature*** 486, 85-89.

Serganov, A., Huang, L. & Patel, D. J. (2008). Structural insights into amino acid binding and gene control by a lysine riboswitch. ***Nature*** 455, 1263-1267.

Serganov, A., Huang, L. & Patel, D. J. (2009). Coenzyme recognition and gene regulation by a FMN riboswitch. ***Nature*** 458, 233-237.

Serganov, A., Keiper, S., Malinina, L., Tereschko, V., Skripkin, E., Hobartner, C., Polonskaia, A., Phan, A. T., Wombacher, R., Micura, R., Dauter, Z., Jaschke, A. & Patel, D. J. (2005). Structural basis for Diels-Alder ribozyme catalyzed carbon-carbon bond formation. ***Nature Struct. Mol. Biol.*** 12, 218-224.

Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. (2006). Structural basis for gene regulation by a riboswitch that senses thiamine pyrophosphate. ***Nature*** 441, 1167-1171.

Serganov, A., Yuan, Y-R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R. & Patel, D. J. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. ***Chem. Biol.*** 11, 1729-1741.

Complementary Publications: Other Labs

Batey, R. T. et al. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. ***Nature*** 18, 411-415.

Butler, E. B. et al. Structural basis of cooperative ligand binding by the glycine riboswitch. ***Chem. Biol***. 18, 293-298.

Garst, A. D. et al. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. ***J. Biol. Chem***. 283, 22347-22351.

Thore, S. et al. (2006). Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. ***Science*** 31, 1208-1211.

2. RNA Silencing Pathways

(2003 to present)

Contributions

RNA silencing, also known as RNA interference, is a conserved biological response to dsRNA that is mediated by siRNAs, which guide the sequence specific degradation of cognate mRNAs. Our long-term goals have been to structurally characterize and mechanistically define molecular recognition events and conformational transitions associated with (1) processing of long dsRNA into siRNAs by the RNase III endonuclease activity of Dicer and (2) guide-strand-mediated cleavage of target mRNAs by Argonaute (Ago), the key component of the RNA-induced silencing complex (RISC), that exhibits RNase H-mediated slicing activity.

2.1 DICER-dsRNA COMPLEXES

The prevalent view of Dicer-mediated cleavage of dsRNAs and micro RNAs (miRNAs) involves cleavage-site selection by measuring a set distance from the 3'-overhang of dsRNA and miRNA termini (3' counting rule). Our goal was to test the generality of these conclusions and explore for alternate mechanisms, should they exist, towards our understanding of Dicer function.

In a collaborative effort championed by the Narry Kim laboratory (Seoul National University), it was established that human Dicer anchors not only the 3'-end (within the PAZ pocket), but also the 5'-phosphorylated end, with the cleavage site determined mainly by the distance (approx. 22-nt) from the 5'-phosphorylated terminus (5' counting rule) (Park et al. *Nature* 2011). Indeed, a basic phosphate-binding pocket (5'-pocket) has been identified within the platform domain in our x-ray-based structure of the hDicer PAZ-platform-connector helix cassette bound to siRNA (Tian et al. *Mol. Cell* 2014), with mutations of the 5'-pocket residues reducing processing efficiency and altering cleavage site selection.

Canonical Dicer's contain a PAZ domain that anchors the end of long dsRNAs, thereby generating siRNAs of a fixed length following RNase III-mediated cleavage, that is dependent on the separation between the PAZ domain and the composite pockets formed by a pair of RNase III domains. Such a mechanism involved processing initiation from the ends and working inwards into the dsRNA. By contrast, budding yeast *K. polysporus* (*Kp*Dicer) lacks a PAZ domain and hence must adopt a non-canonical mechanism for dsRNA cleavage. A combined x-ray structural study by our group and biochemical studies by the David Bartel lab (Whitehead and MIT) established that *Kp*Dicer forms dimers that bind cooperatively along the dsRNA substrate, such that the distance between consecutive active sites determines the length of the siRNA products, thereby initiating processing in the interior and working outwards into the dsRNA (Weinberg et al. *Cell* 2011). This unique mechanism identified for budding yeast Dicers provided insights into the concept of natural protein-based molecular rulers and imparts substrate preference with ramifications for biological function.

2.2 BINARY AND TERNARY ARGONAUTE COMPLEXES WITH GUIDE AND TARGET STRANDS

Our group initially focused on the structural biology of eubacterial Ago's and their complexes with guide and target strands, supplemented by structure-guided biochemical experiments undertaken in the Thomas Tuschl lab (Rockefeller). A major breakthrough was our discovery that eubacterial Ago's preferentially bound 5'-phosphorylated guide DNAs (Yuan et al. *Mol. Cell* 2006) and that diffraction-quality crystals of binary (with guide) and ternary (with guide and target) complexes of thermophilic Ago's could only be grown at elevated temperatures between 35 and 40 C (Wang et al. *Nature* 2008a,b). These advances allowed a range of questions to be addressed regarding the role of Ago as a scaffold for binding guide DNA and target RNA, the conditions under which both ends of the guide remain anchored in their respective pockets, what constitutes nucleation and propagation during guide-target pairing, as well as the nature of the conformational transitions on proceeding from catalytically-incompatible to catalytically-compatible and post-cleavage states.

Eubacterial Ago Complexes. Our x-ray structure of *T. thermophlius* Ago (*Tt*Ago) bound to a 5'-phosphorylated 21-mer guide DNA provided the first molecular view of the nucleic acid-binding channel within the Ago scaffold, as well as details of 5'-phosphate anchoring within the MID pocket, and 2-nt 3'-end anchoring within the PAZ pocket (Wang et al. *Nature* 2008a). These insights were guided by earlier research on anchoring of 5'-phosphate in the Mid pocket (Ma et al. *Nature* 2005; see also Parker et al.

Nature 2005) and 3'-ends in the PAZ pocket (Ma et al. *Nature* 2004; see also Lingel et al. *Nat. Struct. Mol. Biol*. 2004) at the domain level. Base stacking in the binary Ago complex was observed for bases 2 to 10 of the guide strand, with Watson-Crick edges of bases 2 to 6, that form part of the seed segment, directed outwards and accessible for pairing (nucleation process) with complementary bases of the target strand.

Our research was next extended to ternary complexes of *Tt*Ago with 5'-phosphorylated 21-mer guide DNA and complementary 12-mer, 15-mer and 19-mer target RNAs, thereby providing the first insights of guide-target pairing associated with nucleation (Wang et al. *Nature* 2008b) and propagation (Wang et al. *Nature* 2009) processes. The DNA guide-RNA target duplex adopted an A-helical conformation, with pivot-like conformational transitions observed for the N and PAZ domains of Ago on proceeding from the binary to ternary complexes. Importantly, while both ends of the 5'-phosphorylated guide DNA were anchored in their respective pockets for the ternary complex with 12-mer target RNA, the 3'-end was released from the PAZ pocket due to topological restraints for the ternary complexes with longer 15-mer and 19-mer target RNAs. Such a release was accompanied by conformational transitions in three loops that positioned a triad of catalytic aspartate residues from the RNase H-fold adopted by the PIWI domain, together with a pair of bridging Mg^{2+} cations, opposite the cleavable phosphate at the 10'-11' step of the target RNA, thereby generating a catalytically-compatible cleavage state. The above studies on binary and ternary *Tt*Ago-nucleic acid complexes highlight the power of structural biology to monitor individual events in molecular and mechanistic detail related to nucleation, propagation and cleavage steps in the RNA silencing pathway.

An intriguing question relates to why eubacterial and archael Ago's bind 5'-phosphorylated guide DNA rather than guide RNA. This issue has been recently addressed in functional experiments championed by the John van der Oost lab (Wageningen University, The Netherlands) where it was demonstrated that Ago from *T. thermophilus* acts as a barrier for the uptake and propagation of foreign DNA (Swarts et al. *Nature* 2004; see also Olovnikov et al. *Mol. Cell* 2013). It was shown *in vivo* that *Tt*Ago is loaded with 5'-phosphorylated guide DNAs ranging from 13 to 25 nt in length with a strong bias for C as the 5'-residue, which in turn cleave complementary DNA targets. These experiments support the concept that *Tt*Ago functions in host defense by a DNA-guided DNA interference pathway.

Eukaryotic Argonaute Complexes. Subsequent extension of our research on eubacterial Ago-nucleic acid complexes to binary complexes of eukaryotic Ago's with 5'-phosphorylated guide RNAs established that compared to Ago's in inactive conformations, budding yeast *Kp*Ago forms a hydrogen-bonded network that stabilizes an expanded and repositioned loop, which inserts an invariant glutamate into the catalytic pocket, thereby completing a universally conserved RNase H-type catalytic tetrad (Nakanishi et al. *Nature* 2012; with the David Bartel lab).

Humans have four Ago proteins labeled 1 to 4, with only hAgo2 exhibiting slicer activity. Recent efforts by several groups including our own have resulted in structures of human Ago2 (see Schirle and MacRae *Science* 2012; Elkayam et al. *Cell* 2012) and Ago1 (Faehnle et al. *Cell Reports* 2013; Nakanishi et al. *Cell Reports* 2014) bound to endogenous 5'-phosphorylated guide RNAs, thereby raising the question as to why hAgo2 is the sole family member exhibiting slicer activity. X-ray-based structure-function studies have established that reconstitution of the catalytic tetrad (arginine to histidine substitution) and replacement of Pro residues in conserved loop 7 by their counterparts in hAgo2 resulted in a measureable level of slicer activity for hAgo1 (Nakanishi et al. *Cell Reports* 2014; with Thomas Tuschl lab, Rockefeller; see also Faehnle et al. *Cell Reports* 2013; Hauptmann et al. *Nat. Struct. Mol. Biol.* 2013).

In a functional study championed by the Mien-Chie Hung lab (M. D. Anderson Cancer Center), the epidermal growth factor receptor (EGFR), which is the product of a well-characterized oncogene in human cancers, suppresses the maturation of specific tumor suppressor-like mRNAs in response to hypoxic stress through phosphorylation of hAgo2 at Tyr393 (Shen et al. *Nature* 2013). The association between EGFR and hAgo2 is enhanced by hypoxia, leading to elevated hAgo2 Tyr393 phosphorylation, which in turn reduces the binding of Dicer to Ago2 and inhibits miRNA processing from precursor miRNAs to mature miRNAs. Furthermore, Tyr393 phosphorylation of hAgo2 mediates EGFR-enhanced cell survival and invasiveness under hypoxia, while correlating with poorer overall survival in breast cancer patients.

2.3 VIRAL SUPPRESSORS OF RNA SILENCING

Many viruses counter the host antiviral RNA silencing process by evolving suppressor proteins that target individual steps of the silencing pathway. One such viral suppressor, p19 from tomato bushy stunt virus, has been shown to target siRNAs, with our x-ray-based structure of the complex defining the molecular basis for sequestration of the siRNA by this viral suppressor (Ye et al. *Nature*, 2003; see also Vargason et al. *Cell* 2003). p19 forms a homodimer containing an eight-stranded β-sheet whose concave face interacts with both grooves of the RNA duplex. More importantly, a single reading-head α-helix projects from opposite ends of the symmetrical p19 dimer, precisely positioning pairs of tryptophan rings for stacking over the terminal base pairs of bound siRNA, thereby using a caliper-like mechanism to measure the length of the siRNA duplex, resulting in sequestration.

Viral suppressors also target proteins that play key roles in the RNA silencing pathway. Thus, in a project championed by the Nam-Hai Chua lab (Rockefeller), cucumber mosaic virus-encoded 2b suppressor blocks *A. thaliana* Ago1 cleavage activity to inhibit miRNA pathways, attenuate RNA silencing and counter host defense (Zhang et al. *Genes Dev.* 2006).

Publications: Patel Lab

Ma, J-B., Ye, K. & Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. ***Nature*** 429, 318-322.

Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T. & Patel, D. J. (2005). Structural basis for 5'-end-specific recognition of the guide RNA strand by the *A. fugidus* PIWI protein. ***Nature*** 434, 666-670.

Nakanishi, K., Ascano, M., Gogakos, T., Ishibi-Murakami, S., Serganov, A. A., Briskin, D., Morozov, P., Tuschl, T. and Patel, D. J. (2013). Eukaryote-specific insertion elements control human ARGONAUTE slicer activity. ***Cell Reports*** 3, 1893-1900.

Nakanishi, K., Weinberg, D. E., Bartel, D. P. & Patel, D. J. (2012). Structure of yeast Argonaute with guide RNA. ***Nature*** 486, 368-374.

Park, J. E., Heo, I., Tian, Y., Simanshu, D. K., Chang, H., Jee, D., Patel, D. J. & Kim, V. N. (2011). Dicer recognizes the 5'-end of RNA for efficient and accurate cleavage. ***Nature*** 475, 201-205.

Shen, J., Xia, Y., Khotskaya, Y. B., Huo, L., Nakanishi, K., Lim, S-O., Du, Y., Wang, Y., Chang, W-C., Chen, C-H., Hsu, J. L., Lam, Y. C., James, B. P., Liu, C-G., Liu, X., Patel, D. J. & Hung, M. C. (2013). EGFR modulates miRNA maturation in response to hypoxia through phosphorylation of Ago2. ***Nature*** 497, 383-387.

Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Wang, Y., Patel, D. J., Berenguer, J., Brouns, S. J. and van der Oost, J. (2014). DNA-guided DNA interference by prokaryotic Argonaute. ***Nature*** in press.

Tian, Y., Simanshu, D. K., Park, J-E, Heo, I., Ma, J-B., Kim, V. N. & Patel, D. J. (2014). A phosphate-binding pocket within the platform-PAZ cassette of human Dicer. ***Mol. Cell*** in press.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T. & Patel, D. J. (2008b). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. ***Nature*** 456, 921-926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G. S., Tuschl, T. & Patel, D. J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. ***Nature*** 461, 754-761.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. (2008a). Structure of the guide-strand-containing argonaute silencing complex. ***Nature*** 456, 209-213.

Weinberg, D., Nakanishi, K., Patel, D. J. & Bartel, D. P. (2011). The inside-out mechanism of Dicers from budding yeasts. ***Cell*** 146, 262-276.

Ye, K., Malinina, L. & Patel, D. J. (2003). Recognition of siRNA by a viral suppressor of RNA silencing. ***Nature*** 426, 874-878.

Yuan, Y. R., Ma, J. B., Kuryavyi, V., Pei, Y., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T. & Patel, D. J. (2005). Crystal structure of *Aquifex aeolicus* Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. ***Mol. Cell*** 19, 405-419.

Zhang, X., Yuan, Y-R., Pei, Y., Tuschl, T., Patel, D. J. & Chua, N-H. (2006). Cucumber mosaic virus-encoded 2b suppressor inhibits *Arabidopsis* AGO1 cleavage activity to counter plant defense. ***Genes Dev*****.** 20, 3255-3268.

Complementary Publications: Other Labs

Elkayam, E. et al. (2012). The structure of human Argonaute-2 in complex with miR-20a. ***Cell*** 150, 100-110.

Faehnle, C. R. et al. (2013). The making of a slicer: activation of human Argonaute-1. ***Cell Reports*** 3, 1901-1909.

Hauptmann, J. et al. (2013). Tuning catalytically inactive human Argonaute proteins into active slicer active enzymes. *Nat. Struct. Mol. Biol*. **20**, 814-817.

Lingel, A. et al. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. ***Nat. Struct. Mol. Biol.*** 11, 576-577.

Olovnikov, I. et al. (2013). Bacterial Argonaute samples the transcriptome to identify foreign DNA. ***Mol. Cell*** 51, 594-605.

Parker, J. S. et al. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. ***Nature*** 434, 663-666.

Schirle, N. T. and MacRae, I. J. (2012). The crystal structure of human Argonaute2. *Science* 236, 1037-1040.

Vargason, J. M. et al. (2003). Size selective recognition of siRNA by an RNA silencing suppressor. ***Cell*** 115, 799-811.

3. HISTONE MARK-MEDIATED EPIGENETIC REGULATION

(2006 to present)

Contributions

The packaging of DNA within chromosomes, the orderly replication and distribution of chromosomes, the maintenance of genomic integrity, and the regulated expression of genes depend on nucleosomal histone proteins. Our long-term goals are directed toward gaining structural and mechanistic insights into the functional relevance of histone covalent modifications and the molecular principles underlying site-specific readout in a context-dependent manner by writers, readers and erasers of such marks.

Mono-, di- and tri-methylated states of histone lysine residues, together with their acetylated counterparts, are selectively found in different regions of chromatin, thereby implying specialized biological functions for these marks ranging from heterochromatin formation to X-chromosome inactivation and transcriptional regulation. A major challenge in chromatin biology has centered on efforts to define the connection between specific lysine methylation and acetylation states and distinct biological readouts impacting on function. Our efforts have focused on understanding the diversity of mechanisms for site-and state-specific readout of histone lysine methylation and acetylation marks by effector modules.

3.1 READOUT OF HISTONE MARKS

Readout of Kme Marks. An unaddressed challenge in the field has related to whether readers of histone methyl-lysine (Kme) marks used a single recognition mode with common principles or exhibited diversity, thereby expanding on the existing knowledge base of principles underlying molecular recognition.

Earlier studies had identified the second plant homeodomain (PHD) finger of BPTF, the largest subunit of the ATP-dependent chromatin remodeling complex, to be a reader of the H3K4me3 mark, with a methylation state preference of me3>me2>me1. Our x-ray based structural studies defined the intermolecular contacts between the H3K4me3-containing peptide and the BPTF PHD finger, whereby the Kme3 side chain inserted into a ‘surface groove’ aromatic amino acid-lined cage and was stabilized by cation-π interactions (Li et al. *Nature* 2006; with David Allis lab, Rockefeller; see also Pena et al. *Nature* 2006). The specificity for H3 and K4me3 originated in additional contacts involving positioning of the N-terminus in its own pocket, with R2 also being recognized in a ‘surface groove’ pocket adjacent to the K4me3 pocket, with the two separated by an invariant tryptophan ring.

