The Dissertation Committee for Dominic Nathaniel McBrayer Certifies that this is the approved version of the following dissertation:

Synthesis and Evaluation of Rationally Designed Small Molecular Libraries for G-quadruplex Selective DNA Photocleavage

Committee:

Sean Kerwin, Supervisor

Walter Fast

Brent Iverson

Seongmin Lee

Karen Vasquez

Synthesis and Evaluation of Rationally Designed Small Molecular Libraries for G-quadruplex Selective DNA Photocleavage

by

Dominic Nathaniel McBrayer, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2015

Dedication

This work is dedicated to my family. They have always steadfastly supported me in my endeavors, and this was no exception.

Acknowledgements

I would like to thank my supervisor and each of my committee members for their feedback which has been essential in guiding my research. I thank Karen Vasquez and Walter Fast for allowing me access to their laboratory resources, including instruments and reagents. I would also like to thank Michelle Schoonover (formerly from the Kerwin lab) for a range of insightful discussions and early training, Guliang "Graham" Wang from the Vasquez lab for the pSP189 and pSP189G4 plasmids and discussions involving them, Imee Del Mundo from the Vasquez lab for her help in the DMS assays and for general discussions, Gulnar Rawji for discussions about synthesis and photocleavage analysis, and Alesha Stewart from the Fast lab for kindly sharing some of her competent bacterial stocks. NMR and mass spectra were obtained with the help of the UT chemistry department NMR and MS staff respectively.

Synthesis and Evaluation of Rationally Designed Small Molecular Libraries for G-quadruplex Selective DNA Photocleavage

Dominic Nathaniel McBrayer, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Sean Kerwin

DNA can fold into highly stable structural conformations known as G-quadruplexes (G4). Research has suggested that G4 DNA may play a role in the regulation of gene expression and it has been linked with several diseases including Werner's Syndrome, Bloom's Syndrome, and cancer. In order to better understand the biological roles of G4 DNA, small molecules that can be used to selectively detect these structures in a time-dependent manner were designed. These G4 DNA probe candidates bind to and photocleave G4 DNA, allowing analysis of the cleavage products as well as temporal detection of G4 DNA. The synthesis, characterization, and structure activity relationships (SAR) of two rationally designed libraries are discussed. The photocleavage activity and binding selectivity for these compounds is evaluated in both the context of single stranded DNA and plasmids. The development of a program that can be used to search genome sequences for G-triplex (G3) DNA, a potential folding intermediate to G4 DNA that may have its own biological roles, is also discussed.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Deoxyribonucleic Acid (DNA) is a polymer that contains the coding information necessary for all known life. The sequence of DNA's four bases, guanine, adenine, cytosine, and thymine, determines the sequences of transcribed RNA, and indirectly, the amino acid sequences of translated proteins. This has come to be known as the central dogma of biology (1). Studies have revealed a staggering amount of complexity in the many possible products coded for by DNA. The complexity extends further since among these products are those which can recognize and reinterpret the information stored in DNA, for instance as is seen in alternate splicing (2). There is also a great deal of complexity available in the three-dimensional structures that DNA can adopt that extend well beyond the most widely recognized helical structure of B-form DNA. DNA is capable of adopting several other variations on the helix as well as even more complicated secondary structures including hairpins, Holliday junctions, triplexes, and G-quadruplexes (3). These secondary structures can play important roles in the regulation of gene expression, with potential effects at all levels of the central dogma of biology.

G-QUADRUPLEX DNA STRUCTURE

G-quadruplexes (G4s) are a form of secondary conformation that DNA (and RNA) can adopt. G4 structures are formed by the monovalent-ion-stabilized stacking of G-tetrads (also known as G-quartets), which are in turn formed through Hoogsten-bonding of four guanines (4). The monovalent ions most commonly seen in the biological setting are sodium or potassium, although a wide range of monovalent ions have been demonstrated to stabilize G4 structures (5, 6). These ions reside in a channel down the center of the quadruplex (Figure 1.1 B and D). Quadruplexes can be formed both inter and intra- molecularly from a wide variety of different DNA sequences which

follow a similar general sequence pattern. Each of these sequences may form polymorphic G4 structures (6). This polymorphism is exhibited in the many different potential conformations that can form from a given sequence, which are dependent on the arrangements of the non-Hoogsten-bonded bases which loop out along the sides of the stacked tetrads (Figure 1.1 C and D), the concentrations of the stabilizing monovalent ions, and the sequences of the contributing DNA strand(s).