In contrast to the BPTF PHD finger that targets higher lysine methylation states, the L3MBTL1 protein that contains three-malignant brain tumor (MBT) repeats, targets lower lysine methylation states in several histone contexts. Our x-ray-based structure of the complex of Kme2 bound to L3MBTL1 established that the dimethyl-lysine side chain inserted into an aromatic cage within MBT domain 2 using a ‘cavity insertion’ mode, given that the cage was both narrow and deep (Li et al. *Mol. Cell* 2007; with David Allis lab; see also Min et al. *Nat. Struct. Mol. Biol*. 2007). Thus, while the more open and wider ‘surface groove’ aromatic cage can accommodate the Kme3 mark, the more restrictive ‘cavity insertion’ aromatic cage is unable to sterically accommodate the bulkier Kme3 mark.

A third mode of Kme recognition emerged from structure-function studies of the ADD domain of alpha-thalassemia/mental retardation, X-linked (ATRX) protein. The x-ray-based structure of the complex between H3K9me3-containing peptide and the ATRX ADD domain (composed of adjacent GATA and PHD fingers) identified an unanticipated mode of histone mark recognition, namely the utilization of integrated modules to generate a composite and atypical binding pocket at the interface between the GATA and PHD fingers for H3K9me3 recognition (Iwase et al. *Nat. Struct. Mol. Biol.* 2011; with Yang Shi lab, Harvard Medical School; see also Eustermann et al. *Nat. Struct. Mol. Biol*. 2011). In this case, recognition did not involve an aromatic cage, but rather was facilitated by shape-complementarity and formation of atypical C-H••O hydrogen bonds. Notably, H3K9me3-pocket mutants and ATRX syndrome mutants are defective in both H3K9me3 binding and localization at pericentric heterochromatin.

Dysregulated Readout of Kme Marks. The dysregulation of methyl-lysine mark readout has been implicated in several human diseases, including cancers, immune and neurological disorders. Despite its

clinical implications, little has been known mechanistically of the underlying causes of such dysregulation. The fusion of H3K4me3-binding C-terminal PHD finger of JARID1A to nucleoporin-98 (NUP98), a common fusion partner identified in human leukemias, generated potent oncoproteins that arrested haematopoietic differentiation and induced acute myeloid leukemia in murine models. X-ray-based structure-function studies established that this H3K4me3/2-binding PHD finger was essential for leukemogenesis and mutations of a pair of aromatic surface groove cage-forming tryptophan residues that abrogated binding, also abolished leukemic transformations (Wang et al. *Nature* 2009; with David Allis lab). This example highlights how a dysregulated PHD finger perturbs the epigenetic dynamics on developmentally critical loci, impacts on cellular fate decision-making, and facilitates oncogenesis during mammalian development.

ORC1, a component of the origin of replication complex, which mediates pre-DNA replication licensing, contains a bromo adjacent homology (BAH) domain that specifically recognizes the H4K20me2 mark. X-ray-based structure-function studies showed that the H4K20me2 mark, which is enriched at replication origins, is recognized using a 'surface groove' aromatic cage recognition mode (Kuo et al. *Nature* 2012; with Or Gozani lab at Stanford). Functionally, abrogation of ORC1 recognition of H4K20me2 in cells impairs ORC1 occupation at replication origins, ORC chromatin loading and cell-cycle progression. Mutation of the ORC1 BAH domain has been implicated in the aetiology of Meier-Gorlin syndrome (MGS), a form of primordial dwarfism, and ORC1 depletion in zebrafish results in a MGS-like phenotype. Moreover, zebrafish depleted of the H4K20me2 mark have diminished body size, mirroring the phenotype of orc1 morphants. These studies establish the first direct link between histone methylation and the metazoan DNA replication machinery, as well as define a pivotal etiological role for the canonical H4K20me2 mark, via ORC1, in primordial dwarfism.

Spreading of Histone Marks. Polycomb repressive complex 2 (PRC2) regulates pluripotency, differentiation and tumorigenesis through catalysis of H3K27me3 mark on chromatin. NMR-based structure-function studies established that H3K36me3 binding activity is harbored in the Tudor domains of PRC2-associated polycomb-like proteins PHF1 and PHF9, with the H3K36me3 mark captured in a 'surface-groove' aromatic cage pocket. (Cai et al. *Mol. Cell* 2013; with Greg Wang lab at University of North Carolina; see also Musselman et al. *Nat. Struct. Mol. Biol*. 2012). Ectopically expressed PHF1 induced Tudor-dependent stabilization of PRC2 complexes on bulk chromatin and mediated spreading of PRC2 and H3K27me3 mark into active chromatin regions to promote gene silencing and modulate the chromatin landscape during development.

Multivalent Readout of Kme and Kac Marks at the Histone Level. The most abundant dual reader cassette amongst proteins involved in histone mark readout consists of adjacent PHD finger and Bromo domain modules (PHD-Bromo cassette), capable of simultaneous readout of Kme (PHD finger) and Kac (Bromo) marks on the same or different histone tails. We have investigated the PHD-Bromo cassette in TRIM24, TRIM33, BPTF and MLL1, key proteins associated with epigenetic regulation, given that both binding affinity and specificity are enhanced during multivalent readout of histone marks.

TRIM24, which contains a C-terminal PHD-Bromo cassette, is a ligand-dependent corepressor of retinoic acid receptor that interacts with multiple nuclear receptors *in vitro* via an LXXLL motif. Our x-ray-based structural studies established that the TRIM24 PHD-Bromo cassette formed a single functional unit for combinatorial recognition of unmodified H3K4 (PHD finger) and K23ac (Bromo) within the same histone tail (Tsai et al. *Nature* 2010; with Michelle Barton and Mien-Chie Hung labs, M. D. Anderson Cancer Center). Functional studies established that TRIM24 binds chromatin and oestrogen receptor to activate oestrogen-dependent genes associated with cellular proliferation and tumor development, and that aberrant expression of TRIM24 negatively correlated with survival of breast cancer patients.

It has been demonstrated that nodal receptors induce the formation of companion Smad4-Smad2/3 and TRIM33-Smad2/3 complexes. Our x-ray-based structure-function studies established that the PHD-Bromo cassette of TRIM33 facilitates binding of TRIM33-Smad2/3 to unmodified H3K4 and H3K9me3 (PHD finger) and H3K18ac (Bromo) marks on the promoters of mesendoderm regulators, thereby displacing the chromatin compacting factor HP1γ, making nodal response elements accessible to Smad4-Smad2/3 for pol II recruitment (Xi et al. *Cell* 2011; with Joan Massague lab, Memorial Sloan-Kettering Cancer Center). Thus, nodal effectors use the H3K9me3 and H3K18ac marks as a platform to switch master regulators of stem cell differentiation from the poised to the activated state.

Multivalent Readout of Kme and Kac Marks at the Nucleosomal Level. To date, the focus of efforts on readout of histone marks have been at the histone tail level, rather than the nucleosomal level. Our x-ray-based structural studies established that the PHD-Bromo cassette of BPTF combinatorially targets the H3K4me3 mark (PHD finger) and Kac marks on H4 (Bromo) at the histone tail level. Functional studies established that although BPTF Bromo displayed limited discrimination between K12ac, K16ac and K20ac on H4 at the histone peptide level, marked selectivity is observed for only H4K16ac at the nucleosomal level (Ruthenburg et al. *Cell* 2011; with David Allis lab). Further, this unique *trans*-histone modification pattern resides within a single nucleosomal unit in human cells and that the PHD-Bromo cassette colocalizes with the H3K4me3 and H4K16ac marks in the genome. This critical study outlines principles underlying recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions.

PHD-Bromo Cassette as a Regulatory Platform. The mixed lineage leukemia 1 (MLL1) gene is a frequent target for recurrent chromosomal translocations, resulting in transformation of hematopoietic precursors into leukemia stem cells. X-ray- and NMR-based structure-function studies have highlighted the role of the MLL1 PHD3-Bromo cassette as a regulatory platform, orchestrating MLL1 binding of H3K4me3/2 marks and cyclophilin Cyp33-mediated repression through histone deacetylase recruitment (Wang et al. *Cell* 2010; with the David Allis lab). In essence, the peptidylprolyl isomerase domain of Cyp33 regulated the conformation of MLL1 through *cis* to *trans* isomerization of a proline within the linker element the PHD-Bromo cassette, thereby disrupting the PHD-Bromo interface and facilitating binding to the interface of the Cyp33 RRM domain.

Readout of Unmodified Lys and Arg. An emerging challenge relates to what recognition principles have evolved to target unmodified lysines and arginines, in contrast to their methylated and acetylated counterparts. There are now several examples of recognition of unmodified lysine residues by reader modules (see Ooi et al. *Nature* 2007; Lan et al. *Nature* 2007). We have validated such recognition based on our x-ray structural studies mentioned above related to H3 tail recognition by the PHD fingers of TRIM24 (Tsai et al. *Nature* 2010) and TRIM33 (Xi et al. *Cell* 2011), as well as by the ADD domain of ATRX (Iwase et al. *Nat. Struct. Mol. Biol.* 2011), all of which recognize unmodified H4K4. The structures establish that recognition of the ε-ammonium group of unmodified lysines required hydrogen bond network alignments involving side chains of acidic amino acids and backbone carbonyl groups.

WDR5 protein, a core component of SET1-family MLL1 complexes that achieve transcriptional activation, is essential for vertebrate development, HOX gene activation and global H3K4 trimethylation. Our x-ray-based structure-function studies of the H3 tail-WDR5 complex established that the side chain of unmodified H3R2 inserted into the central cavity of the cylindrical WD40-type scaffold, with the guanidinium group of the inserted arginine sandwiched between a pair of phenylalanine rings and anchored in place through backbone carbonyl-mediated hydrogen bonds (Ruthenburg et al. *Nat. Struct. Mol. Biol*. 2006; with Gregory Verdine lab at Harvard and David Allis lab; see also Couture et al. *Nat. Struct. Mol. Biol*. 2006).

Our x-ray-based structure-function studies identified the PHD finger of UHRF1, an important regulator of DNA CpG methylation, as a reader of unmodified H3R2. Recognition of H3R2 is mediated through an extensive intermolecular hydrogen bonded network, with methylation of H3R2, but not H3K4 or H3K9, disrupting complex formation (Rajakumara et al. *Mol. Cell* 2011; with Yang Shi lab). Functionally, UHRF1's ability to repress its target genes is dependent on PHD finger binding to unmodified H3R2.

Erasers of Histone Marks. The existing paradigm that histone lysine methylation marks were stable was overturned by the discovery of LSD1 and jumonji family histone lysine demethylases (KDMs). Jumonji family KDMs are Fe^{2+} and α-ketogluterate-dependent oxygenases that are essential components of regulatory transcription chromatin complexes. These jumonji-domain KDMs demethylate lysine residues in histones in a methylation-state and sequence-specific context. A x-ray-based structure-guided small molecule and chemoproteomics approach has elucidated the functional role of the H3K27me3-specific KDM6 subfamily members JMJD3 and UTX (Kruidenier et al. *Nature* 2012; championed by the GlaxoSmithKline group, Stevenage, England). These studies identified a small molecule catalytic site inhibitor (GSK-J1) that is selective for the H3K27me3-specific KDM6 subfamily demethylases, with this inhibitor binding in a novel manner to reduce lipopolysaccharide-induced

proinflammatory cytokine production by human primary macrophages, a process that is dependent on both JMJD3 and UTX.

3.2 PROTEIN-PROTEIN INTERACTIONS IMPACTING ON EPIGENETIC REGULATION

Peptide-protein and protein-protein interactions play key roles in epigenetic regulation. Our group has investigated several examples of such interactions on systems ranging from the DNA damage response, to acute myeloid leukemia, as well as histone chaperones to chromatin remodelers as outlined below.

Novel Tyrosine Kinase Regulates H2A.X-mediated DNA Damage Response. The inability to repair double-strand breaks in DNA of eukaryotic cells can lead to genomic instability, carcinogenesis and cell death. During the double-strand break response, mammalian chromatin undergoes reorganization demarcated by H2A.X Ser139 phosphorylation (γ-H2A.X). Functional studies identified a new regulatory mechanism mediated by Williams-Beuren syndrome transcription factor (WSTF), a component of the WICH chromatin remodeling complex, that phosphorylates Tyr142 of H2A.X, with WSTF activity having an important role in regulating events critical for the DNA damage response (Xiao et al. *Nature* 2009; with David Allis lab).

Control of Leukemogenesis by AML-ETO Transcription Factor Complex. AML1-ETO, a fusion protein generated by the t(8;21) translocation in acute myeloid leukemia, resides in and functions through a stable transcription factor complex (AETFC). Within the AEFTC complex, AML1-ETO oligomerization is required for a specific interaction between the NHR2 domain and a novel NHR2-binding (N2B) motif in E proteins. X-ray-based structure-function studies of the NHR2-N2B complex reveals a unique interaction pattern in which an N2B peptide makes direct contact with side chains of two NHR2 domains as a dimer, providing a novel model of how transcription factors create a new protein-binding interface through dimerization/oligomerization (Sun et al. *Nature* 2013; with Robert Roeder lab at Rockefeller). Importantly, disruption of this interaction by multiple point mutations abrogates AML1-ETO-induced haematopoietic stem/progenitor cell self-renewal and leukemogenesis.

Histone Chaperones. Histone chaperones represent a structurally and functionally diverse family of histone-binding proteins that prevent promiscuous interactions of histones before their assembly into chromatin. DAXX is a metazoan histone chaperone specific for the evolutionary conserved histone variant H3.3. X-ray-based structure-function studies on the DAXX histone-binding domain bound to histone H3.3/H4 dimer establish that DAXX wraps around the H3.3/H4 dimer, with complex formation accompanied by structural transitions in the H3.3/H4 fold (Elsasser et al. *Nature* 2012; with David Allis lab; see also Liu et al. *Nat. Struct. Mol. Biol*. 2012). Further, DAXX uses an extended α-helical conformation to compete with major inter-histone, DNA and chaperone ASF1 interactions sites. Structural studies identified recognition elements that readout H3.3-specific residues, while functional studies addressed the contributions of Gly90 in H3 and Gln225 in DAXX to chaperone-mediated H3.3 variant recognition specificity.

Role of Chromatin Remodeler SMARCA3 in p11-dependent anti-depressive action. Selective serotonin reuptake inhibitors (SSRIs) are currently the most widely used class of antidepressants. Nevertheless, our knowledge of the molecular mechanisms underlying the efficacy of long-term treatment with SSRIs and the pathophysiology of depression is still rudimentary. p11 is a pivotal regulator of depression-like behaviors and a mediator of antidepressant responses. Annexin A2 (AnxA2) is a well-characterized binding partner for p11. Our efforts have focused on identifying proteins that target the p11-AnxA2 complex, as an approach towards contributing mechanistic insights to our current understanding of the SSRI/p11 signaling pathway. The protein p11, through unknown mechanisms, is required for behavioral and cellular responses to selective serotonin reuptake inhibitors (SSRIs). Our structure-function studies established that SMARCA3, a chromatin-remodeling factor, is a target for the p11/annexin A2 heterotetrameric complex. The SMARCA3 peptide bound within a hydrophobic channel in the p11-annexin A2 heterotetramer, resulting in an increase in the DNA-binding affinity of SMARCA3 and its localization to the nuclear matrix faction (Oh et al. *Cell* 2013; championed by Paul Greengard lab, Rockefeller). The available behavioral studies indicate a central role for the SMARCA3 chromatin-remodeling factor in the SSRI/p11 signaling pathway and suggest an approach to the development of improved antidepressant therapies.

Publications: Patel Lab

Cai, L., Rothbart, S. B., Lu, R., Xu, B., Tripathy, A., Chen, W-Y., Zheng, D., Patel, D. J., Allis, C., Strahl, B. D., Song, J. & Wang, G. (2013). An H3K36me3-containing Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. ***Mol. Cell*** 49, 571-582.

Elsasser, S. J., Huang, H., Lewis, P. W., Allis, C. D. & Patel, D. J. (2012). DAXX histone chaperone envelops an H3.3/H4 dimer for H3.3-specific recognition. ***Nature*** 491, 560-565.

Iwase, S., Xiang, B., Ghosh, S., Ren, T., Lewis, P. W., Cochrane, J. C., Allis, C. D., Picketts, D. J., Patel, D. J., Li, H. & Shi, Y. (2011). ATRX links atypical histone methylation recognition mechanisms to human cognitive function. ***Nat. Struct. Mol. Biol****.* 18, 769-776.

Kruidenier, L., Chung, C., Cheng, Z., Liddle, J., Bantscheff, M., Bountra, C., Bridges, A., Che, K., Diallo, H., Eberhard, D., Hutchinson, S., Joberty, G., Jones, E., Katso, R., Leveridge, M., Mosley, J., Rowland, P., Ramirez-Molina, C., Schofield, C. J., Sheppard, R., Smith, J. E., Swales, C., Tanner, R., Thomas, P., Tumber, A., Drewes, G., Oppermann, U., Patel, D. J., Lee, K., & Wilson, W. (2012). A selective H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. ***Nature*** 488, 404-408.

Kuo, A. J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J., Patel, D. J. & Gozani, O. (2012). ORC1 BAH domain links dimethylation of H4K20 to DNA replication licensing and Meier-Gorlin syndrome. ***Nature*** 484, 115-119.

Li, H., Fischle, W., Wang, W. K., Duncan, E. M., Liang, L., Murakami-Ishibe, S., Allis, C. D. & Patel, D. J. (2007). Structural basis for lower lysine methylation state-specific readout by MBT repeats and an engineered PHD finger module*.* ***Mol. Cell*** 28, 677-691.