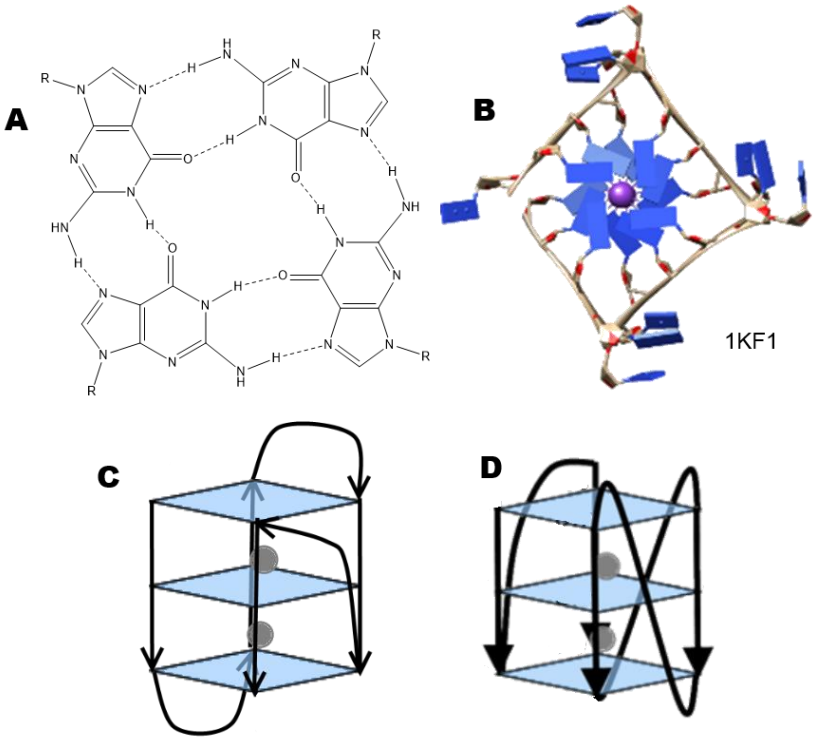


Figure 1.1: G-quadruplex structural features. (A) G-tetrad, dashed lines indicate the locations of Hoogsten hydrogen bonding. R is where the ribose ring of the DNA backbone attaches to the guanine base. (B) Crystal structure image of a G-quadruplex looking down the center ion channel (7). The monovalent ions are depicted as purple spheres, the bases are shown as blue rectangles, and the DNA backbone is in gray. (C) Diagram of a mixed-type G-quadruplex containing both antiparallel and parallel loop arrangements. Arrows indicate the directionality of the DNA sequence from 5' to 3'. The shaded rectangles represent G-tetrads. (D) Diagram of a parallel G-quadruplex. Gray circles represent the monovalent ions in the center ion channel.

Conformations are most commonly referred to in terms of the directionality of the DNA sequence following the 5' to 3' convention, in part due to the use of circular

dichroism (8–10) to identify the formation and loop orientation of G4 structures. When the directionality of the DNA strands at the edges of the quadruplex are in the same direction, the quadruplex is considered to be parallel stranded. If the directionality alternates, it is instead considered to be anti-parallel. More complicated structures can be formed by the mixing of patterns of parallel and anti-parallel arrangements (11, 12).

When discussing G4 structure sequence, the term G-tract has been used to refer to the segments of consecutive G's. For example, GGGTTAGGG could be said to have two (non-overlapping) or four (overlapping) G-tracts of size 2nt and/or two G-tracts of size 3nt. Bioinformatics predictions of G4 DNA have typically used G-tracts of size 3 (13) or larger (although there have been some G4 structures formed *in vitro* with G-tracts of size 2) with a variable number of nucleotides composing the loops, typically between 1 and 15nt. However, the loop length has been extended up to 40nt in some cases given evidence that G4 structures can still form even with loops that are quite large (14). Thus the sequence pattern that is predictive of G4 DNA structures is: $G_xN_yG_xN_yG_xN_yG_x$, where x is greater than or equal to two and y is greater than or equal to one. The value of each x and each y can be considered to be independent of each other. However, the number of tetrads formed is typically limited by the smallest of the G-tracts incorporated into the structure with the remaining Gs from the larger tracts occupying the loops, although bulges within G-tracts have recently shown to be possible (15).

Several other structures are associated with G-quadruplex formation. I-motifs have been proposed to form in the C-rich complement strand of the DNA involved in forming G4 structures (16). Although I-motifs are most stable under more acidic conditions, molecular crowding experiments, and the effects of supercoiling have suggested that I-motifs are not necessarily disfavored under physiological conditions (17, 18). G-triplexes (not to be confused with regular triplexes) are structures similar to G-

quadruplexes, however, they are formed by the association of the guanines in only three G-tracts as compared to the four G-tracts needed for G-quadruplexes. G-triplexes have been proposed as possible folding intermediates to G-quadruplexes (19).

SIGNIFICANCE OF G-QUADRUPLEXES

While G-quadruplexes have been known to form *in vitro* for many years (20), their association with several diseases, such as cancer, Werner's Syndrome, and Bloom's Syndrome (21) have increased interest in understanding G4 DNA and its influence on biological processes.

Analysis of the sequences of the human genome has given support for the importance of G4-forming structures in humans. It was determined that sequences capable of adopting these structures occur much less often than would be expected by chance in the case of mRNA (13), suggesting the presence of selective pressure. Studies of the sequences have also shown that a preponderance of sequences predicted to form G4 structures are found in or near the promoter regions of proto-oncogenes (22, 23), suggesting a role in gene regulation and in cancer.

There have also been several helicases that favor G4 structures as a substrate (Table 1.1 shows helicases that have been shown to act on intermolecular G4 structures). Mutations in many of these helicases are linked to human hereditary diseases (21), many of which include a greatly heightened risk for cancer among their symptoms.

Despite this strong indirect evidence, it was some time before the formation of G4 structures in a living system could be verified. This was achieved using fluorescently labeled antibodies that had been raised to recognize G4 structures. G4 DNA was first located in ciliates and bacteria (24, 25), and recently, staining with antibodies also allowed for the visualization of G4 DNA and RNA within cultured human cells (26, 27).

Table 1.1: Intermolecular G-quadruplex Helicases

Helicase Family	Helicase Name	Organism	Direction	Intermolecular (Substrate)	Reference
Dna2 Nuclease/Helicase (SF1)	Dna2	*S. cerevisiae*	5'-->3'	G'2>G4	(28)
	hDna2	*Human*	5'-->3'	G'2, G4	(28)
Pif1 (SF1)	Pif1	*Human*	5'-->3'	G4 (DNA)	(29, 30)
	Pif1	*S. cerevisiae*	5'-->3'	G4 (DNA)	(30, 31)
	Pif1	*Bacteroides sp.*	5'-->3'	G4 (DNA)	(30)
UvrD (SF1)	Srs2	*S. cerevisiae*	3'-->5'	G4 (DNA)	(30)
DEAH/RHA (SF2)	G4 Resolvase-1 (G4R1)/RHAU/ DHX36/MLEL1	*Human*	3'-->5'	G'2 (DNA & RNA), G4 (RNA) > G4 (DNA)	(32–35)
	DHX9/NDHII/RHA	*Human*	3'-->5'	G'2 (DNA) < G'2 (RNA)	(32)
	Gp41	*phage T4*	5'-->3'	G4 (DNA)	(30)
Rad3/XPD Family DEXH (SF2)	FANCJ	*Human*	5'-->3'	G'2, G4 (DNA)	(36, 37)
	DDX11/CHLR1/WABS1	*Human*	5'-->3'	G'2>G4 (DNA)	(38)
RecQ (SF2)	BLM	*Human*	3'-->5'	G4, G'2 (DNA)	(39–43)
	WRN	*Human*	3'-->5'	G'2>G4 (DNA), Not G'2 (RNA)	(32, 39, 40, 44)
	Sgs1	*S. cerevisiae*	3'-->5'	G4, G'2 (DNA)	(41, 45–47)
	RecQ	*E. coli*	3'-->5'	G4, G'2 (DNA)	(42, 48)
	RecQ1	*Human*	3'-->5'	No	(37, 43)
	StyRecQL	*Stylonychia*	ND	Assumed Telomeric	(49)
Large T Antigen (SF3)	Large T-antigen	*SV40*	3'-->5'	G'2>G4 (DNA)	(50)
DnaB-like (SF4)	DnaB	*E. coli*	5'-->3'	G4 (DNA)	(30)

Understanding of the specific biological functions and mechanisms of G4 DNA is still lacking. G4 structures could present physical barriers to polymerases, which could allow G4 structure to prevent gene expression. This possibility has driven the development of many G4 ligands, which, by stabilizing the G4 structure, are expected to

decrease polymerase activity. In particular, this has been done in hopes of inhibiting the action of telomerase in cancer cells (51), giving rise to potential anti-cancer therapeutics. However, some experiments have suggested that the presence of a G4 structure is required for full gene expression (52). This implies that the roles of G4 structures may be more complex and are likely dependent on the context in which they are found.

G-QUADRUPLEX LIGANDS

One of the early G4 ligands reported was telomerastatin (Figure 1.2), so named for its ability to inhibit telomerase activity (53), and which differs from many G4 ligands in its large, macrocyclic structure. Porphyrins, such as TMPyP4 (Figure 1.2), represent another set of compounds that have been studied extensively as G4 ligands (47, 54–58). Additional examples of other G4 ligands include 360A (Figure 1.2), which was shown to be selective for G4s formed from human telomeric sequence over duplex DNA (53, 59), and perylene diimides (50, 60–62).