Li, H., Ilin, S., Wang, W. K., Wysocka, J., Allis, C. D. & Patel, D. J. (2006). Molecular basis for site-specific readout of H3 lysine 4 trimethylation by the BPTF PHD finger. ***Nature*** 442, 91-95.

Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., Guo R., Wu, F., Li, H., Lan, F., Shi, Y., Xu, Y., Patel, D. J. & Shi, Y. (2011). PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. ***Mol. Cell*** 43, 275-284.

Ruthenberg, A. J., Wang, W., Graybosch, D. M., Li, H., Allis, C. D., Patel, D. J. & Verdine, G. L. (2006). Histone H3 lysine 4 methylation state recognition and presentation by the WDR5 module of the MLL1 complex. ***Nat. Struct. Mol. Biol*****.** 13, 704-712.

Ruthenburg, A., Li, H., Milne, T., Dou, Y., McGinty, R. K., Yuen, M., Muir, T. W., Patel, D. J. & Allis, C. D. (2011). Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. ***Cell*** 145, 692-706.

Sun, X-J., Wang, Z., Wang, L., Jiang, Y., Chen, W-Y., Melnick, A., Patel, D. J., Nimer, S. D. & Roeder, R. G. (2013). A stable transcription factor complex nucleated by dimeric AML1-ETO controls leukemogenesis. ***Nature*** 500, 93-97.

Tsai, W-W., Wang, Z., Yiu, T. T., Akdemir, K. C., Xia, W., Winter, S., Tsai, C-Y., Shi, X., Schwarzer, D., Plunkett, W., Aronow, B., Gozani, O., Fischle, W., Hung, M. C., Patel, D. J. & Barton, M. C. (2010). TRIM24 links recognition of a non-canonical histone signature to breast cancer. ***Nature*** 468, 927-932.

Wang, G. G., Song, J., Wang, Z., Dormann, H. L., Casadio, F., Li, H., Luo, J., Patel, D. J. & Allis, C. D. (2009). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. ***Nature*** 459, 847-851.

Wang, Z., Song, J., Milne, T. A., Wang, G. G., Li, H., Allis, C. D. & Patel, D. J. (2010). Pro isomerization in MLL1 PHD3-Bromo cassette connects H3K4me3 readout to CyP33 and HDAC-mediated repression. ***Cell*** 141, 1183-1194.

Xi, Q., Wang, Z., Zaromytidou, A., Zhang, X. H., Chow-Tsang, L-F., Liu, J. X., Kim, H., Monova-Todorova, K., Kaartinen, V., Studer, L., Mark, W., Patel, D. J. & Massague, J. (2011). A poised chromatin platform for Smad access to master regulators. ***Cell*** 147, 1511-1524.

Xiao, A., Li, H., Shechter, D., Ahn, S. H., Fabrizio, L., Erajument-Bromage, H., Murakami-Ishibe, S., Wang, B., Tempst, P., Hofmann, K., Patel, D. J., Elledge, S. J. & Allis, C. D. (2009). WSTF regulates the DNA damage response of H2A.X via a novel tyrosine kinase activity. ***Nature*** 457, 57-62.

Complementary Publications: Other Labs

Couture, J. F. et al. (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. ***Nat. Struct. Mol. Biol.*** 13, 698-703.

Eustermann, S. et al. (2011). Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. ***Nat. Struct. Mol. Biol***. 18, 777-782.

Lan, F. et al. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* 448, 718-722.

Liu, C. P. et al. (2012). Structure of the variant histone H3.3-H4 heterodimer in complex with its chaperone DAXX. ***Nat. Struct. Mol. Biol***. 19, 1287-1292.

Min, J. et al. (2007). L3MBTL1 recognition of mono- and dimethylated histones. ***Nat. Struct. Mol. Biol***. 14, 1229-1230.

Musselman et al. (2012). Molecular basis of H3K36me3 recognition by the Tudor domain of PHF1. ***Nat. Struct. Mol. Biol***. 19, 1266-1272.

Oh, Y-S., Gao, P., Lee, K., Ceglia, I., Zhang, X., Ahn, J-H., Chait, B. T., Patel, D. J., Kim, Y. & Greengard, P. (2013). SMARCA3, a chromatin remodeling factor, is required for p11-dependent anti-depressive action. ***Cell*** 152, 831-843.

Ooi, S. K. et al. (2007). DNMT2L connects unmethylated lysine 4 and histone H3 to *de novo* methylation of DNA. **Nature** 448, 714-717.

Pena, P. V. et al. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. ***Nature*** 442, 100-103.

4. DNA METHYLATION MARK-MEDIATED EPIGENETIC REGULATION

(2011 to present)

Methylation of cytosine in the CpG context has pronounced effects on gene expression with DNA methylation patterns established during embryonic development and then faithfully maintained during subsequent somatic cell division. The basic principles underlying the setting up and maintenance of DNA methylation patterns remains an area of intense research, given that perturbation of DNA methylation patterns impacts on a range of human diseases. Our current efforts are focused on structure-function studies of writers of DNA methylation marks and their complexes with unmodified and hemimethylated DNAs.

Establishment and maintenance of DNA methylation in *A. thaliana* involves the RNA-directed DNA methylation pathway. Our efforts have been directed towards understanding the role of RNA polymerases pol-IV and pol-V and their interacting protein partners in mediating RNA-directed DNA methylation.

Contributions

4.1 READOUT OF DNA METHYLATION MARKS

Mammalian Maintenance DNA Methyltransferases. Maintenance of genetic methylation patterns in mammals is mediated primarily by DNA methyltransferase-1 (DNMT1), thereby helping to regulate gene expression, with an impact on genome imprinting and X-chromosome inactivation. The X-ray-based structure of a human DNMT1 construct (contains CXXC, BAH1/2 and methyltransferase domains) bound to dsDNA containing unmethylated CpG sites has identified an autoinhibitory mechanism, in which unmethylated CpG dinucleotides are occluded from the active site to ensure that only hemimethylated CpG sites undergo methylation (Song et al. *Science* 2011). This is achieved by positioning the CXXC-BAH1 linker between the DNA and the catalytic pocket containing methyl donor S-adenosylmethionine, thereby preventing looping out of the to-be-methylated cytidin. In addition, autoinhibition is also mediated through a stabilizing interaction between the tip of a loop projecting from the BAH2 domain and the target recognition domain (TRD) of the methyltransferase, thereby holding the TRD in a retracted position off the DNA and preventing its insertion into the major groove to facilitate looping out of the to-be-methylated cytosine. The concept of DNMT1 functioning in part by an autoinhibitory mechanism has also emerged from the crystal structure of a larger DNMT1 construct (includes the RFD domain) in the free state (see Takeshita et al. *Proc. Natl. Acad. Scis. USA* 2011)

The x-ray-based structure of a shorter human DNMT1 construct (contains BAH1/2 and methyltransferase domains) bound covalently to dsDNA containing a single hemimethylated CpG site has provided insights into formation of a productive complex (Song et al. *Science* 2012). In this complex, the methyl group of the meC is positioned within a shallow concave hydrophobic surface, while three loops of the methyltransferase are inserted into the grooves of the DNA, thereby facilitating looping out of the cytidine on the target strand and its covalent anchoring within the catalytic pocket. The DNA is distorted at the hemimethylated CpG step, with side chains from recognition and catalytic loops inserting through both grooves to fill an intercalation-type cavity associated with a staggered dual base flip-out on partner strands. The above structural studies establish how a combination of autoinhibitory and productive mechanisms ensures the high fidelity of DNMT1-mediated maintenance DNA methylation.

Plant Maintenance DNA Methyltransferases. Both histone modification and DNA methylation exert epigenetic control over gene expression. A combined x-ray-based structure-function study has been undertaken on the complex of H3K9me2 peptide and chromomethylase3 (CMT3), a plant-specific methyltransferase responsible for CHG methylation in *A. thaliana* (Du et al. *Cell* 2012; with Steve Jacobsen lab at UCLA Medical School). ZMET2 (a maize analog of CMT3) adopts a triangular architecture with its BAH and chromo domains positioned at the apices of the triangle. In the structure of the complex, H3K9me2 peptides are bound within ‘surface groove’ aromatic cages for both the BAH and chromo domains. Mutations that abolish either interaction disrupt CMT3 binding to nucleosomes, and exhibit a complete loss of CMT3 activity *in vivo*. Further, CMT3 genome locations are nearly perfectly correlated with H3K9me2 marks, with CMT3 stably associated with H3K9me2-containing nucleosomes.

4.2 RNA-DIRECTED DNA METHYLATION IN PLANTS

In *A. thaliana*, DNA methylation is established by the domains rearranged methyltransferase (DRM2) and targeted by 24-nt siRNAs through a pathway termed RNA-directed DNA methylation (RdDM). This pathway requires two plant-specific RNA polymerases, pol-IV, which functions to initiate siRNA biogenesis, and pol-V, which functions to generate scaffold transcripts that recruit downstream RdDM factors. We have initiated a structure-function research program to identify proteins that mediate the role of pol-IV and pol-V in RNA-directed DNA methylation by solving the structures of complexes of such proteins with histone peptides containing methylation marks.

Role of pol-IV in RNA-directed DNA Methylation. Our X-ray-based structure-function studies have shown that the Sawadee homeodomain homolog 1 (SHH1), a pol-IV interacting protein, adopts a unique tandem Tudor-like fold and functions as a dual lysine reader, probing for both unmethylated K4 and methylated K9 modification on the histone H3 tail (Law et al. *Nature* 2013; with Steve Jacobsen lab at UCLA Medical School; see also Zhang et al. *Proc. Natl. Acad. Scis. USA* 2013). Mutation studies establish that key residues within both lysine-binding pockets of SHH1 are required *in vivo* to maintain siRNA and DNA methylation levels, as well as pol-IV occupancy at RdDM targets.

Role of pol-V in RNA-directed DNA Methylation. RNA-directed DNA methylation (RdDM) in *A. thaliana* depends on the synthesis of non-coding RNAs by pol-V. Genome-wide studies of pol-V association with chromatin redundantly required SUVH2 and SUVH9, proteins that resemble histone lysine methyltransferases, but lack activity given that the structure of apo-SUVH9 lacks a substrate-binding cleft and lacks a properly formed SAM pocket necessary for normal catalysis (Johnson et al. *Nature* 2014; with Steve Jacobsen lab). Given that SUVH2 and SUVH9 both contain SRA domains capable of binding methylated DNA, it appears likely that they function in recruiting pol-V through DNA methylation. Consistent with this model, functional studies establish that mutation of DNA methyltransferase 1 (MET1) causes loss of DNA methylation and a nearly complete loss of pol-V at its normal locations.

Publications: Patel Lab

Du, J., Zhong, X., Barnatavichute, Y. V., Stroud, H., Feng, S., Caro, E., Vashisht, A. A., Terragni, J., Chin, H. G., Tu, J., Hetzel, J., Wohlschlegel, J. A., Pradhan, S., Patel, D. J. & Jacobsen, S. E. (2012). Dual binding of chromomethylase BAH and chromo domains to H3K9me2-containing nucleosomes in the targeting of DNA methylation. ***Cell*** 151,167-180.

Johnson, L. M., Du, J., Chodavarapu, R. K., Feng, S., Zhong, X., Hale, C. J., MarsoG., Segal, D. J., Pellergrini, M., Patel, D. J. and Jacobsen, S. E. (2014). SRA/SET domain proteins link RNA polymerase V binding to DNA methylation. ***Nature*** in press.

Law, J. A., Du, J., Hale, C. J., Feng, S., Krajewski, K., Strahl, B. D., Patel, D. J. & Jacobsen, S. E. (2013). SHH1 recruits RNA polymerase-IV to RNA-directed DNA methylation targets. ***Nature*** 498, 385-389.

Song, J., Rechkoblit, O., Bestor, T. H. & Patel, D. J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. ***Science*** 331, 1036-1040.

Song, J., Teplova, M., Ishibe-Murakami, S. & Patel, D. J. (2012). Structural principles underlying DNMT1-mediated DNA methylation. ***Science*** 335, 709-712.

Complementary Publications: Other Labs

Takeshita, K. et al. (2011). Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (DNMT1). ***Proc. Natl. Acad. Scis. USA*** 108, 9055-9059.

Zhang, H. et al. (2013). DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of pol IV. ***Proc. Natl. Acad. Scis. USA*** 110, 8290-8295.

5. PATTERN RECOGNITION RECEPTORS OF CYTOSOLIC NUCLEIC ACIDS

(2010 - present)

Host defense against infection by viral and bacterial pathogens is critically dependent on the initiation and maintenance of the finely tuned innate immune response, a rapid protective response that is coupled to subsequent adaptive immunity, thereby providing long-term protection based on immunological memory. The innate immune response is equipped with a number of pattern recognition receptors that detect characteristic microbial components, ranging from unmethylated CpG DNA, double-stranded RNA and 5'-triphosphorylated RNA. Our most recent efforts have been primarily focused on the metazoan second messenger cyclic-GAMP (cGAMP) produced by DNA-activated cyclic GAMP-AMP synthase (cGAS), together with elucidation of the principles underlying activation of hSTING by cGAMP and potential targeting of hSTING by anti-viral small molecules.

Contributions

5.1 PATTERN RECOGNITION RECEPTORS OF CYTOSOLIC dsDNA

Metazoan Cytosolic dsDNA-Sensing by cGAS. Recent studies from the Zhijian Chen lab (Univ. Texas Southwestern Medical Center, Dallas) identified cGAMP generated from GTP and ATP by activated cytosolic dsDNA sensor cGAMP synthase (cGAS) as the metazoan second messenger responsible for triggering an interferon response (see Sun et al. *Science* 2013; Wu et al. *Science* 2013). Our x-ray-based structural studies demonstrated that upon dsDNA binding, cGAS is activated through conformational transitions, resulting in formation of a catalytically competent and accessible nucleotide-binding pocket for generation of cGAMP containing G(2',5')pA and A(3',5')pG phosphodiester linkages, designated c[G(2',5')pA(3',5'p] (Gao et al. *Cell* 2013a; see also Diner et al. 2013; Civril et al. *Nature* 2013; Ablasser et al. *Nature* 2013; Zhang et al. *Mol. Cell* 2013). Structural, biochemical (with Thomas Tuschl lab at Rockefeller), cell biological (with Winfried Barchet and Gunther Hartmann labs at University Hospital, Bonn) and chemical (with Roger Jones lab at Rutgers, New Jersey) establish that chemistry within the binding pocket occured in a stepwise manner through initial generation of 5'-pppG(2',5')pA prior to cyclization to c[G(2',5')pA(3',5')p], thereby identifying a founding member of a family of metazoan 2',5'-containing cyclic dinucleotides distinct from bacterial all 3',5' cyclic dinucleotides.

Activation of Adaptor hSTING by cGAMP. Our x-ray-based structural studies of adaptor hSTING activation by cGAMP establish that the hSTING dimer adopts a 'closed' conformation on binding c[G(2',5')pA(3',5'p] and c[G(2',5')pA(2',5'p] linkage isomers, as does the mSting dimer on binding c[G(2',5')pA(3',5')p], c[G(3',5')pA(3',5')p] linkage isomers , as well as the antiviral drug DMXAA. Comparing hSTING to mSting, 2',5'-linkage-containing cGAMP isomers were more specific triggers of the interferon pathway compared to the all 3',5'-linkage isomer (Gao et al. *Cell* 2013b; with the Tuschl, Barchett/Hartmann and Jones labs; see also Ablasser et al. *Nature* 2013; Zhang et al. *Mol. Cell* 2013). Importantly, a pair of unique point mutations were identified which when placed within the cyclic dinucleotide-binding site of hSTING rendered it sensitive to the otherwise mouse-specific drug DMXAA (Gao et al. *Cell* 2013b; with the Barchett/Hartmann lab).

5.2 PATTERN RECOGNITION RECEPTORS OF 5'-ppp-RNA

5'-ppp dsRNA Pattern Recognition by RIG-I. RIG-I is a cytosolic helicase that senses 5'-ppp RNA contained in negative-strand viruses and triggers the innate antiviral immune response. Our x-ray-based structure of the RIG-I C-terminal domain (CTD) bound to blunt-end double-stranded 5'-ppp RNA established how a lysine-rich basic cleft within the RIG-I CTD sequesters the 5'-triphosphorylated end of the bound RNA (Wang et al. *Nat. Struct. Mol. Biol.* 2010; with Gunther Hartmann and Thomas Tuschl labs; see also Lu et al. *Structure* 2010). Key intermolecular contacts are retained in full-length RIG-I under *in vivo* conditions, as evaluated from the impact of binding pocket RIG-I mutations and 2'-OCH_3 RNA modifications on the interferon response.

Publications: Patel Lab

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B. L., Zillinger, T., Serganov, A., Jones, R. A., Hartmann, G., Tuschl, T. and Patel, D. J. (2013a). Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. ***Cell*** 153, 1094-1107.

Gao, P., Ascano, M., Zillinger, T., Wang, Y., Dai, P., Serganov, A. A., Gaffney, B. L., Shuman, S., Jones, R., Deng, L., Hartmann, G., Barchet, W., Tuschl, T. and Patel, D.J. (2013b) Structure-function analysis of STING activation by c[G(2',5')pA(3',5')p] and targeting by DMXAA. ***Cell*** 154, 748-762.

Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H., Juranek, S., Sheng, G., Micura, R., Tuschl, T., Hartmann, G. & Patel, D. J. (2010). Structural and functional insights into 5'-ppp-RNA pattern recognition by the innate immune receptor RIG-I. ***Nat. Struct. Mol. Biol.*** 17, 781-787.

Publications: Other Labs

Ablasser, A. et al. (2013). cGAS produces a 2',5'-linked cyclic dinucleotide second messenger that activates STING. ***Nature*** 498, 380-384.

Civril, F. et al. (2013). Structural mechanism of cytosolic DNA sensing by cGAS. ***Nature*** 498, 332-337.

Diner, E. J. et al. (2013). The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. ***Cell Reports*** 3, 1355-1361.