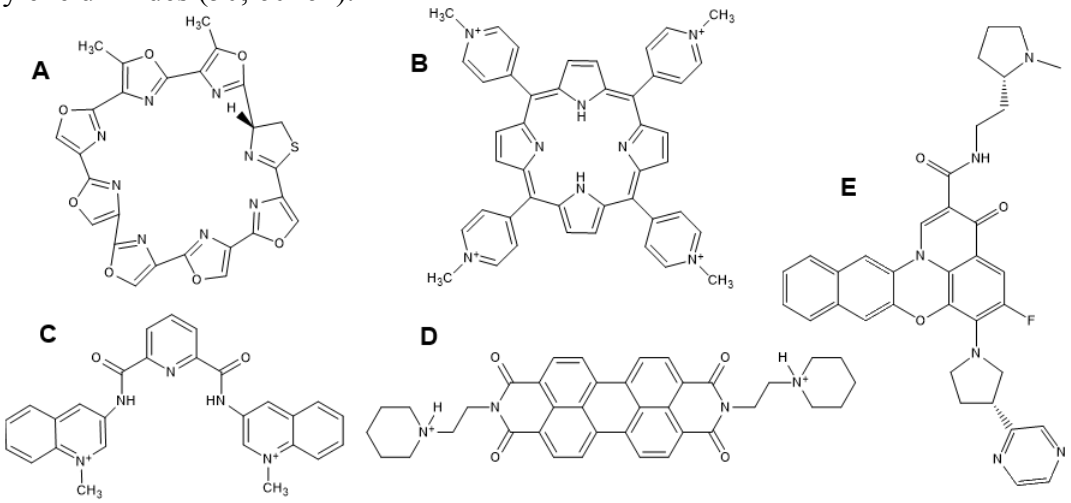


Figure 1.2: Example G4 Ligands. (A) Telomeristatin, (B) TMPyP4, (C) 360A, (D) PIPER, a perylenediimide, (E) Quarfloxin, which has entered phase II clinical trials

These compounds share several key features which grant them varying degrees of binding selectivity for G4 structures. They have planar aromatic regions capable of pi-stacking with the guanines composing the tetrads at either end of the G4 DNA. Also, most often incorporate positively charged peripheral groups that are capable of interacting with the negatively charged phosphate backbone of the DNA.

While most G4 ligands likely stack on the tetrads at the ends of the quadruplex, some researchers have proposed binding through intercalation (47), or association with the loops as additional potential mechanisms. For instance, Distamycin A differs from what has come to be the more "canonical" design with its long chain of aromatically-linked rings. This chain binds to the G4 structure through association with the grooves on the sides of the structure created by the loops of the quadruplex (63).

The use of G4 ligands in therapeutics is still in its infancy. One major reason for this is that many of the molecular features that favor the formation of the G4 DNA-ligand complex also tend to make it more difficult to use the ligand as a therapeutic. The planar aromatic character of these compounds can make them prone to aggregation, limiting their solubility. The polymorphism of G-quadruplex targets also makes it difficult to avoid interactions with unintended regions in other genes. However, at least one compound, Quarfloxin (CX-3543), has reached the phase II clinical trials for use as an anti-cancer drug (Figure 1.2). Quarfloxin has been shown to interact selectively with G4 DNA over duplex DNA and has also shown preference for certain G4s in the genome (64).

RATIONALE FOR DEVELOPMENT OF G-QUADRUPLEX PHOTOCLEAVAGE AGENTS

Detection using small molecular probes has several advantages over other methods for detecting G-quadruplexes. While antibody-based detection is highly

specific, the large proteins needed to detect G4 structures can make delivery of the probe difficult. Currently, these probes have been limited to cell cultures that are fixed and stained for observation. Small molecules are more easily delivered to living cells and could potentially be used to detect G4 structures in still-living cells, and, perhaps, could even detect them *in vivo*. Small molecules can also be cheaper to produce in large quantities; using them could be more cost effective than protein-based approaches (65).

Small molecules that also have photoactive effects (such as photocleavage and/or inherent fluorescence) have the additional advantage of an inducible signal that can be used for the temporal detection of G-quadruplexes. Temporal detection is crucial if these probes are ever to be applied to living systems and would enable the determination of the longevity of G4 DNA structures in various biological contexts. Fluorescence-based probes have a significant advantage because their signal is unlikely to add additional toxicity to the compound. In comparison, photocleavage compounds come with an increased risk of heightened toxicity due to damage dealt to the DNA. However, this damage has the potential to give more information if the cleaved DNA is analyzed. DNA cleavage patterns have been used in the past to aid in interpretation of interactions between G4 DNA and a ligand (47, 62). The destructiveness of photocleavage compounds also has potential therapeutic applications, provided the compound can be made specific enough to preferentially target G4 structures associated with the disease state.

WORK CONTAINED HEREIN

This dissertation describes the synthesis of several libraries of G4-selective small molecule photocleavage agents in the effort to develop probes for the use of the study of the biological roles of G4 DNA. For most of the compounds discussed, a strategy of

combining a known G4 DNA moiety with known photocleavage moieties through "click" chemistry (66) was utilized. The photocleavage activity of the library compounds were assessed both through the use of PAGE and by measuring the reduction in FRET quenching in dual-labeled oligonucleotides used in the studies. The extent of the selectivity of photocleavage was investigated using plasmid nicking assays. The binding specificity of representatives of the libraries, as indicated by changes in T_m, was also measured.

CHAPTER 1 REFERENCES

1. Crick,F. (1970) Central Dogma of Molecular Biology. *Nature*, **227**, 561–563.

2. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.

3. Saini,N., Zhang,Y., Usdin,K. and Lobachev,K.S. (2013) When secondary comes first – The importance of non-canonical DNA structures. *Biochimie*, **95**, 117–123.

4. Burge,S., Parkinson,G., Hazel,P., Todd,A. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.

5. Federiconi,F., Ausili,P., Fragneto,G., Ferrero,C. and Mariani,P. (2005) Locating counterions in guanosine quadruplexes: A contrast-variation neutron diffraction experiment in condensed hexagonal phase. *J. Phys. Chem. B*, **109**, 11037–11045.

6. Lipps,H. and Rhodes,D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.

7. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.

8. Paramasivan,S., Rujan,I. and Bolton,P.H. (2007) Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods*, **43**, 324–331.

9. Vorlíčková,M., Kejnovská,I., Sagi,J., Renčiuk,D., Bednářová,K., Motlová,J. and Kypr,J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64–75.

10. Randazzo,A., Spada,G.P. and da Silva,M.W. (2013) Circular Dichroism of Quadruplex Structures. *Quadruplex Nucleic Acids*, **330**, 67–86.

11. Lane,A., Chaires,J., Gray,R. and Trent,J. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.

12. Qin,Y. and Hurley,L. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **90**, 1149–1171.

13. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.

14. Mukundan,V.T. and Phan,A.T. (2013) Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.

15. Martadinata,H. and Phan,A.T. (2014) Formation of a Stacked Dimeric G-Quadruplex Containing Bulges by the 5′-Terminal Region of Human Telomerase RNA (hTERC). *Biochemistry*, **53**, 1595–1600.

16. Brooks,T.A., Kendrick,S. and Hurley,L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.*, **277**, 3459–3469.

17. Sun,D. and Hurley,L.H. (2009) The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J Med Chem*, **52**, 2863–2874.

18. Benabou,S., Ferreira,R., Aviñó,A., González,C., Lyonnais,S., Solà,M., Eritja,R., Jaumot,J. and Gargallo,R. (2014) Solution equilibria of cytosine- and guanine-rich sequences near the promoter region of the n-myc gene that contain stable hairpins within lateral loops. *Biochim. Biophys. Acta BBA - Gen. Subj.*, **1840**, 41–52.

19. Bardin,C. and Leroy,J.L. (2008) The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.*, **36**, 477–488.

20. Gellert,M., Lipsett,M. and Davies,D. (1962) Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.*, **48**, 2013–&.

21. Wu,Y. and Brosh Jr.,R.M. (2010) G-quadruplex nucleic acids and human disease. *FEBS J.*, **277**, 3470–3488.

22. Hurley,L.H. (2002) DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer*, **2**, 188–200.

23. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 2105.

24. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 8572–8577.

25. Cahoon,L.A. and Seifert,H.S. (2009) An Alternative DNA Structure Is Necessary for Pilin Antigenic Variation in Neisseria gonorrhoeae. *Science*, **325**, 764–767.

26. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.

27. Biffi,G., Di Antonio,M., Tannahill,D. and Balasubramanian,S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.*, **6**, 75–80.

28. Masuda-Sasa,T., Polaczek,P., Peng,X.P., Chen,L. and Campbell,J.L. (2008) Processing of G4 DNA by Dna2 helicase/nuclease and replication protein a (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. *J. Biol. Chem.*, **283**, 24359–24373.

29. Sanders,C.M. (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem. J.*, **430**, 119–128.

30. Paeschke,K., Bochman,M.L., Garcia,P.D., Cejka,P., Friedman,K.L., Kowalczykowski,S.C. and Zakian,V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.

31. Ribeyre,C., Lopes,J., Boule,J.-B., Piazza,A., Guedin,A., Zakian,V.A., Mergny,J.-L. and Nicolas,A. (2009) The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. *Plos Genet.*, **5**, e1000475.

32. Chakraborty,P. and Grosse,F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair*, **10**, 654–665.

33. Creacy,S.D., Routh,E.D., Iwamoto,F., Nagamine,Y., Akman,S.A. and Vaughn,J.P. (2008) G4 Resolvase 1 Binds Both DNA and RNA Tetramolecular Quadruplex

with High Affinity and Is the Major Source of Tetramolecular Quadruplex G4-DNA and G4-RNA Resolving Activity in HeLa Cell Lysates. *J. Biol. Chem.*, **283**, 34626 –34634.