Lu, C. et al. 2010. The structural basis of 5' triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. ***Structure*** 18, 1032-1043.

Sun, L. et al. (2013). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. ***Science*** 339, 786-791.

Wu, J. et al. (2013). Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. ***Science*** 339, 826-830.

Zhang, X. et al. (2013). Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. ***Mol. Cell*** 51, 226-235.

6. NON-VESICULAR TRAFFICKING BY LIPID TRANSFER PROTEINS

(2004 - present)

Lipid transfer proteins are important in membrane vesicle biogenesis and trafficking, signal transduction and immunological presentation processes. The conserved and ubiquitous mammalian glycolipid transfer proteins (GLTPs) serve as potential regulators of cell processes mediated by glycosphingolipids (GSLs), ranging from differentiation and proliferation to invasive adhesion, neurodegeneration and apoptosis. We have initiated a structural biology program towards defining a framework for understanding how GLTPs acquire and release GSLs during lipid intermembrane transfer and presentation processes. This research has recently been extended to transfer of ceramide-1-phosphate by its transfer protein CPTP.

Contributions

6.1 GLYCOSPHINGOLIPID TRANSFER PROTEINS

The conserved and ubiquitous mammalian glycolipid transfer proteins (GLTPs) serve as potential regulators of cell processes mediated by glycosphingolipids (GSLs), ranging from differentiation and proliferation to invasive adhesion, neurodegeneration and apoptosis. Our x-ray based structure determination of GLTP in the free- and lactosylceramide-bound states established that GLTP adopts a newly-identified two-layer all α-helical topology, with GSL binding specificity achieved through anchoring of the lactosyl ring through hydrogen bond networks and hydrophobic contacts within a sugar headgroup recognition center, while encapsulating both lipid chains in a precisely oriented manner within a 'molded-to-fit' hydrophobic tunnel (Malinina et al. *Nature* 2004; with Rhoderick Brown lab at Hormel Institute). Mutation and functional analysis of residues lining the sugar headgroup center and within the hydrophobic tunnel support a framework for understanding of how GLTPs acquire and release GSLs during lipid intermembrane transfer and presentation processes.

A comparison of structures of GLTPs bound to GSLs of diverse acyl chain length, unsaturation and sugar composition established that acyl chain length structure and occupancy of the hydrophobic tunnel dictate partitioning between sphingosine-in and sphingosine-out binding modes (Malinina et al. *PLoS Biol.* 2006; with Rhoderick Brown lab). It appeared that a cleft-like conformational gating mechanism, involving two inter-helical loops and one α-helix of GLTP, could enable the glycolipid chains to enter and leave the tunnel in the membrane-associated state.

The potential development of 'engineered GLTPs' with enhanced specificity for select GSLs could provide a potential new therapeutic approach for targeting GSL-mediated pathologies. To this end, the structure of a D48V mutant of GLTP bound to N-nervonoyl-sulfatide revealed the molecular basis for specific anchoring of sulfatide (3-O-sulfo-galactosylceramide) (Samygina et al. *Structure* 2010; with the Lucy Malinina lab at CIC bioGUNE, Bilbao).

6.2 PHOSPHOSPHINGOLIPID TRANSFER PROTEINS

Ceramide-1-phosphate (C1P) and sphingosine-1-phosphate (S1P) have emerged as key regulators of cell growth, survival, migration and inflammation. To modulate eicosanoid action and avoid the damaging effects of chronic inflammation, cells require efficient targeting, trafficking and presentation of C1P to specific cellular sites. A ubiquitously expressed lipid transfer protein CPTP has now been shown from x-ray based structure-function studies to specifically transfer C1P between membranes, with structures of C1P-CPTP complexes establishing binding of C1P through a novel surface-localized, phosphate head group recognition center connected to an interior hydrophobic channel that adaptively expands to ensheath differing-length lipid chains using a cleft-like gating mechanism (Simanshu et al. *Nature* 2013; with Rhoderick Brown lab and Charles Chalfont lab at Virginia Commonwealth University, Richmond). RNA interference-induced CPTP depletion results in C1P decrease in plasma membranes and increase in the Golgi complex, thereby stimulating cytosolic phospholipase $A_2\alpha$ release of arachidonic acid, and triggering pro-inflammatory eicosanoid production.

The accelerated-cell-death 11 (acd11) mutant of *A. thaliana* provides a general model for studying immune response activation and localized cellular suicide that halts the spread of pathogen infection in plants. Our x-ray-based structure-function studies established that ceramide-1-phosphate (C1P) binds to ACD11 (like CPTP) via a surface-localized phosphate headgroup recognition center

connected to an interior hydrophobic pocket that adaptively accomodates lipid chains via a cleft-like gating mechanism (Simanshu et al. *Cell Reports* 2014; with Rhoderick Brown lab). A π-bulge near the lipid-binding cleft distinguishes apo-ACD11 from other GLTP folds. In acd11 mutants, normally low C1P levels become elevated, but the relatively abundant cell death inducer, phytoceramide, rises acutely.

Publications: Patel Lab

Malinina, L., Malakhova, M. L., Kanack, A. T., Brown, R. E. & Patel, D. J. (2006). The liganding mode of glycolipid transfer protein is controlled by glycolipid acyl structure. ***PLoS Biol.*** 4, 1996-2011.

Malinina, L., Malakhova, M. L., Teplov, A., Brown, R. E. & Patel, D. J. (2004). Structural basis for glycosphingolipid transfer specificity. ***Nature*** 430, 1048-1053.

Samygina, V., Popov. A. N., Cabo-Bilbao, A., Ochoa-Lizarralde, B., Goni-de-Cerio, F., Zhai, X., Molotkovsky, J. G., Patel, D. J. Brown, R. E., & Malinina, L. (2010). A designer human glycolipid transfer protein with enhanced transfer selectivity for sulfatide. ***Structure*** 19, 1625-1634.

Simanshu, D. K., Kamlekar, R. K., Wijesinghe, D. S., Zou, X., Zhai, X., Mishra, S. K., Molotkovsky, J. G., Malinina, L., Hincliffe, E. H., Chalfant, C. E., Brown, R. E. & Patel, D. J. (2013). Nonvesicular trafficking by ceramide-1-phosphate transfer protein regulates eicosanoid production. ***Nature*** 500, 463-467.

Simanshu, D., Zhai, X., Munch, D., Hofius, D., Markham, J. E., Bielawski, J., Bielawski, A., Slotte, J. P., Malinina, L., Molotkovsky, J. G., Mundy, J. E., Patel, D. J. and Brown, R. E. (2014). Arabidopsis accelerated-cell-death 11 (ACD11) is a ceramide-1-phosphate transfer protein that regulates levels of lipid inducer, phytoceramisde. ***Cell Reports*** in press.

7. PROTEIN-RNA COMPLEXES ON DISEASE-RELATED SYSTEMS

(2006 to present)

Our laboratory has had a long-term interest towards understanding the principles underlying protein-RNA recognition, with the focus on those complexes with identified sequence specific recognition of RNA targets by proteins of disease interest. A wealth of new and novel information has emerged on the principles of intermolecular recognition and protein dimerization in facilitating protein-RNA complex formation. The available structure-based information on intermolecular contacts in turn has guided functional experiments that probe the impact of mutations of interfacial residues in both the protein and RNA on complex formation.

Contributions

7.1 RECOGNITON SPECIFICITY IN PROTEIN-RNA COMPLEXES

La Autoantigen-RNA Complexes Implicated in Lupus and Sjogren's Syndrome. La has been identified as an autoantigen in patients with systemic lupus erythematosus and Sjogren's syndrome. X-ray-based structural studies by our group on the N-terminal La and RRM1 domains of La protein bound to RNAs containing UUU 3'-ends established that this UUU segment, in a splayed-apart orientation, is sequestered within a basic and aromatic amino acid-lined cleft between the La and RRM1 motifs (Teplova et al. *Mol. Cell* 2006). The structure of the complex explained the molecular basis underlying protection of 3'-UUU termini of nascent RNA polymerase transcripts during downstream processing and maturation events from exonuclease digestion.

Alternate Splicing Regulation by MBNL1 and CUGBP1 on Myotonic Dystrophy. Muscleblind-like (MBNL) protein and CUG-binding protein 1 (CUGBP1) regulate multiple aspects of nuclear and cytoplasmic processing, with implications for onset of myotonic dystrophy. Our x-ray-based structural studies have defined the intermolecular contacts between tandem zinc fingers ZnF3/4 of MBNL1 and its consensus r(CGCUGU) target site, with both ZnF3 and ZnF4 targeting GC steps, with site-specific recognition mediated by a network of hydrogen bonds primarily with the backbone of the protein (Teplova et al. *Nat. Struct. Mol. Biol*. 2008). The anti-parallel alignment of GC-binding segments was supportive of a chain-reversal loop trajectory for MBNL1 bound pre-mRNA targets.

X-ray-based structural studies by our group have also been undertaken on tandem RNA recognition motifs RRM1/2 of CUGBP1 and its consensus target sites containing tandem UGU(U/G) elements (Teplova et al. *Structure* 2010). The structure of the complex establishes that both RRM domains target tandem UGUU motifs, with the UG step adopting a left-handed Z-RNA conformation, with the *syn* guanosine recognized through Hoogsteen edge-protein backbone hydrogen bonding interactions. The above studies are consistent with MBNL1- and CUGBP1-mediated targeting of looped RNA segments proximal to splice-site junctions potentially contributing to pre-mRNA alternative-splicing regulation.

Nucleocytoplasmic Export of Retroviral Genomic RNA. mRNA export is mediated by the TAP-p15 heterodimer, which also binds to the constitutive transport element (CTE) in retroviral RNAs. Our x-ray-based structural and cellular studies on the RNA recognition and leucine-rich repeat motifs of TAP bound to one symmetrical half of the CTE RNA have defined the recognition principles whereby the L-shaped conformations of the protein and RNA are involved in a mutual molecular embrace on complex formation (Teplova et al. *Nat. Struct. Mol. Biol.* 2011; with Elisa Izaurralde lab, Max-Planck Institute, Tubingen). The impact of structure-guided mutations on binding affinities *in vitro* and transport assays *in vivo* have defined the principles by which CTE RNA subverts the mRNA export receptor TAP, thereby facilitating the nuclear export of viral genomic RNAs, and more generally, provides insights on cargo RNA recognition by mRNA export receptors.

Nova Tandem KH domain-RNA Complexes Implicated in POMA Syndrome. Nova onconeural antigens are neuron-specific RNA-binding proteins implicated in paraneoplastic opsoclonus-myoclonus-ataxia (POMA) syndrome. Our x-ray-based structural and functional studies of tandem KH1/2 domains of Nova bound to an *in vitro*-selected RNA hairpin containing a r(UCAG-UCAC) binding site established

intermolecular contacts between the KH1 domain and the second UCAC repeat, with the RNA scaffold buttressed by interaction between the adjacent UCAY repeats (Teplova et al. *Structure* 2011; with Robert and Jennifer Darnell lab, Rockefeller). By contrast, the KH2 domains are involved in dimerization, with the antiparallel alignment of KH1 and KH2 domains creating a scaffold that could facilitate target pre-mRNA looping on Nova complex formation, thereby providing insights into Nova's function in splicing regulation.

FMRP RGG Peptide Bound to Quadruplex-Duplex Junctional RNA. Fragile X mental retardation protein (FMRP) is a regulatory RNA binding protein that binds with high affinity to guanosine-rich RNAs capable of G-quadruplex formation. Loss of FMRP function leads to fragile X syndrome, the most common form of human mental retardation. Our solution based NMR structural and functional studies of the complex of an arginine-glycine-rich RGG peptide of FMRP bound to an *in vitro* selected guanosine-rich sc1 RNA established that the G-quadruplex forced a sharp turn in the bound peptide at the quadruplex-duplex junction (Phan et al. *Nat. Struct. Mol. Biol.* 2011; with Robert and Jennifer Darnell lab) Specificity and stabilization of complex formation was associated with shape complementarity, supplemented by intermolecular hydrogen bonds between a pair of arginines and the major groove edges of a pair of guanosines on partner strands of the duplex. These conclusions were validated following filter-binding assays of mutations of peptide and RNA residues involved in intermolecular recognition.

STAR Domain-containing Quaking Proteins Involved in Myelination. Mammalian Quaking (QKI) is an evolutionary conserved STAR domain-containing RNA-binding protein (contains Qua1, KH and Qua2 domains), which post-transcriptionally regulates target genes essential for developmental processes and myelination. Our x-ray-based structural and cellular studies have defined the principles underlying r(YUAAY) RNA target selection by STAR domain proteins QKI and its *C. elegans* homolog GLD-1 (Teplova et al. *Genes Dev.* 2013; with Thomas Tuschl lab). The KH and Qua2 motifs of the STAR domain synergize to specifically interact with the bases and sugar-phosphate backbone of the YUAAY RNA recognition elements (RREs). Qua-1-mediated homodimerization generates a scaffold that enables concurrent recognition of two RREs, thereby targeting tandem RREs present in many QKI-targeted transcripts. Expression of structure-guided QKI mutants in human embryonic kidney cells significantly decreased the abundance of QKI target mRNAs.

MazF Toxin-RNA and MazF Toxin-MazE Antitoxin Complexes. MazF is a mRNA interferase, which on undergoing activation under stress conditions, cleaves mRNA in a sequence-specific manner, resulting in cellular growth arrest. MazF is inactivated under normal growth conditions through binding to its cognate antitoxin MazE. Our x-ray-based structural and functional studies of the *B. subtilis* MazF-mRNA complex defined the molecular basis underlying sequence-specific recognition of the UACAU segment by a MazF dimer (Simanshu et al. *Mol. Cell* 2013; with Masayori Inouye lab at Robert Wood Johnson Medical School, NJ). The RNA in an extended alignment is bound along the RNA-binding interface between subunits of the MazF dimer. In addition, x-ray structural studies by our group of the *B. subtilis* MazF-MazE complex defined formation of a heterohexameric $(MazF)_2$-$(MazE)_2$-$(MazF)_2$ alignment, and established positioning of the C-terminal helical segment of MazE within the RNA-binding channel of the MazF dimer, thereby preventing mRNA binding and cleavage by MazF under non-stress conditions.

Publications: Patel Lab

Phan, A. T., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R. B. and Patel, D. J. (2011). Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. ***Nat. Struct, Mol. Biol.*** 18, 796-804.

Simanshu, D. K., Yamaguchi, Y., Park, J-H., Inouye, M. and Patel, D. J. (2013). Structural insights into mRNA recognition by toxin MazF and its regulation by antitoxin MazE in B. subtilis. ***Mol. Cell*** 52, 447-458.

Teplova, M. & Patel, D. J. (2008). Structural insights into RNA recognition by the alternate splicing regulator muscleblind-like MBNL1. ***Nat. Struct. Mol. Biol.*** 15, 1343-1351.

Teplova, M., Hafner, M., Teplov, D., Essig, K., Tuschl, T. and Patel, D. J. (2013). Structure-function studies of STAR family Quaking proteins bound to their *in vivo* RNA target sites. ***Genes Dev***. 27, 928-940.

Teplova, M., Malinina, L., Darnell, J. C., Song, J., Lu, M., Abagyan, R., Musunuru, K., Teplov, A., Burley, S. K., Darnell, R. B. & Patel, D. J. (2011). Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. ***Structure*** 19, 930-944.

Teplova, M., Song, J., Gaw, H. Y., Teplov, V. & Patel, D. J. (2010). Structural insights into RNA recognition by the CUG binding protein 1. ***Structure*** 18, 1364-1367.

Teplova, M., Wohlbold, L., Kim, N. Y., Izaurralde, E. & Patel, D. J. (2011). Structure-function studies of nucleocytoplasmic transport of retroviral genomic RNA by mRNA export factor TAP. ***Nat. Struct. Mol. Biol*****.** 18, 990-998.

Teplova, M., Yuan, Y. R., Phan, A. T., Malinina, L., Ilin, S., Teplov, A. & Patel, D. J. (2006). Structural basis for recognition and sequestration of UUU_{OH} 3'-terminii of nascent mRNA polymerase III transcripts by La autoantigen. ***Mol. Cell*** 21, 75-85.

8. LIGAND- AND PEPTIDE-RNA COMPLEXES

(1995 to 2007)

Peptide-RNA recognition plays a key role in processes ranging from transcription enhancement of viral RNAs to antitermination in bacteriophage systems. In several cases, complex formation can be investigated at the minimalist level by restricting studies to peptide recognition of a stem-loop RNA, making such systems amenable to structural characterization by heteronuclear NMR methods using isotopically ^{15}N,^{13}C-labeled peptide and RNA. Structural studies undertaken in our group have identified the role of β-hairpins and α-helices in targeting RNA stem-loops, together with positioning of the opposing RNA face for additional potential interactions.

Our group has also undertaken comprehensive NMR-based structural studies on HIV-1 Rev peptide bound to their natural and *in vitro*-selected RNA targets. The challenge has been to address the extent to which RNA architecture can dictate the conformation of the bound peptide from studies of a single peptide bound to two distinct RNA targets. Conversely, it remains to be determined whether peptide-mediated conformational switches can occur in RNA targets following studies of two distinct peptides bound to the same RNA target. By studying the complexes of basic arginine-rich viral peptides bound to their RNA targets, it should be possible to elucidate the diversity of pairing alignments between arginine side chains and guanosine base edges to complex formation.

Given the high affinity of ligand-RNA aptamer complexes, such systems constitute attractive targets for understanding the diversity of the code for intermolecular recognition. To this end, we initiated a program for study of cofactors and antibiotics to their *in vitro* selected RNA aptamer targets as part of an effort to define the principles associated with RNA folding and recognition. Such efforts could improve our understanding of the biological functions of RNA, as well as for the design of RNA sequences with specific catalytic activities. Aminoglycoside antibiotics are known to target complex RNA structural motifs such as the bacterial ribosome, spliceosome and catalytic RNAs. A question of special interest is how polycationic charged amines on the aminoglycosides are able to discriminate between alternate RNA folding topologies. A related question addressed whether antibiotics use the same recognition principles to target RNA aptamers as their natural target sites on ribosomal RNA.