34. Vaughn,J.P., Creacy,S.D., Routh,E.D., Joyner-Butt,C., Jenkins,G.S., Pauli,S., Nagamine,Y. and Akman,S.A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **280**, 38117–38120.

35. Lattmann,S., Giri,B., Vaughn,J.P., Akman,S.A. and Nagamine,Y. (2010) Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.*, **38**, 6219 –6233.

36. London,T.B.C., Barber,L.J., Mosedale,G., Kelly,G.P., Balasubramanian,S., Hickson,I.D., Boulton,S.J. and Hiom,K. (2008) FANCJ Is a Structure-specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts. *J. Biol. Chem.*, **283**, 36132–36139.

37. Wu,Y., Rawtani,N., Thazhathveetil,A.K., Kenny,M.K., Seidman,M.M. and Brosh,R.M. (2008) Human Replication Protein A Melts a DNA Triple Helix Structure in a Potent and Specific Manner. *Biochemistry*, **47**, 5068–5077.

38. Wu,Y., Sommers,J.A., Khan,I., de Winter,J.P. and Brosh,R.M. (2012) Biochemical Characterization of Warsaw Breakage Syndrome Helicase. *J. Biol. Chem.*, **287**, 1007–1021.

39. Mohaghegh,P., Karow,J.K., Brosh Jr,R.M., Bohr,V.A. and Hickson,I.D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.*, **29**, 2843 –2849.

40. Li,J.L., Harrison,R.J., Reszka,A.P., Brosh,R.M., Bohr,V.A., Neidle,S. and Hickson,I.D. (2001) Inhibition of the Bloom's and Werner's syndrome helicases by G-quadruplex interacting ligands. *Biochemistry*, **40**, 15194–15202.

41. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.

42. Huber,M.D., Duquette,M.L., Shiels,J.C. and Maizels,N. (2006) A conserved G4 DNA binding domain in RecQ family helicases. *J. Mol. Biol.*, **358**, 1071–1080.

43. Popuri,V., Bachrati,C.Z., Muzzolini,L., Mosedale,G., Costantini,S., Giacomini,E., Hickson,I.D. and Vindigni,A. (2008) The Human RecQ Helicases, BLM and

RECQ1, Display Distinct DNA Substrate Specificities. *J. Biol. Chem.*, **283**, 17766 –17776.

44. Fry,M. and Loeb,L.A. (1999) Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)(n). *J. Biol. Chem.*, **274**, 12797–12802.

45. Han,H., Bennett,R.J. and Hurley,L.H. (1999) PIPER, a G4 DNA-interactive compound, prevents the unwinding of G-quadruplex DNA by yeast Sgs1 helicase. *Clin. Cancer Res.*, **5**, 3852S–3852S.

46. Han,H.Y., Bennett,R.J. and Hurley,L.H. (2000) Inhibition of unwinding of G-quadruplex structures by Sgs1 helicase in the presence of N,N ’-bis[2-(1-piperidino)ethyl]-3,4,9,10-perylenetetracarboxylic diimide, a G-quadruplex-interactive ligand. *Biochemistry*, **39**, 9311–9316.

47. Han,H., Langley,D.R., Rangan,A. and Hurley,L.H. (2001) Selective Interactions of Cationic Porphyrins with G-Quadruplex Structures. *J. Am. Chem. Soc.*, **123**, 8902–8913.

48. Wu,X. and Maizels,N. (2001) Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Res.*, **29**, 1765–1771.

49. Postberg,J., Tsytlonok,M., Sparvoli,D., Rhodes,D. and Lipps,H.J. (2012) A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication. *Gene*, **497**, 147–154.

50. Tuesuwan,B., Kern,J.T., Thomas,P.W., Rodriguez,M., Li,J., David,W.M. and Kerwin,S.M. (2008) Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents. *Biochemistry*, **47**, 1896–1909.

51. Chen,J., Sperry,J., Ip,N. and Brimble,M. (2011) Natural products targeting telomere maintenance. *MedChemComm*, **2**, 229–245.

52. Wei,P.-C., Wang,Z.-F., Lo,W.-T., Su,M.-I., Shew,J.-Y., Chang,T.-C. and Lee,W.-H. (2012) A cis-element with mixed G-quadruplex structure of NPGPx promoter is essential for nucleolin-mediated transactivation on non-targeting siRNA stress. *Nucleic Acids Res.* **2012.**

53. Cian,A.D. and Mergny,J.-L. (2007) Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. *Nucleic Acids Res.*, **35**, 2483–2493.

54. Han,F.X., Wheelhouse,R.T. and Hurley,L.H. (1999) Interactions of TMPyP4 and TMPyP2 with Quadruplex DNA. Structural Basis for the Differential Effects on Telomerase Inhibition. *J. Am. Chem. Soc.*, **121**, 3561–3570.