Contributions

8.1 PEPTIDE-RNA COMPEXES

BIV Tat-TAR RNA Complex. In lentiviruses such as HIV and its bovine counterpart BIV, the activating (Tat) protein enhances transcription of the viral RNA by complex formation with the trans-activation response (TAR) RNA element. The NMR-based solution structure by our group of a 17-mer Tat peptide bound to a 28-mer TAR RNA established that the edge-wise orientation of the bound β-peptide hairpin of Tat penetrated deep into the RNA major groove on complex formation (Ye et al. *Chem. Biol*. 1995; see also Puglisi et al. *Science* 2005). The RNA fold was defined, in part, by two uracil bulged bases, one of which adopted a looped out conformation that widened the major groove, while the other forms a U•(A-U) triple that buttressed the RNA helix. A set of specific intermolecular hydrogen bonds between arginine side chains and the major groove edges of guanosine bases and backbone phosphates contributeed to sequence specificity.

P22 N-Peptide-boxB RNA Complex. The N-protein of lambdoid bacteriophages interacts with the N-utilization (nut) site on the nascent phage transcript, and together with host factors, plays an essential role in transcriptional anti-termination. Our NMR-based solution structure of the 20-mer P22 basic peptide of the N-protein bound to the 15-mer boxB RNA established that the N-peptide in a bent α-helical conformation packed tightly through hydrophobic and hydrogen bonding interactions against the major groove face of the boxB RNA hairpin, orienting the opposite face for potential interactions with host factors and/or RNA polymerase (Cai et al. *Nat. Struct. Biol.* 1998; see also Legault et al. *Cell* 1998). The guanidinium group of a key arginine is hydrogen-bonded to the guanosine of a loop-closing sheared G•A non-canonical pair and to an adjacent backbone phosphate, thereby contributing to sequence specificity of recognition.

8.2 PEPTIDE-RNA APTAMER COMPEXES

HIV-1 Rev Peptide-RNA Aptamer Complex. The viral protein, regulator of viral expression (Rev), controls nuclear to cytoplasmic export of intron-containing mRNAs that encode for viral structural proteins. *In vitro* selection approaches have identified RNA aptamers with a common consensus sequence that bind to the basic Rev peptide with even higher affinity than the Rev response element (RRE) RNA. The NMR-based solution structure by our group of the 17-mer HIV-1 Rev peptide bound to a high affinity 35-mer RNA aptamer established that the α-helical arginine-rich basic Rev peptide targeted a widened RNA major groove centered about adjacent G•A and reversed A•A non-canonical pairs (Ye et al. *Nat. Struct. Biol.* 1996; see also Battiste et al. *Science* 1996). The complex was stabilized by formation of a U•(A-U) triple, with the Hoogsteen-paired uracil base sandwiched between two arginines side chains.

RNA Architecture Dictates Bound HIV-1 Rev Peptide Conformation. NMR-based solution structural approaches have been applied by our group to compare complex formation of the arginine-rich peptide domain of HIV-1 Rev bound to two distinct RNA aptamers to determine whether RNA architecture can dictate the conformation of the bound peptide. Indeed, the HIV-1 Rev peptide adopts either an α-helical or extended conformation dependent on the RNA target on complex formation. Therefore, arginine-rich peptides can adapt distinct secondary folds to complement the tertiary folds of their RNA targets (Ye et al. *Chem. Biol*. 1999). This contrasts with protein-RNA complexes in which elements of RNA secondary structures adapt to fit within the tertiary folds of their protein targets.

HTLV-1 Rex Peptide-RNA Aptamer Complex. NMR-based solution structural studies by our group of the complex of the 16-mer Rex protein of HTLV-1 bound to a 33-mer RNA aptamer established that the peptide in a predominantly S-shaped extended conformation threads through a channel formed by the shallow and widened RNA major groove and a looped out guanosine (Jiang et al. *Structure* 1999). Binding specificity is associated with intermolecular hydrogen bonding between the guanidinium groups of three non-adjacent arginines and the guanosine base edges of three adjacent G-C pairs.

Peptide-triggered Conformational Switch in HIV-1 RRE RNA Complexes. To address the question as to how two different peptides recognize the same RNA target, NMR-based solution structural studies by our group were undertaken to compare the structures of the HIV-1 Rev peptide and an *in vitro* selected high affinity arginine-rich peptide both bound to the same stem IIB of the Rev response element (RRE) RNA. A peptide-triggered conformational switch was observed in the RRE RNA, whereby an unpaired uridine base points out into solvent in the Rev peptide-RRE RNA complex, but is stabilized inside the RNA deep groove by stacking with an arginine side chain in the *in vitro*-selected peptide –RRE RNA complex (Gosser et al. *Nat. Struct. Mol. Biol*. 2001).

8.3 LIGAND-RNA APTAMER COMPLEXES

Cofactor-RNA Aptamer Complexes. The solution-based NMR structure by our group of flavin mononucleotide (FMN) bound to a 35-mer RNA aptamer established zippering up of the internal loop and insertion of the planar isoalloxaxine ring of FMN between a G•G non-canonical pair and G•(U-A) triple (Fan et al. *J. Mol. Biol.* 1996). Specificity was associated with hydrogen bonding of the uracil like edge of the isoalloxazine ring of FMN to the Hoogsteen edge of an adenosine at the intercalation site.

The solution-based NMR structure by our group of AMP bound to a 40-mer RNA aptamer containing a 11-nt bulge established that the RNA adopted an L-shaped conformation in the bound state, with the AMP bound at the junction within a GNRA-like turn motif (Jiang et al. *Nature* 1996; see also Dieckmann et al. *RNA* 2006). The adeninose base of bound AMP was sandwiched between a pair of purines, with its Watson-Crick edge paired with the minor groove edge of a guanosine base.

Antibiotic-RNA Aptamer Complexes. Aminoglycoside antibiotics represent an important family of drugs that target ribosomal RNA by inhibiting protein synthesis. Our NMR-based solution structure of the aminoglycoside antibiotic tobramycin bound to a stem-loop RNA resulted in partial zippering up of the 14-nt hairpin-loop through formation of four Watson-Crick and one non-canonical pair on complex formation (Jiang and Patel, *Nat. Struct. Biol*. 1998) The bound tobramycin was sandwiched between the floor of the zippered-up widened major groove and a looped out guanosine base and anchored in place through a

network of intermolecular hydrogen bonds between amine groups of the antibiotic and base edges and phosphate groups of the RNA.

Aminoglycoside antibiotics are known to target complex RNA structural domains and discriminate between alternate folding topologies. The aminocyclitol antibiotic streptomycin has had a dramatic impact on medical practice and treatment. X-ray-based structural studies by our group were undertaken on a complex of streptomycin and an RNA aptamer containing a pair of asymmetrical internal loops. The RNA in the complex adopted a distinct divalent cation-stabilized fold involving a series of S-shaped backbone turns anchored by canonical and non-canonical pairs and triples (Tereshko et al. *Chem. Biol.* 2003). The streptose ring of bound streptomycin was encapsulated by stacked arrays from both loops at the elbow of the L-shaped RNA architecture. Specificity was achieved by direct hydrogen bonds between all streptose functional groups and base edges that line the inner walls of the cylindrical binding pocket.

The ribosome is a target for many antibiotics that interfere with bacterial protein synthesis by binding to ribosomal RNA components. Aminoglycoside antibiotics interact with the decoding site in 16S RNA, thereby decreasing the fidelity of translation. X-ray-based structural studies by our group of the complex of apramycin bound to the eukaryotic decoding site RNA established that the drug bound in the deep groove of the RNA within a continuously stacked helical segment comprising non-canonical C•A and G•A pairs and a bulged-out adenosine (Hermann et al. *Blood, Cells, Molecules and Diseases* 2007). The binding mode of apramycin to the human decoding-site RNA is distinct from aminoglycoside antibiotic recognition of the bacterial decoding site RNA.

Publications: Patel Lab

Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. & Patel, D. J. (1998). Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. ***Nat. Struct. Biol****.* 5, 203-212.

Fan, P., Suri, A. K., Fiala, R., Live, D. & Patel, D. J. (1996). Molecular recognition in the FMN-internal loop RNA aptamer complex. ***J. Mol. Biol.*** 258, 480-500.

Gosser, Y., Hermann, T., Majumdar, A., Hu, W., Frederick, R., Jiang, F., Xu, W. & Patel, D. J. (2001). Peptide-triggered conformational switch in HIV-1 RRE RNA complexes*.* ***Nat. Struct. Biol.*** 8, 146-150.

Hermann, T., Tereshko, V., Skripkin, E. & Patel, D. J. (2007). Apramycin recognition by the human ribosomal decoding site. ***Blood, Cells, Molecules, and Diseases*** 38, 193-198.

Jiang, F. Kumar, R. A., Jones, R. & Patel, D. J. (1996). Structural basis for RNA folding and recognition in the AMP-RNA aptamer complex. ***Nature*** 382, 183-186.

Jiang, F., Gorin, A., Hu, W., Majumdar, A., Baskerville, S., Ellington, A. & Patel, D. J. (1999). Anchoring an extended HTLV-1 Rex peptide within the RNA major groove containing junctional base triples, ***Structure* 7**, 1461-1472.

Jiang, L. & Patel, D. J. (1998). Solution structure of the tobramycin-RNA aptamer complex. ***Nat. Struct. Biol.*** 5, 769-774.

Tereshko, V., Skripkin, E. & Patel, D. J. (2003). Encapsulating streptomycin within a small 40-mer RNA. ***Chem. Biol*.** 10, 175-187.

Ye, X., Gorin, A., Ellington, A. D. & Patel, D. J. (1996). Deep penetration of an α-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. ***Nat. Struct. Biol.*** 3, 1026-1033**.**

Ye, X., Gorin, A., Frederick, R., Hu, R., Majumdar, A., Xu, W., McLendon, G., Ellington, A. & Patel, D. J. (1999). RNA architecture dictates bound peptide conformation. ***Chem. Biol****.* 6, 657-669.

Ye, X., Kumar, R. A. & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide - TAR RNA complex. ***Chem. Biol****.* 2, 827-840.

Complementary Publications: Other Labs

Battiste, J. L. et al. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. ***Science*** 273, 1547-1551.

Dieckmann, T. et al. (2006). Solution structure of an ATP-binding RNA aptamer reveals a novel fold. ***RNA*** 2, 628-640.

Legault, P. et al. (1998). NMR structure of the bacteriophage lambda N peptide/box B RNA complex: recognition of a GNRA fold by an arginine-rich motif. ***Cell*** 93, 289-299.

Puglisi, J. D. et al. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. ***Science*** 270, 1200-1203.

9. DNA LESIONS AND THEIR PROCESSING BY BYPASS POLYMERASES (1992 to 2011)

Benzo[a]pyrene [BP], a 'bay' region PAH, is metabolized in mammalian cells to highly reactive, mutagenic and tumorigenic chiral BP diol epoxide [BPDE] derivatives, which on covalent binding to cellular DNA, exhibit a spectrum of mutational and carcinogenic activities. BPDE can covalently bind to the exocyclic N^2-amino group of G to generate BP-N^2G adducts, whereby the BP ring is directed towards the minor groove of DNA. By contrast, benzo[c]phenanthrene [BPh], a more crowded 'fjord' region PAH, can covalently bind to the exocyclic N^6-amino group of A to generate BPh-N^6A adducts whereby the BPh ring is directed towards the major groove of DNA. The challenge to be addressed relates to how bulky covalent 'bay' and 'fjord' region PAHs are accommodated along the DNA helix and the extent to which contributions of chirality of the four stereo centers of the benzylic ring dictate these alignments. The resulting PAH alignments could occur with either minimal or substantial perturbations of the DNA helix and would in turn determine the potential for recognition and processing by the repair machinery.

There is a great deal of interest in the stalling of replicative polymerases at bulky DNA lesion sites and how such blockage is overcome by specific bypass polymerases, either in an error-free or error-prone manner. The challenge has been to identify mechanistic insights into the translocation mechanics during a cycle of binding and incorporation of nucleoside triphosphates opposite lesion sites. The focus of our lab has been on oxoG, [AF]G and me-N^1G/me-N^3C lesions and their processing by the bypass polymerase Dpo4, including formation of semitargeted mutations during the extension step.

Contributions

9.1 COVALENT CARCINOGENIC PAH DNA LESIONS

Minor Groove Alignment of BP-N^2G Adduct Without Disruption of Modified Pair. Synthesis and purification protocols have allowed generation of the highly tumorigenic (+)-*trans-anti*-[BP]G adduct in a central (C-[BP]G-C)•(G-C-G) sequence context at the 11-mer duplex level (collaboration on synthesis and structural studies with the Nicholas Geacintov and Suse Broyde labs at New York University). Our NMR structural studies undertaken on this site-specifically incorporated chiral lesion established that the BP ring is positioned in the minor groove and directed towards the 5'-end of the modified strand, with one face of the BP ring facing towards the minor groove while the opposite face is exposed to solvent (Cosman et al. *Proc. Natl. Acad. Scis. USA* 1992). The helix is minimally perturbed through widening of the minor groove centered about the binding site without impacting on the Watson-Crick alignment of the [BP]G-C pair at the lesion site, as well as flanking G-C base pairs.

Our structural studies were next extended to the tumorigenic (-)-*trans-anti*-[BP]G adduct in which the chirality is inverted at all four positions on the benzylic ring of BP relative to the (+)-*trans-anti*-[BP]G adduct. NMR structural studies of the (-)-*trans-anti*-[BP]-G adduct in the same sequence context established that the BP ring spans both strands in the minor groove of the minimally perturbed duplex and was directed towards the 3'-end of the modified strand (de los Santos et al. *Biochemistry* 1992). Both the [BP]G-C and flanking G-C base pairs retained their integrity. The observed opposite orientational alignments between (+)-*trans-anti*- and (-)-*trans-anti*-[BP]G adducts originate in the chiral characteristics of the two BPDE enantiomers and may have profound influences on the interaction of cellular enzyme systems with these opposing alignments of this bulky lesion.

Intercalation of the BP-N^2G Adduct with Base Displacement. Our NMR-based studies next focused on the (+)-*cis-anti*-[BP]-G adduct in which the chirality is inverted solely at the covalent linkage position on the benzylic ring of BP relative to the (+)-*trans-anti*-[BP]G adduct. NMR structural studies of the (+)-*cis-anti*-[BP]G adduct in the central (C-[BP]G-C)•(G-C-G) sequence context established that the BP ring was intercalated from the minor groove direction between flanking intact G-C base pairs. The modified guanosine of [BP]G was displaced into the minor groove with its plane parallel to the helix axis, resulting in disruption of the [BP]G-C pair (Cosman et al. *Biochemistry* 1993a).

Related NMR-based structural studies by our group on the (-)-*cis-anti*-[BP]G adduct in the same sequence context established that the BP ring is intercalated from the major groove direction between

flanking intact G-C base pairs, while displacing the modified guanosine of [BP]G into the major groove, with its plane parallel to the helix axis, resulting in disruption of the [BP]G-C pair (Cosman et al *Biochemistry* 1996). Thus, inversion of chirality at all four positions on the benzylic ring of BP results in opposite orientations of the BP lesion for the intercalation with base displacement alignment, as it does for the minor groove alignment outlined above.

Intercalation of the BPh-N^{6}A Adduct Without Disruption of Modified Pair. Our NMR-based structural studies of the more crowded 'fjord' region (+)-*trans-anti*-[BPh]A adduct in the central (C-[BPh]A-C)•(G-T-G) sequence context established that the BPh ring intercalated to the 5'-side of the [BPh]A lesion site without disruption of either G-C or [BPh]A-T base pairs on either side of the intercalation site (Cosman et al. *Biochemistry* 1993b). This unidirectional intercalation without disruption of the modified base pair is achieved through buckling of the [BPh]A-T base pair, displacement of the covalent linkage bond between BPh and A from the plane of the A base, the specific pucker adopted by the benzylic ring of BPh, and most importantly, the propeller-like non-planar geometry for the aromatic ring system of BPh.

Our follow up NMR-based structural studies on the (-)-*trans-anti*-[BPh]A adduct in the same sequence context established that the BPh ring intercalated to the 3'-side of the [BPh]A lesion site without disruption of either G-C or [BPh]A-T base pairs on either side of the intercalation site (Cosman et al. *Biochemistry* 1995). Once again, inversion of chirality at all four positions on the benzylic ring results in opposite orientations of the BPh lesion for 'intercalation without disruption of the modified pair' alignment (reviewed in Geacintov et al. *Chem. Rev. Toxicol*. 1997).

9.2 ADDITIONAL COVALENT CARCINOGENIC LESIONS

Aminofluorene-C^{8}G-DNA Adducts. The aromatic amine 2-aminofluorene [AF] and its acetyl counterpart [AAF] are potent carcinogens that are present in the environment in tobacco smoke, automobile exhaust and as byproducts of various industrial processes. The major *in vivo* adducts are to the C^8 position of guanosine to form [AF]-C^{8}G and [AAF]-C^{8}G lesions. Overall, our NMR-based structural studies have identified two distinct motifs adopted by the [AF]G lesion positioned opposite C, with their distribution dependent on flanking sequence. These alignments are major groove positioning of the AF ring without disruption of the modified [AF]G-C and flanking base pairs, and intercalation of the [AF] ring between intact flanking base pairs and displacement of the modified guanosine into the major groove (reviewed in Patel et al. *Chem. Res. Toxicol.* 1998; with Suse Broyde lab, New York University). The equilibrium shifted towards the [AF]-intercalated with base displacement conformation for cases where the [AF]G is positioned opposite a deletion site, as well as for the case where the [AF]G adduct is replaced by the [AP]G adduct, that involved by the larger aromatic ring system of 1-aminopyrene. The working hypothesis has been that the non-distorting external alignment of the lesion might represent structures that escape repair, while distorting conformations that involve intercalation of the planar lesion are accessible and processed by the repair machinery.