55. Vialas,C., Pratviel,G., Claparols,C. and Meunier,B. (1998) Efficient Oxidation of 2‘-Deoxyguanosine by Mn-TMPyP/KHSO5 to Imidazolone dIz without Formation of 8-Oxo-dG. *J. Am. Chem. Soc.*, **120**, 11548–11553.

56. Kovaleva,O.A., Shchyolkina,A.K., Mamaeva,O.K., Ol’shevskaya,V.A., Makarenkov,A.V., Semeikin,A.S., Shtil,A.A., Borisova,O.F. and Kaluzhny,D.N. (2013) Complexes of antiparallel telomeric G-quadruplex d(TTAGGG)4 with carboxymethyl tetracationic porphyrins. *Mol. Biol.*, **47**, 453–460.

57. Nagesh,N., Sharma,V.K., Ganesh Kumar,A. and Lewis,E.A. (2009) Effect of Ionic Strength on Porphyrin Drugs Interaction with Quadruplex DNA Formed by the Promoter Region of C-myc and Bcl2 Oncogenes. *J. Nucleic Acids*, **2010**.

58. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.

59. Granotier,C., Pennarun,G., Riou,L., Hoffschir,F., Gauthier,L.R., De Cian,A., Gomez,D., Mandine,E., Riou,J.-F., Mergny,J.-L., *et al.* (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.*, **33**, 4182–4190.

60. Kern,J.T., Thomas,P.W. and Kerwin,S.M. (2002) The Relationship between Ligand Aggregation and G-Quadruplex DNA Selectivity in a Series of 3,4,9,10-Perylenetetracarboxylic Acid Diimides. *Biochemistry*, **41**, 11379–11389.

61. D’Ambrosio,D., Reichenbach,P., Micheli,E., Alvino,A., Franceschin,M., Savino,M. and Lingner,J. (2012) Specific binding of telomeric G-quadruplexes by hydrosoluble perylene derivatives inhibits repeat addition processivity of human telomerase. *Biochimie*, **94**, 854–863.

62. Schoonover,M. and Kerwin,S.M. (2012) G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*, **20**, 6904–6918.

63. Moore,M.J.B., Cuenca,F., Searcey,M. and Neidle,S. (2006) Synthesis of distamycin A polyamides targeting G-quadruplex DNA. *Org Biomol Chem*, **4**, 3479–3488.

64. Bidzinska,J., Cimino-Reale,G., Zaffaroni,N. and Folini,M. (2013) G-Quadruplex Structures in the Human Genome as Novel Therapeutic Targets. *Molecules*, **18**, 12368–12395.

65. Imai,K. and Takaoka,A. (2006) Comparing antibody and small-molecule therapies for cancer. *Nat. Rev. Cancer*, **6**, 714–727.

66. Kolb,H.C., Finn,M.G. and Sharpless,K.B. (2001) Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.*, **40**, 2004–2021.

Chapter 2: Compound Design and Synthesis

The compounds synthesized belong to two different structural families. The first structural family includes the 360A-based compounds. Both the first and second generation compounds use structures similar to that of 360A (1) as the basis for the binding or "targeting" moiety. An additional group of 360A-based compounds uses a similar approach but differs from the other two libraries in the incorporation of an amine-arm linker. The second structural family is the Perylenediimide (PDI)-based compounds. Compounds which are discussed in more detail in this and subsequent chapters have been numbered for ease of reference. Those compounds which are only discussed in this chapter are numbered with a "misc" prefix. See Appendix B for the synthetic protocols used in the preparation of the discussed compounds.

BACKGROUND

360A has been shown to be selective for G4 structures adopted by human telomeric DNA over duplex DNA (roughly 16X even before correction for potential binding sites) by competition dialysis experiments (1). This high selectivity of 360A for G4 structures is conducive to it serving as the targeting portion of the probe candidates to be synthesized. Although the first generation library most closely resembles 360A, the second generation library uses approaches similar to another analogous compound (2) that has been used to fluorescently label potential G4 sites in stained cells and which has some analogues with anticancer activity (3). All of these compounds also are reasonable scaffolds because functionalization of the pyridine ring or the quinolone moieties allows for many possible variations when producing a library.

Perylene Diimides (PDIs) are brightly colored compounds that have found use in dyes and cosmetics. The aromatic, planar structure of these compounds also makes them

good candidates as ligands for G4 DNA. PDIs have multiple sites that are accessible for functionalization, particularly at either end of the molecule and within the bay region. While identifying the optimal conditions can be difficult, asymmetric functionalization is possible (4) and increases the number of potential members in prepared libraries. However, because some of these compounds have shown intrinsic DNA photocleavage activity (5), asymmetric functionalization is not strictly necessary for the development of a library of PDI-based G4 DNA selective photocleavage compounds. Several PDIs have also been shown to be able to inhibit a G4 helicase through interaction with quadruplexes (6) and have previously been shown to be ligands for G4 structures (7).