Tamoxifin-N^{2}G Adduct. The non-steroid anti-estrogen tamoxifen [TAM] is a potent adjunct chemotherapeutic agent for treatment of breast cancer. Tamoxifen has been shown to form covalent complexes with DNA and its stereoisomeric adducts can be mutagenic, with their activities being conformation dependent. Our NMR-based structural studies were undertaken on the covalent *cis*-isomer (S-epimer) of [TAM]-N^{2}G in the (C-[TAM]G-C)•(G-C-G) sequence context at the 11-mer duplex level. The S-*cis* TAM lesion is accommodated within a widened minor groove on both sides of the lesion site without disruption of either the Watson-Crick aligned [TAM]G-C pair or flanking G-C pairs (Shimotakahara et al. *J. Mol. Biol*. 2000; with Geacintov and Broyde labs). The DNA helix axis is bent by 30 degrees and is directed away from the minor groove adduct site, thereby potentially compromising the fidelity of the minor groove polymerase scanning machinery.

9.3 PROCESSING OF DNA DAMAGE LESIONS BY BYPASS POLYMERASES

The structural biology research on lesion processing by bypass polymerases has been pioneered by the Wei Yang laboratory based on their original studies with the Dpo4 bypass polymerase (see Ling et al. 2001, 2003, 2004)

Bypass of oxoG Lesions. 7,8-dihydro-8-oxo-guanosine (oxoG), the predominant lesion formed following oxidative damage of DNA by reactive oxygen species, is processed differently by replicative and bypass polymerases. Our kinetic primer extension studies demonstrate that the bypass polymerase Dpo4 preferentially inserts C opposite oxoG, and also preferentially extends from the oxoG-C base pair, thus achieving error-free bypass of this lesion. Our crystal structures of preinsertion binary, insertion ternary and postinsertion binary complexes of oxoG-modified template-primer DNA and Dpo4 have provided detailed insights into the translocation mechanics of the bypass polymerase accompanying nucleoside triphosphate binding and covalent nucleobase incorporation during a complete cycle of nucleotide incorporation (Rechkoblit et al. *PLoS Biol*. 2006; with Geacintov and Broyde labs).

Bypass of AF-C^8-G Lesions. The aromatic amine carcinogen 2-aminofluorene forms adducts at the C^8 position of guanosine ([AF]-C^{8}G) that are bypassed by the Y-family polymerase Dpo4 via error-free and error-prone mechanisms. Our crystal structures of the Dpo4 ternary complex, with the 3'-terminal primer C base opposite [AF]G in the *anti* conformation and with the AF moiety positioned in the major groove, reveal both accurate and misalignment-mediated mutagenic extension pathways (Rechkoblit et al. *Nat. Struct. Mol. Biol.* 2010; with Broyde and Geacintov labs). Further extension leads to semitargeted mutations via this proposed polymerase-guided mechanism.

Bypass of Base-modified me-N^{1}G and me-N^{3}C Lesions. DNA is susceptible to alkylation damage that modifies the Watson-Crick edge of the bases. The bypass polymerase Dpo4 is strongly inhibited by 1-methylguanosine (me-N^{1}G) and 3-methylcytidine (me-N^{3}C) adducts, with nucleotide incorporation opposite these lesions being predominantly mutagenic, and extension after insertion of both correct and incorrect bases introducing additional base substitution and deletion errors. Our crystal structures of Dpo4 ternary extension complexes with correct and mismatched 3'-terminal primer bases opposite the lesions revealed that both me-N^{1}G and me-N^{3}C remain positioned inside the helix, while both correct and incorrect pairing partners exhibit pronounced primer terminal nucleotide distortion, being primarily evicted from the DNA helix when opposite me-N^{1}G or misaligned when pairing with me-N^{3}C (Rechkoblit et al. *Structure* 2011; with John Essigmann lab, MIT).

Publications: Patel Lab

Cosman, M. Hingerty, B. E., Luneva, N., Amin, S. Geacintov, N. E., Broyde, S. & Patel, D. J. (1996). Solution conformation of the (-)-*cis*-*anti*-benzo[a]pyrenyl-dG adduct opposite dC in a DNA duplex: Intercalation of the covalently attached BP ring into the helix with base displacement of the modified deoxyguanosine into the major groove. ***Biochemistry*** 35, 9850-9863.

Cosman, M., de los Santos, C., Fiala, R., Hingerty, B. E., Ibanez, V., Luna, E., Harvey, R., Geacintov, N. E., Broyde, S., and Patel, D. J. (1993a). Solution conformation of the (+)-*cis*-*anti*-[BP]dG adduct in a DNA duplex: Intercalation of the covalently attached benzo[a]pyrenyl ring into the helix and displacement of the modified deoxyguanosine. ***Biochemistry*** 32, 4145-4155.

Cosman, M., de los Santos, C., Fiala, R., Singh, S. B., Hingerty, B. E., Ibanez, V., Margulis, L. A., Live, D., Geacintov, N. E., Broyde, S., and Patel, D. J. (1992). Solution conformation of the major adduct between the carcinogen (+)-*anti*-benzo[a]pyrene diol epoxide and DNA. ***Proc. Natl. Acad. Scs. USA***. 89, 1914-1918.

Cosman, M., Fiala, R., Hingerty, B. E., Laryea, A., Lee, H., Harvey, R. G., Amin, S., Geacintov, N. E., Broyde, S. and Patel, D. J. (1993b). Solution conformation of the (+)-*trans*-*anti*-[BPh]dA adduct opposite dT in a DNA duplex: Intercalation of the covalently attached benzo[c]phenanthrene to the 5'-side of the adduct site without disruption of the modified base pair. ***Biochemistry*** 32, 12488-12497.

Cosman, M., Laryea, A., Fiala, R., Hingerty, B. E., Amin, S., Geacintov, N. E., Broyde, S., & Patel, D. J. (1995). Solution conformation of the (-)-*trans*-*anti*-[BPh]dA opposite dT adduct in a DNA duplex: Intercalation of the covalently attached benzo[c]phenanthrenyl ring to the 3'-side of the adduct site and comparison with the (+)-trans-anti[BPh]dA opposite dT stereoisomer. ***Biochemistry*** 34, 1295-1307.

de los Santos, C., Cosman, M., Hingerty, B. E., Ibanez, V., Margulis, L. A., Geacintov, N. E., Broyde, S. and Patel, D. J. (1992). Influence of chirality on solution conformations of benzo[a]pyrene-diol-epoxide-DNA covalent adducts: The (-)-*trans*-*anti*-[BP]G•C adduct structure and comparison with the (+)-*trans*-*anti* enantiomer. ***Biochemistry*** 31, 5245-5252.

Geacintov, N. E., Cosman, M., Hingerty, B. E., Amin, S., Broyde, S. and Patel, D. J. (1997). NMR solution

structures of stereoisomeric covalent polycyclic aromatic carcinogen-DNA adducts: Principles, patterns and diversity. ***Chem. Res. Toxicol***.10, 111-146.

Patel, D. J., Mao, B., Gu, Z., Hingerty, B. E., Gorin, A., Basu, A. K. and Broyde, S. (1998). NMR solution structures of covalent aromatic amine-DNA adducts and their mutagenic relevance. ***Chem. Res. Toxicol.*** 11, 391-407.

Rechkoblit, O., Delaney, J. C., Essigmann, J. M. & Patel, D. J. (2011). Implications for damage recognition during Dpo4-mediated mutagenic bypass of m1G and m3C lesions. ***Structure*** 19, 821-832.

Rechkoblit, O., Kolbanovskiy, A., Malinina, L., Geacintov, N. E., Broyde, S. & Patel, D. J. (2010). Mechanism of error-free bypass and semi-targeted mutagenic bypass of an aromatic amine lesion by Y-family polymerase Dpo4. ***Nat. Struct. Mol. Biol*.** 17, 379-388.

Rechkoblit, O., Malinina, L., Cheng, Y., Kuryavyi, V., Broyde, S., Geacintov, N. & Patel, D. J. (2006). Stepwise translocation of Dpo4 polymerase during error-free bypass of oxoG lesion. ***PLoS Biology*** 4, 25-42.

Shimotakahara, S., Gorin, A., Kolbanovskiy, A., Kettani, A., Hingerty, B., Amin, S., Broyde, S., Geacintov, N. & Patel, D. J. (2000). Accommodating covalently bound S-*cis*-tamoxifen-N^2-guanine adduct in a widened DNA minor groove. ***J. Mol. Biol.*** 302, 377-393.

Complementary Publications: Other Labs

Ling, H. et al. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. ***Cell*** 107, 91-102.

Ling, H. et al. (2003). Replication of a cis-syn thymine dimer at atomic resolution. ***Nature*** 424, 1083-1087.

Ling, H. et al. (2004). Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with DNA polymerase. ***Proc. Natl. Acad. Scis. USA*** 24, 2265-2269.

10. MULTI-STRANDED DNA ARCHITECTURES AND IMPACT OF HELICAL IMPERFECTIONS (1975 to 2012)

High resolution NMR is a powerful technique to study the conformations and dynamics of nucleic acids in solution. We have applied this technique to study the fundamentals of DNA duplex structure and dynamics such as fraying at the ends of helices, as well as to search for and identify new conformations such as dinucleotide repeats characteristic of alternate DNA handedness, and non-canonical pairings that stabilize formation of parallel-stranded DNA. The NMR method is also amenable to the monitoring of consequences related to helical imperfections in DNA such as mismatches, bulges and junctions and has the ability to define the helical distortions and altered pairing alignments required for accommodation of such modifications.

We have applied exchangeable and nonexchangeable proton NMR to initially monitor formation of pyrimidine•(purine-pyrimidine) and purine-rich•(purine-pyrimidine) triplexes, and subsequently used measurable NOE distance connectivity data to define their conformations in solution. These structures have identified the directionalities of the third strand, the base triple pairing alignments and the helical distortions associated with incorporation of non-isomorphous base triples.

We have also developed NMR-based techniques and approaches to define the solution structures of G-quadruplexes containing G•G•G•G tetrads (G-tetrads). Such efforts have identified novel folding topologies including double-chain reversal loops, V-shaped scaffolds, mixed G•C•G•C and A•T•A•T tetrads and A•(G•G•G•G)•A hexad platforms. The diversity of G-quadruplex folding topologies have emerged from our structural studies of G-rich tracts in the human telomere sequence, in c-myc and c-kit promoters, and in intronic sequences. We have also demonstrated the capacity for monomeric and dimeric G-quadruplexes to target protein scaffolds, including a basic channel/canyon formed between subunits of a dimer of dimers of HIV-1 integrase.

Contributions

10.1 DNA DUPLEXES, HANDEDNESS AND HELICAL IMPERFECTIONS

Fraying at the Ends of DNA Duplexes. Exchangeable imino proton (guanosine N^1H and thymidine N^3H) NMR resonances were monitored by our group for fraying effects at the ends of a self-complementary d(ATGCAT) hexanucleotide duplex in H_2O solution. Analysis of imino proton NMR chemical shifts and line width changes as a function of temperature yielded thermodynamic parameters for the fraying process (Patel and Hilbers, *Biochemistry* 1975). The symmetry-related internal A-T base pairs break in a coupled manner at low ionic strength, with coupling removed in the presence of Mg^{2+} cations. By contrast, the symmetry-related terminal A-T base pairs break independently of each other both in the absence and presence of Mg^{2+} cations.

Alternating Dinucleotide Repeats Adopted by d(C-G)$_n$ in High Salt. Proton and phosphorus NMR spectra of the d(C-G)$_n$ duplex have been recorded by our group at low and 4 M NaCl conditions to monitor the salt-dependent structural transition that resulted in the inversion of the circular dichroism spectrum. The NMR spectra in high salt conditions required the d(C-G)$_n$ duplex to adopt an alternating DNA conformation for which the symmetry repeats every two base pairs (Patel et al. *Proc. Natl. Acad. Scis. USA.* 1979). More specifically, the chemical shift parameters demonstrate that every other glycosidic torsion angle and phosphodiester linkage adopted a different conformation than that observed in B-DNA.

Subsequent crystal structural studies from the Alexander Rich lab (MIT) established that d(C-G)$_3$ adopts a left-handed Z-helix with a dinucleotide repeat (see Wang et al. *Science* 1981). The observed *syn* glycosidic angle at guanosine and *anti* glycosidic angle at cytidine observed in the crystal structure was validated by our NMR-based nuclear Overhauser enhancement (NOE) measurements in solution (Patel et al. *Proc. Natl. Acad. Scis. USA*. 1982). Further, phosphorus NMR can monitor the transition for d(C-G)$_n$ between low-salt B-DNA (single phosphorus resonance) and high-salt Z-DNA (two resolved phosphorus resonances) forms.

Impact of Helical Imperfections in DNA Duplexes. The relaxation lifetime of imino protons were monitored by our group using saturation recovery NMR experiments on a self-complementary d(CGCGAATTCGCG) 12-mer duplex, and its counterparts containing symmetrically positioned G•T Wobble pairs and adenosine bulge bases. Comparison of the resulting imino proton exchange lifetimes showed that the destabilizing effect of the G•T mismatch pair affected the opening rate of only the nearest-neighbor base pairs, while the extra adenosine had a more pronounced impact and affected the opening rate of all base pairs (Pardi et al. *Biochemistry* 1982; with Ignatio Tinoco lab at UC-Berkeley).

Earlier NMR and x-ray studies on DNA duplexes containing G•A non-canonical pairs had identified G(*anti*)•A(*anti*) and G(*anti*)•A(*syn*) pairing alignments. By contrast, our NMR studies demonstrated a pH dependent switch from G(*anti*)•A(anti) pairing at neutral pH to protonated G(*syn*)•A(*anti*) pairing at acidic pH (Gao and Patel, *J. Am. Chem. Soc*. 1988). The observed variability of repair efficiencies for G•A non-canonical pairs in DNA duplexes could reflect more than one type of pairing alignment and their differential recognition by the repair system.

Impact of Bulges and Junctions on DNA Topology. Our solution NMR-based studies on a DNA duplex containing an ATA three-base bulge loop established that all three unpaired bases adopt *anti* conformations and are stacked into the duplex (Rosen et al. *Biochemistry* 1992). The duplex is significantly kinked at the bulge site, with bending occurring in the direction away from the bulge-containing strand.

Solution NMR-based studies have been undertaken by our group to examine the stability of junctional base pairs in a three-way junction with two unpaired cytidines at the branch point. The experimental data supported formation of two of the three junctional base pairs, as well as a pair-wise stacking arrangement between two of the three helices in the junction (Rosen et al. *Biochemistry* 1993). In three-way junctions, the presence of unpaired bases at the branch point, facilitated simultaneous formation of base pairing and stacking within the branch point.

Parallel-stranded DNA Duplexes. Solution NMR studies on the self-complementary d(TCGA) duplex at acidic pH by our group established formation of a parallel-stranded duplex containing all *anti* glycosidic torsion angles stabilized by reversed C•C^+, reversed G•G (aligned through minor groove edges) and reversed A•A (aligned through major groove edges) non-canonical pairs (Wang and Patel, *J. Mol. Biol.* 1994; see also Robinson et al. *Biochemistry* 2002). The parallel-stranded duplex is stabilized by intra-strand base stacking at the C-G step and inter-strand base stacking at the G-A step.

10.2 DNA TRIPLEXES

NMR-based Characterization of DNA Triplexes. Solution proton NMR spectra have been recorded by our group on pyrimidine•(purine-pyrimidine) DNA triplexes containing one oligopurine and two oligopyrimidine strands at acidic pH. The observed NOE distance constraints establish that the extra oligopyrimidine strand resides in the major groove through Hoogsteen pairing, is aligned parallel to the Watson-Crick aligned purine strand, with the triplex stabilized through formation of T•(A-T) and C^+•(G-C) base triples (de los Santos et al. *Biochemistry* 1989; see also Rajagopal and Feigon, *Nature* 1989). The results demonstrate that pyrimidine•(purine-pyrimidine) DNA triple helices can be monitored at the individual base-triple level with distinct markers differentiating between Watson-Crick and Hoogsteen pairing alignments.

Solution proton NMR studies by our group have been recorded on purine-rich•(purine-pyrimidine) DNA triplexes containing two oligopurine and one oligopyrimidine strands at neutral pH. The observed NOE distance connectivities establish that the extra oligopurine-rich strand resides in the major groove through Hoogsteen pairing, is aligned anti-parallel to the Watson-Crick aligned purine strand, with the triplex stabilized through G•(G-C), A•(A-T) and T•(A-T) base triples (Radhakrishnan et al. *J. Mol. Biol.* 1991). The results demonstrate that Watson-Crick and Hoogsteen pairing can be readily monitored at the individual base-triple level in purine-rich•(purine-pyrimidine) DNA triple helices.

DNA Triplexes Containing Pyrimidine Third Strands. The NMR-based solution structure has been determined by our group for a pyrimidine•(purine-pyrimidine) DNA triplex containing a guanosine residue in the otherwise all pyrimidine third strand capable of G•(T-A) base triple formation. The formation of the G•(T-A) triple within an otherwise pyrimidine•(purine-pyrimidine) DNA triplex formed by T•(A-T) and

C^+•(G-C) base triples resulted in conformational realignments in and around the G•(T-A) triple (Radhakrishnan and Patel, *Structure* 1994).

DNA Triplexes Containing Purine-rich Third Strands. The NMR-based solution structure has been determined by our group for a purine-rich•(purine-pyrimidine) DNA triplex containing guanosines, adenosines and thymidines in the purine-rich third strand. The structure defined the G•(G-C), A•(A-T) and T•(A-T) pairing alignments in the purine-rich•(purine-pyrimidine) DNA triplex and identified the structural discontinuities in the third purine-rich strand associated with the non-isomoporphism of the G•(G-C) and A•(A-T) versus T•(A-T) base triples (Radhakrishnan and Patel, *Structure* 1993).

10.3 DNA G-QUADRUPLEXES

NMR-based Characterization of G-Quadruplexes. Proton NMR studies of G-rich containing DNA sequences in H_2O exhibited guanosine imino protons between 10.5 and 12.0 ppm, as well as identified NOE distance connectivities between imino and H8 protons of guanosines, characteristic features of G•G•G•G tetrad formation through Hoogsteen pairing alignment around the G-tetrad (Wang et al. *J. Mol. Biol*. 1991; see also Smith and Feigon, *Nature* 1992). Internal G-tetrads within a G-quadruplex exhibited much slower guanosine imino proton exchange rates compared to their terminal G-tetrad counterparts. The magnitude of the distance constraints between the guanosine imino proton and its own H8 proton readily allowed identification of *anti* versus *syn* glycosidic torsion angles. These studies established that guanosines adopted all *anti* glycosidic torsion angles for parallel-stranded G-quadruplexes (Wang and Patel, *Biochemistry* 1992), while a mixture of *anti* and *syn* glycosidic torsion angles were observed for anti-parallel aligned G-quadruplexes (Wang and Patel, *Structure* 1993).

G-Quadruplex Scaffolds Adopted by the Human Telomeric Repeat. Our NMR-based structural studies have defined the solution structure of the four-repeat human telomere sequence $d(AG_3(T_2AG_3)_3]$ in Na^+ cation solution. This sequence, which contains four AG_3 repeats, folded intramolecularly into a G-quadruplex stabilized by three stacked G-tetrads, which are connected by two edge-wise loops and a central diagonal loop (Wang and Patel *Structure* 1993). The alignments around individual G-tetrads are *syn•syn•anti•anti*, with each G-rich strand having adjacent strands aligned both parallel and anti-parallel to it, resulting in one wide, two medium and one narrow grooves for the fold of this G-quadruplex.

The NMR-based solution structure of an intramolecular G-quadruplex formed by the four-repeat human telomere $d(T_2AG_3)_4$ sequence in physiologically relevant K^+ solution by our group adopted a (3+1) folding topology, where three G tracts are oriented in one direction and the fourth in the opposite direction. Further, two of the connecting loops are of the double-chain-reversal type and one of the edge-wise type, resulting in two *syn•anti•anti•anti* G-tetrads and one *anti•syn•syn•syn* G-tetrad (Luu et al. *J. Am. Chem. Soc*. 2006; see also Ambrus et al. *Nucleic Acids Res*. 2006; Xu et al. *Biorg. Med. Chem*. 2006). Thus, the four T_2AG_3 repeat-containing human telomere G-quadruplex adopts distinct topologies in Na^+ and K^+ containing solutions.

G-Quadruplex Scaffolds Formed by the G-rich c-myc Promoter. The nuclease-hypersensitivity element III in the c-myc promoter remains a promising anti-cancer target since it largely controls transcriptional activation of the c-myc oncogene. Our NMR-based solution studies established that the four G-rich tracts in the human c-myc promoter form an intramolecular propeller-type parallel-stranded G-quadruplex in K^+-containing solution (Phan et al. *J. Am. Chem. Soc*. 2004; see also Ambrus et al. *Biochemistry* 2005). The G-quadruplex topology contains a core of three-stacked *anti•anti•anti•anti* G-tetrads and three double-chain-reversal loops bridging three G-tetrad layers, two of which contain a single connecting residue.

Extension of our NMR-based solution studies to five G-rich tracts in the human c-myc promoter established formation of a G-quadruplex involving a core of three stacked G-tetrads formed by four parallel G-tracts with all *anti* guaosnines and a snapback 3'-end *syn* guanosine (Phan et al. *Nat. Chem. Biol*. 2005). The cationic porphyrin TMPyP4 was shown to stack over a terminal G-tetrad and stabilize this G-quadruplex fold.

G-Quadruplex Scaffolds Formed by the G-rich c-kit Promoter. The c-kit oncogene is an important target in the treatment of gastrointestinal tumors. Our NMR-based studies have determined the solution structure of the intramolecular G-quadruplex formed by the four G-rich tracts of the human c-kit1

promoter in K^+ solution. Unexpectedly, an isolated guanosine is involved in G-tetrad core formation to generate an all parallel-stranded G-quadruplex (Phan et al. *J. Am. Chem. Soc.* 2007; see also Wei et al. *Nucleic Acids Res*. 2012). The topology is further defined by two single-residue double-chain-reversal loops, a five-residue stem-loop and a two-residue loop, thereby providing a highly specific platform for small molecule targeting of this scaffold.

Extension of the NMR-based solution studies by our group have resulted in the solution structure determination of the intramolecular dimeric G-quadruplex formed by the four G-rich tracts of the human c-kit2 promoter in K^+ solution. This dimeric G-quadruplex adopted an all parallel-stranded topology where individual G-rich strands span a pair of three stacked G-tetrad layers separated by a sandwiched non-canonical A•A pair (Kuryavyi et al. *Nucleic Acids Res*. 2010). This dimeric G-quadruplex topology has potential implications for strand exchange during recombination

G-Quadruplexes Containing G•C•G•C and A•T•A•T Tetrads. The fragile X syndrome $d(CGG)_n$ triplet repeat has been shown to form a stable G-quadruplex in monovalent cation solution. NMR-based structural studies on $d(GCGGT_3GCGG)$ in Na^+ cation solution established formation of a G-quadruplex through dimerization of two symmetry-related hairpins containing edge-wise T_3 loops at opposite ends. The G-quadruplex topology contains internal all-*anti* G•C•G•C tetrads sandwiched between terminal G•G•G•G tetrads (Kettani et al. *J. Mol. Biol*. 1995). The G•C•G•C tetrads are formed through bifurcated hydrogen bonding along the major groove edges through alignment of a pair of Watson-Crick G-C pairs.

Our NMR-based structural studies have now identified G-quadruplexes that are composed of a G•G•G•G tetrad, a G•C•G•C tetrad and a slipped A•T•A•T tetrad, with the latter formed by alignment of Watson-Crick A-T base pairs along the major groove edges of opposing adenine residues (Zhang et al. *J. Mol. Biol*. 2001). These studies highlight the versatility of the G-quadruplex architecture, a topology accessible to a range of mixed tetrad alignments in addition to the canonical G-tetrad.

G-Quadruplexes Containing A•(G•G•G•G)•A Hexads. Our NMR-based structural studies on the guanosine- and adenosine-rich d(GGAGGAG) sequence containing GGA repeats has identified A•(G•G•G•G)•A hexad formation involving adenosine recognition of the exposed minor groove edge of opposing guanosines within a G•G•G•G tetrad through sheared G•A non-canonical pair formation (Kettani et al. *J. Mol. Biol*. 2000). Such a A•(G•G•G•G)•A hexad provides a stable platform for stacking interactions.

G-Quadruplex Formed by a Human Intronic G-rich Sequence. Our NMR-based solution structure of the human ch1 intronic sequence composed of four G-rich tracts adopted a unique intramolecular G-quadruplex scaffold where the first guanosine is positioned within the central G-tetrad and also contains a V-shaped loop that spans three G-tetrad planes without containing a bridging nucleotide (Kuryavyi and Patel, *Structure* 2010). In addition, a guanosine base is intercalated between an extended GG step, a feature observed in common with the catalytic site of group I introns.

pilE G-rich Sequence Forms a G-Quadruplex Essential for pilin Antigenic Variation. The four G-rich tract sequence located upstream of the *N. gonorrhoeae* pilin expression locus (pilE) was shown to be necessary for initiation of pilin antigenic variation, a recombinant based, high-frequency, diversity-generating system. Our NMR-based solution studies established that the pilE sequence adopts all-parallel stranded monomeric and 5'-end-stacked dimeric G-quadruplexes in monovalent cation solution (Kuryavyi et al. *Structure* 2012; with Hans Seifert, Northwestern University School of Medicine). The monomeric G-quadruplex containing single residue double-chain-reversal loops can be modeled without steric clashes into the 3-nt DNA-binding site of RecA, thereby potentially facilitating specialized recombination reactions leading to pilin diversification.

Interlocked G-Quadruplex Inhibitors of HIV-1 Integrase. A four G-rich tract sequence named 93del was identified by *in vitro* selection based on its potent nanomolar inhibition of HIV-1 integrase. Our NMR-based solution studies established formation of an unusually stable dimeric G-quadruplex in K^+ solution. Each monomeric unit involved an all parallel-stranded alignment, where all guanosines are *anti* except for G1, which is *syn*, and where an A•(G•G•G•G) pentad is sandwiched between two G-tetrads. The interlocked dimer formation was achieved by G1 of one monomer forming a G-tetrad with three other guanosines from the other monomer (Phan et al. *Proc. Natl. Acad. Scis. USA*. 2005; with the Simon Litvak and Marie Andreola lab at University of Bordeaux). We proposed a model based on a molecular

docking approach where the interlocked dimeric 93del G-quadruplex was positioned within a basic channel/canyon formed between subunits of a dimer of dimers of HIV-1 integrase.

Publications: Patel Lab

de los Santos, C., Rosen, M. and Patel, D. J. (1989). NMR studies of DNA $(R+)_n•(Y-)_n•(Y+)_n$ triple helices in solution: Imino and amino proton markers of T•(A-T) and C^+•(G-C) base triple formation. ***Biochemistry*** 28, 7282-7288.

Gao, X., and Patel, D. J. (1988). G(syn)•A(anti) mismatch formation in DNA dodecamers at acidic pH: pH dependent conformational transition of G•A mispairs detected by proton NMR. ***J. Am. Chem. Soc.*** 110, 5178-5182.

Kettani, A., Gorin, A., Majumdar, A., Skripkin, E., Zhao, H., Jones, R. and Patel, D. J. (2000). A dimeric DNA interface stabilized by stacked A•(G•G•G•G)•A hexads and coordinated monovalent cations. ***J. Mol. Biol.*** 297, 627-644.

Kettani, A., Kumar, R. A., & Patel, D. J. (1995). Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. ***J. Mol. Biol.*** 254, 638-656.

Kuryavyi, V. & Patel, D. J. (2010). Solution structure of an unique G-quadruplex scaffold adopted by a guanosine-rich human intronic sequence. ***Structure*** 18, 73-82.

Kuryavyi, V., Cahoon, L. A., Seifert, H. S. & Patel, D. J. (2012). RecA-binding *pilE* G4 sequence essential for pilin antigenic variation forms parallel-stranded monomeric and 5'-end stacked dimeric G-quadruplexes. ***Structure*** 20, 2090-2102.

Kuryavyi, V., Phan, A. T. & Patel, D. J. (2010). Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human *c-kit2* promoter. ***Nucleic Acids Res.*** 38, 6757-6773.

Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L. & Patel, D. J. (2006). Structure of the human telomere in K^+ solution: An intramolecular (3+1) G-quadruplex scaffold. ***J. Am. Chem. Soc.*** 128, 9963-9970.

Pardi, A., Morden, K. M., Patel, D. J., and Tinoco, I., Jr. (1982). Kinetics for exchange of imino protons in the d(CGCGAATTCGCG) double helix and in two similar helices that contain a G•T base pair d(CGTGAATTCGCG) and an extra adenine d(CGCAGAATTCGCG). ***Biochemistry*** 2**1**, 6567-6574.

Patel, D. J., and Hilbers, C. W. (1975). Proton NMR investigations of fraying in double-stranded d(ATGCAT) in H_2O solution. ***Biochemistry*** 14, 2651-2656.

Patel, D. J., Canuel, L. L. and Pohl, F. M. (1979). 'Alternating B-DNA' conformation for the oligo(dG-dC) duplex in high salt solution. ***Proc. Natl. Acad. Scis. USA*** 76, 2508-2511.

Patel, D. J., Kozlowski, S. A., Nordheim, A., and Rich, A. (1982). Right-handed and left-handed DNA: Studies of B-DNA and Z-DNA by using proton nuclear Overhauser effect and phosphorus NMR. ***Proc. Natl. Acad. Scis. USA.*** 79, 1413-1417.

Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S. & Patel, D. J. (2007). Structure of an unprecedented G-quadruplex scaffold adopted by the human *c-kit* promoter. ***J. Am. Chem. Soc.*** 129, 4386-4392.

Phan, A. T., Kuryavyi, V., Gaw, H. Y. & Patel, D. J. (2005). Small molecule interaction with a five-guanine tract G-quadruplex structure from the human c-myc promoter. ***Nat. Chem. Biol.*** 1, 167-173.

Phan, A. T., Kuryavyi, V., Ma, J. B., Andreola, M. L. & Patel, D. J. (2005). An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. ***Proc. Natl. Acad. Scis. USA*** 102, 634-639.

Phan, A. T., Modi, Y. S. & Patel, D. J. (2004). Propeller-type parallel-stranded G-quadruplexes in the human *c-myc* promoter. ***J. Am. Chem. Soc.*** 126, 8710-8716.

Radhakrishan, I., and Patel, D. J. (1993). Solution structure of a purine•purine•pyrimidine DNA triplex containing G•GC and T•AT triples. ***Structure*** 1, 135-152.

Radhakrishnan, I., and Patel, D. J. (1994). Solution structure of a pyrimidine•purine•pyrimidine DNA triplex containing T•AT, C^+•GC and G•TA triples. ***Structure*** 2, 17-32.

Radhakrishnan, I., de los Santos, C. and Patel, D. J. (1991). NMR structural studies of intramolecular purine•purine•pyrimidine DNA triplexes in solution: triple pairing alignments and strand direction. ***J. Mol. Biol.*** 221, 1403-1418.

Rosen, M. A., and Patel, D. J. (1993). Structural features of a three-stranded DNA junction containing a C-C junctional bulge. ***Biochemistry*** 32, 6576-6587.

Rosen, M. A., Shapiro, L. and Patel, D. J. (1992). Solution structure of a trinucleotide A-T-A bulge loop within a DNA duplex. ***Biochemistry*** 31, 4015-4026.

Wang, Y. and Patel, D. J. (1992). Guanine resides in d(T_2AG_3) and d(T_2G_4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. ***Biochemistry*** 31, 8112-8119.

Wang, Y., & Patel, D. J. (1994). Solution structure of the d(T-C-G-A) duplex at acidic pH: A parallel-stranded helix containing C^+•C, G•G and A•A pairs. ***J. Mol. Biol.*** 242, 508-526.

Wang, Y., and Patel, D. J. (1993). Solution structure of the human telomeric repeat d($AG_3(T_2AG_3)_3$] G-tetraplex. ***Structure*** 1, 263-282.

Wang, Y., de los Santos, C., Gao, X., Greene, K., Live, D. and Patel, D. J. (1991). Multinuclear NMR studies of Na cation-stabilized complex formed by d(GGTTTTCGG) solution. ***J. Mol. Biol.*** 222, 819-832.

Zhang, N., Majumdar, A., Gorin, A., Kettani, A., Chernichenko, N., Skripkin, E., & Patel, D. J. (2001). Dimeric DNA quadruplex containing major groove aligned A•T•A•T and G•C•G•C tetrads stabilized by inter-subunit Watson-Crick A•T and G•C pairs. ***J. Mol. Biol.*** 312, 1073-1088.

Complementary Publications: Other Labs

Ambrus, A. et al. (2005). Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. ***Biochemistry*** 44, 2048-2058.

Ambrus, A. et al. (2006). Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel-antiparallel strands in potassium solution. ***Nucleic Acids Res***. 34, 2723-2735.

Rajagopal, P. and Feigon.J. (1989). Triple-strand formation in the homopurine:hompyrimidine deoxyribonucleotides d$(G-A)_4$ and d$(C-T)_4$. ***Nature*** 339, 637-640.

Robinson, H. et al. (1992). Unusual DNA conformation at low pH revealed by NMR: parallel-stranded DNA duplex with homo base pairs. ***Biochemistry*** 31, 10510-10517.

Smith, F. W. and feigon, J. (1992). Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. ***Nature*** 356, 164-168.

Wang, A. J. et al. (1981). Left-handed double helical DNA: variations in the backbone conformation. ***Science*** 211, 171-176.

Wei, D. et al. (2012). Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop formation. ***Nucleic Acids Res***. 40, 4691-4700.

Xu, Y. et al. (2006). The new models of the human telomere d(AGGG$(TTAGGG)_3$] in K^+ solution. ***Bioorg. Med. Chem.*** 14, 5584-5591.

11. DRUG-DNA COMPLEXES

(1974 to 1990)

We have developed NMR approaches to monitor the sequence-specific binding of groove-binders and intercalators to DNA. For each complex, we have determined the range of intermolecular contacts, thereby accounting for sequence-specificity of recognition and stabilization to complex formation. In addition, the impact of drugs bound at adjacent sites on DNA have been probed from both structural and imino proton hydrogen exchange measurements thereby providing insights into the potential propagation of conformational perturbations to segments flanking the binding site.

11.1 NON-COVALENT DRUG-DNA COMPLEXES

Intercalators and Bisintercalators. The antibiotic actinomycin D is composed of a phenoxazone ring to which are attached a pair of pentapeptide lactone rings that project on either side of the chromophore. We have monitored the proton and phosphorus NMR of the self-complementary d(ATGCAT) duplex on addition of actinomycin D. The selective upfield chemical shifts observed for the guanosine imino protons on complex formation are consistent with intercalation of the planar phenoxazone ring of actinomycin between the central G-C base pairs of the duplex and positioning of the pentapeptide lactone rings within the minor groove (Patel, *Biochemistry* 1974; see also Sobell and Jain, *J. Mol. Biol.* 1972). Such an intercalative mode of binding was confirmed by observing downfield shifts for a pair of phosphorus resonances on complex formation.

The antibiotic echinomycin is composed of two quinoxaline-2-carboxylic acid chromophores attached to a cross-bridged cyclic octapeptide dilactone containing both L- and D-amino acids. An NMR-based solution structure has been determined by our group for the bisintercalator echinomycin bound to self-complementary d(ACGT) and d(TCGA) duplexes. The large upfield guanosine imino proton and large downfield phosphorus chemical shifts on complex formation, together with intermolecular NOE distance constraints, are reflective of the two quinoxaline chromophores of echinomycin bisintrcalating from the minor groove surrounding the CG step (Gao and Patel, *Biochemistry* 1988; see also Wang et al. *Science* 1984). Further, this step adopts the C(C3'-*endo*)-G(C2'-*endo*) sugar pucker geometry characteristic of intercalation sites.

Nogalamycin belongs to the anthracycline antibiotic family with nogalose and bicyclic sugars as appendages attached at either end of the aglycone chromophore. The NMR-based solution structure has been determined by our group for nogalamycin bound to the self-complementary d(AGCATGCT) duplex. Chemical shift changes on complex formation and intermolecular NOE distance constraints establish that the aglycione chromophore intercalated between (CA)•(TG) steps with the long axis of the aglycone perpendicular to the long axis of the flanking base pairs, such that the nogalose and bicyclic sugars are positioned in the minor and major grooves, respectively (Zhang and Patel, *Biochemistry* 1990).

Monomeric and Dimeric Drugs Targeted to the DNA Minor Groove. The basic oligopeptide netropsin binds to the minor groove of A-T rich regions in DNA. Features of the NMR-based solution structure of the netropsin-d(CGCGAATTCGCG) complex have been identified by our group based on intermolecular NOE distance constraints between the pyrrole ring and side chain methylene protons of netropsin and the minor groove H2 proton of adenosines of the AATT segment in the complex (Patel, *Proc. Natl. Acad. Scis. USA*. 1982; see also Kopka et al. *J. Mol. Biol*. 1985).

The antitumor agent chromomycin A_3 is composed of a partially aromatic chromophore, to which are attached a hydrophilic side chain, a disaccharide segment (sugars A-B) and a trisaccharide segment (sugars C-D-E). The NMR-based solution structure has been determined by our group for chromomycin A_3 bound to the self-complementary d(TTGGCCAA) duplex. The observed set of intermolecular NOE distance constraints established that chromomycin A_3 forms an unprecedented dimer that targets a wider and shallower minor groove centered about the internal (GGCC)•(GGCC) segment (Gao and Patel, *Biochemistry* 1989). Further, each hydrophilic edge of the chromophore is located next to the GGCC half-site and each C-D-E trisaccharide chain extended towards the 3'-direction of the duplex. The A-B disaccharide and the hydrophilic side chain are directed towards the phosphate backbone.

11.2 COVALENT DRUG-DNA COMPLEXES

Enediyne Drugs Targeted to the DNA Minor Groove. Esperamycin A_1 and calicheamycin γ_1^I are enediyne antibiotics possessing antitumor activity associated with their ability to bind and, following activation, affect strand cleavage of DNA. The NMR-based solution structure has been determined by our group for esperamycin A_1 bound to the self-complementary d(CGGATCCG) duplex. Esperamycin targets the DNA minor groove with its anthranilate moiety intercalating into the helix at the (GG)•(CC) step. The anthranilate intercalator and the minor groove binding A-B-C trisacchride moieties rigidly anchor the enediyne such that the pro-radical centers are proximal to their anticipated proton abstraction sites (Kumar et al. *J. Mol. Biol.* 1997a). The thiomethyl B sugar is buried deep in an edge-wise manner in the minor groove with its two faces sandwiched between the walls of the groove. There is little perturbation of the right-handed DNA duplex other than unwinding of the helix at the intercalation site and widening of the minor groove at the binding site.

The NMR-based solution structure has been determined by our group for calicheamicin γ_1^I bound to the self-complementary DNA duplex containing a central d(AGGA)•d(TCCT) segment. Calicheamicin γ_1^I binds to the minor groove of the DNA and in the process positions the enediyne ring to abstract hydrogen atoms from partner strands leading to double-strand cleavage. The sequence specificity of intermolecular recognition has been identified with the aryltetrasaccharide component of the drug which binds in an extended conformation, with its thio sugar B and thiobenzoate ring C inserted in an edge-wise manner deep into the minor groove with their faces sandwiched between the walls of the groove (Kumar et al. *J. Mol. Biol*. 1997b). Sequence-specific binding was facilitated by the complementarity of the fit through hydrophobic and hydrogen bonding interactions between the enediyne and the floor and the walls of the minor groove of a minimally perturbed DNA helix.

11.3 IMPACT OF ADJACENT DRUG BINDING SITES ON DNA

Adjacently-bound Intercalators and Groove Binders. The NMR proton and phosphorus parameters measured by our group establish that the self-complementary d(CGCGAATTCGCG) duplex can simultaneously accommodate actinomycin intercalated at GC steps towards either end and netropsin positioned in the minor groove spanning the central AATT segment, with some mutual interaction between antibiotic binding sites (Patel et al. *Proc. Natl. Acad. Scis. USA.* 1981).

The exchange lifetimes of imino protons of self-complementary d(CGCGAATTCGCG) in H_2O solution have been monitored by NMR in the presence of bound netropsin and in the presence of bound actinomycin D (Pardi et al. *Biochemistry* 1983; with Ignatio Tinoco lab at UC-Berkeley). On formation of the netropsin-DNA complex, there is not only a large kinetic stabilization of the A-T base pairs within the netropsin-bound AATT segment but also a significant stabilization of the flanking G-C base pairs. For the actinomycin complex, where the phenoxazone ring intercalates between GC steps and the pentapeptide lactone rings are positioned in opposite directions within the minor groove, there is an increase in the lifetime of G-C base pairs at the binding site, accompanied by a slight kinetic destabilization for A-T base pairs within the central AATT segment.

Publications: Patel Lab

Gao, X. and Patel, D. J. (1989). Solution structure of the chromomycin-DNA complex. ***Biochemistry*** 28, 751-762.

Gao, X., and Patel, D. J. (1988). NMR studies of echinomycin bisintercalation complexes with d(A1-C2-G3-T4) and d(T1-C2-G3-A4) duplexes in aqueous solution: Sequence dependent formation of Hoogsteen A1•T4 and Watson Crick T1•A4 base pairs flanking bisintercalation site. ***Biochemistry*** 27, 1744-1751.

Kumar, R. A., Ikemoto, N. & Patel, D. J. (1997a). Solution structure of the esperamicin A_1-DNA complex. ***J. Mol. Biol.*** 265, 173-186.

Kumar, R. A., Ikemoto, N. & Patel, D. J. (1997b). Solution structure of the calicheamicin γ_1^I-DNA complex. ***J. Mol. Biol***. 265, 187-201.

Pardi, A., Morden, K. M., Patel, D. J., and Tinoco, I., Jr. (1983). The kinetics for exchange of the imino protons of the d(CGCGAATTCGCG) double helix in complexes with the antibiotics netropsin and/or actinomycin. ***Biochemistry*** 22, 1107-1113.

Patel, D. J. (1974). Peptide antibiotic-oligonucleotide interactions. NMR investigations of complex formation between actinomycin D and d(ATGCAT) in aqueous solution. ***Biochemistry*** 13, 2396-3402.

Patel, D. J. (1982). Antibiotic-DNA interactions: Intermolecular nuclear Overhauser effects in the netropsin-d(CGCGAATTCGCG) complex in solution. ***Proc. Natl. Acad. Scis. USA.*** 79, 6424-6428.

Patel, D. J., Kozlowski, S. A., Rice, J. A., Broka, C., and Itakura, K. (1981). Mutual interaction between adjacent dG•dC actinomycin binding sites and dA•dT netropsin binding sites on the self-complementary d(CGCGAATTCGCG) duplex in solution. ***Proc. Natl. Acad. Scis. USA.*** 78, 7281-7284.

Zhang, X. and Patel, D. J. (1990). Solution structure of the nogalamycin-DNA complex. ***Biochemistry*** 29, 9451-9466.

Complementary Publications: Other Labs

Kopka, M. L. et al. (1985). Binding of an antitumor drug to DNA: netropsin and C-G-C-G-A-A-T-T-brC-G-C-G. ***J. Mol. Biol***. 183, 553-563.

Sobell, H. M. and Jain, S. C. (1972). Stereochemistry of actinomycin binding to DNA II: Detailed molecular model of actinomycin-DNA complex and its implications. ***J. Mol. Biol***. 68, 21-34.

Wang, A. H. et al. (1984). The molecular structure of a DNA-triostin complex. ***Science*** 14, 1115-1121.

12. EARLY INDEPENDENT RESEARCH

(1969 - 1976)

We capitalized on access to high field NMR instrumentation in the late 1960s to address challenging problems in the small molecule, peptide, protein and RNA fields during the beginning stage of our scientific career. Our first application of NMR was to retinal, the prosthetic group of rhodopsin, with the emphasis on monitoring the factors that contribute to the extent of charge and spin delocalization along the polyene chain for various isomeric (*cis/trans*) retinals.

These studies were next extended to cation-binding cyclic depsipeptides and peptides to monitor conformational interconversions in the free state and conformational transitions leading to formation of the compact singular complex with bound cations. Such studies provided insights into contributions of hydrophobic and hydrogen bonding interactions to conformational stability and cation-acceptor interactions and *cis-trans* proline peptide bond isomerizations to complex formation.

Most importantly, we pioneered the recording of NMR spectra of exchangeable NH protons in proteins and nucleic acids in H_2O solution with the emphasis on slowly exchanging amide protons (8 to 10 ppm) in proteins and imino protons in nucleic acids (11 to 12.5 ppm). Thus, the slowly exchanging amide proton markers could be monitored as a function of oxidation state of cytochrome c, thereby providing insights into the redox-based chemistry of this important heme protein. Our introduction of NMR studies of imino protons in tRNA has revolutionized applications of NMR to the nucleic acid field, as it opened up opportunities for the application of exchangeable proton NMR approaches to study the solution structure and dynamics of DNA, RNA and their complexes with small molecules, as well as protein-nucleic acid complexes.

Contributions

12.1 PROTON NMR STUDIES OF SMALL MOLECULES

The access to high-resolution NMR instruments (220 MHz) at the end of the 1960's provided opportunities for us to monitor small molecules, such as retinals, that existed in several isomeric (*cis/trans*) states along the polyene chain. Given that proton resonances could be assigned unambiguously, NMR approaches were shown to have the potential to monitor and address both charge and spin delocalization down the polyene chain.

Retinals. High-resolution proton NMR spectra were recorded by our group for all-*trans*-retinal, and its 13-*cis*, 11-*cis*, 9-*cis*, 9,13-*cis* and 11,13-*cis* isomers. An analysis of chemical shifts and coupling constants established that the steric strain between H10 and CH_3-13 in 11-*cis* retinal could be relieved by a skew geometry about the C12-C13 single bond (Patel, *Nature* 1969a).

Charge delocalization for various retinal isomers were studied by our group by following proton and methyl NMR chemical shifts upon addition of trifluoroacetic acid (Patel, *Nature* 1969b). These studies were extended to spin delocalization in various retinal isomers using proton and methyl NMR chemical shift changes following binding of Ni(II) and Co(II) acetylacetonates to retinyl-N-methyl Schiff's bases. These experiments have provided the magnitude and sign of the spin densities at the carbon atoms in the π system of the all-*trans*, 9-*cis* and 13-*cis* retinal isomers (Patel and Shulman, *Proc. Natl. Acad. Scis. USA*. 1970). Comparison of experimental with molecular orbital calculations have provided independent insights into spin delocalization down the polyene chain.

Cofactors. The methylene protons at the 4 position of the nicotinamide ring of NADH have been followed by high resolution NMR as a function of temperature in our group. These methylene protons are chemically and magnetically non-equivalent in the folded form of NADH in the absence of substrate and enzyme, thereby accounting for the specific recognition by certain enzymes of one or the other of the methylene protons (Patel, *Nature* 1969c).

12.2 PROTON NMR STUDIES OF CYCLIC PEPTIDES

The inherent flexibility of linear peptides have limited NMR-based structural studies of their conformations in solution. By contrast, cyclization limits conformational mobility in cyclic peptides,

thereby allowing structural studies to be undertaken on cation-binding cyclic depsipeptides such as valinomycin (K^+-binder) and cyclic proline-containing peptides such as antamanide (Na^+-binder). Of particular interest are the extent of conformational plasticity of these cyclic peptides in the free state and the conformational transitions associated with formation of a compact topology in the cation-bound state

Cyclic Depsipeptides. Valinomycin is a cyclic 12-membered depsipeptide composed of alternating amino acid and ester residues with both L- and D-configurations. Proton NMR chemical shifts and coupling constants, as well as conformational calculations, have been applied by our group to monitor changes in the valinomycin conformation as a function of solvent polarity (Patel and Tonneli, *Biochemistry* 1973; see also Bystrov et al. *Eur. J. Biochem.* 1977). These NMR studies have been extended to define the solution structure of the valinomycn-K^+ complex.

Cyclic Proline-containing Peptides. Antamanide, a 10-mer cyclic peptide containing a pair of symmetry-related Pro-Pro steps, is known to be an ionophore that selectively binds Na^+ cation. Proton NMR studies of antamanide-Na^+ complex by our group are consistent with formation of two *cis*-Pro and two *trans*-Pro bonds, thereby restricting the conformation of the cyclic peptide in the Na^+ bound state (Patel, *Biochemistry* 1973; see also Karle, *Biochemistry* 1974).

12.3 PROTON NMR STUDIES OF PROTEINS AND tRNA IN H_2O SOLUTION

The non-exchangeable protons of proteins and nucleic acids span a crowded region of the NMR spectrum between 0 and 10 ppm in D_2O solution. To monitor exchangeable NH protons in proteins and nucleic acids, NMR spectra have to be recorded in H_2O solution, and even then the protons have to exchange slowly with bulk solvent, to detect narrow resonances. Indeed, slowly exchanging amide protons in proteins and amino protons of nucleic acids can be detected and assigned in the 8 to 10 ppm range. Our efforts have focused on the 10 to 15 ppm NMR spectral window where we have detected slowly exchanging NH protons for both proteins and nucleic acids, with nucleic acid imino protons of uridine/thymidine and guanosine participating in hydrogen bonds providing critically important markers between 11 and 12.5 ppm. This breakthrough has been especially important for studying the structure and dynamics of tRNA, riboswitches and ribozymes in solution.

Proteins. Exchangeable protons have been observed for diamagnetic oxygen-bound myoglobin in H_2O solution in a spectral window between 10 and 15 ppm. These resonances have been assigned by our group based on the chemical shift dependence as a function of species, as well as spin, pH and temperature and provide useful markers for following myoglobin function (Patel et al. *Proc. Natl. Acad. Scis. USA*. 1970).

Exchangeable proton NMR spectra also have been used to monitor slowly exchanging backbone amide protons of cytochrome c in the reduced and oxidized states between 8 and 10 ppm, with the 12 resolvable amide protons originating from the same residues in both oxidation states (Patel and Canuel, *Proc. Natl. Acad. Scis. USA.* 1976). These slowly-exchangeable amide protons were studied by our group as a function of pH and temperature to demonstrate sequential exchange, thereby ruling out a resistant cluster of protons that exchange together.

tRNAs. A number of exchangeable protons have been observed by our group in the high -esolution proton NMR spectrum of yeast $tRNA^{Phe}$ between 11 and 15 ppm in H_2O solution. These protons arise from imino protons of guanosine (N^1H of G) and thymidine (N^3H of T) involved in slow exchange due to base pairing and/or shielding from solvent (Kearns et al. *Nature* 1971). Hence, once these imino protons, which are reflective of G-C and A-U pair formation, can be assigned to specific base pairs in tRNA, they promise to provide critical markers of base pairing that are distributed throughout the tRNA clover-leaf fold. Indeed, the imino proton spectral resolution improved significantly with the advent of access to even higher field NMR instrumentation.

Publications: Patel Lab

Kearns, D. R., Patel, D. J., and Shulman, R. G. (1971). High resolution NMR studies of hydrogen-bonded protons in tRNA in H_2O. ***Nature*** 229, 338-339.

Patel, D. J. (1969a). 220 MHz proton NMR spectra of retinals. ***Nature*** 221, 825-828.

Patel, D. J. (1969c). 220 MHz proton NMR spectrum of NADH. ***Nature*** 221, 1239-1241.

Patel, D. J. (1969b). Charge delocalization in the prosthetic group of visual pigments. ***Nature*** 224, 799-800.

Patel, D. J. (1973). Conformation of antamanide-sodium in solution. ***Biochemistry*** 12, 677-678.

Patel, D. J. and Shulman, R. G. (1970). Spin delocalization in retinals. ***Proc. Natl. Acad. Scis. USA***. 65, 31-38.

Patel, D. J. and Tonelli, A. E. (1973). Solvent-dependent conformations of valinomycin in solution. ***Biochemistry*** 12, 486-495.

Patel, D. J., Kampa, L., Shulman, R. G., Yamane, T., and Wyluda, B. J. (1970). Proton NMR studies of myoglobin in H_2O. ***Proc. Natl. Acad. Scis. USA.*** 67, 1109-1115.

Patel, D. J. and Canuel, L. L. (1976). NMR studies of slowly exchanging peptide protons in cytochrome c in aqueous solution. ***Proc. Natl. Acad. Scis. USA.*** 73, 1398-1402.

Complementary Publications: Other Labs

Bystrov, V. F. et al. (1977). Refinement of the solution conformation of valinomycin with the acid coupling constants from the ^{13}C-NMR spectra. ***Eur. J. Biochem***. 78, 63-82.

Karle, I. (1974). The conformation of the sodium complex of a biologically active analog of antamanide in the crystalline state. ***Biochemistry*** 13, 2155-2162.