FIRST GENERATION LIBRARY

The first generation library adopted a direct approach that used the 360A scaffold with a linker region terminating in an alkyne extending up from position 6 on the central pyridine ring. This was then reacted with an azide-functionalized photoactive group via the copper-catalyzed "click" reaction (8), joining both the "targeting" moiety and the "warhead" moiety via a linker containing a triazole ring. Functionalization at position 6 on the pyridine ring was utilized due ease of synthetic feasibility although functionalization of the quinoline arms was also briefly explored (see below) because of the limited information available on which functionalization location would be optimal using these scaffolds. The linker lengths were varied in order to identify the optimal positioning of the "warhead" moiety. The general synthetic pathway used is depicted in Figures 2.1 and 2.2. The photocleavage moieties that were chosen act through several different photocleavage mechanisms. Excited benzophenones undergo efficient intersystem crossing, leading to DNA cleavage exclusively through the triplet excited state (9). Excited anthraquinones cleave DNA through electron transfer (10), while

naphthalimides cleave duplex DNA both through electron transfer and through singlet oxygen generation (11).

The synthesis of the first generation compounds began with chelidamic acid monohydrate (**1**), which was converted to the known diester **2** (12) through reaction with thionyl chloride in methanol. This diester was then subjected to Mitsunobu coupling with a series of terminal alkyne alcohols giving compounds **3a-d**, which differ from each other only in the carbon chain linker length. The terminal alkyne alcohols were all commercially obtained except for non-8-yn-1-ol (**misc_2**), which was prepared by reacting 1-octyne with paraformaldehyde in THF to make the 2-alkynyl alcohol **misc_1** (13), which was then subjected to a lithium metal facilitated "zipper" reaction in 1,3-diaminopropane to produce **misc_2** (14). The linker-functionalized diesters **3a-d** were treated with 3-aminoquinoline and trimethylaluminum in 1,2-dichloroethane (DCE), giving the diamides **4a-d**. These diamides were then subjected to "click" reaction conditions in DMF with different azide-functionalized photoactive groups to afford the naphthalimides (reaction with azide **5**) **8a-b**, the benzophenones (reaction with azide **6**) **9a-d**, and the anthraquinones (reaction with azide **7**) **10a-b**. DMF proved to be superior to aqueous ^{t}BuOH as a solvent for these couplings, although it necessitated the use of copper triflate as a source of copper due to the low solubility of copper sulfate in DMF, which was exacerbated by the requirement for stoichiometric copper. Stoichiometric copper was needed due to the copper-chelating propensity of the pyridine-2,6-dicarboxamide (15, 16) based scaffold. Addition of copper led to an immediate shift in the coloration of the solution to a deep green and product could only be isolated once 1.1 equivalents of copper were introduced, likely because of the formation of a 1:1 compound-copper complex that depleted the amount of available copper catalyst. The formation of these complexes also complicated purification and characterization of the

product triazoles as they needed to be disrupted in order to obtain free compound (see Appendix B). The additional steps required to disrupt the complex also contributed to the modest yields of these coupling reactions.

Finally, methylation of the triazoles **8-10** in the presence of excess methyl triflate in chloroform gave trimethylated products **11a-b**, **12a-d**, and **13a-b**. The somewhat unexpected formation of the trimethylated products was indicated by the presence of two different N-methyl resonances for these compounds in the ^{1}H NMR spectra and the presence of $(M\text{-}OTf)^{+}$ ions in the MALDI MS and $(M\text{-}3TfO)^{3+}$ ions in the ESI-MS. The location of these methyl groups was established indirectly through the following experiment: A small-scale methylation of **4a** with methyl triflate followed by "click" reaction with azide **5** in water/^{t}BuOH gave a crude dimethylated product, which was characterized by ^{1}H NMR and LRMS. Comparison of the ^{1}H NMR chemical shifts of this product against those of trimethylated **11a** showed substantial differences in the linker methylene peaks best explained by methylation on the triazole ring of **11a**. It should be noted that attempts to conduct this order of reactions on a larger scale failed due to very poor mass recovery from the final copper-catalyzed coupling reactions. This was due to the difficulties in isolating these water-soluble compounds, especially in the presence of the copper salts employed in the final coupling conducted in aqueous ^{t}BuOH.

The methylation of the triazole groups of **11-13** was originally thought to have been the cause of the low binding affinity observed with most of the first generation compounds (Chapter 3). However, results of the second generation library suggest that methylation on the linker's triazole ring likely aided in binding, at least for those compounds incorporating longer linkers.

The use of a Mitsunobu reaction to functionalize the pyridine ring proved to be the most effective as compared to alkoxylation through substitution of a chlorine as considered in the amine arm linker library (see below).

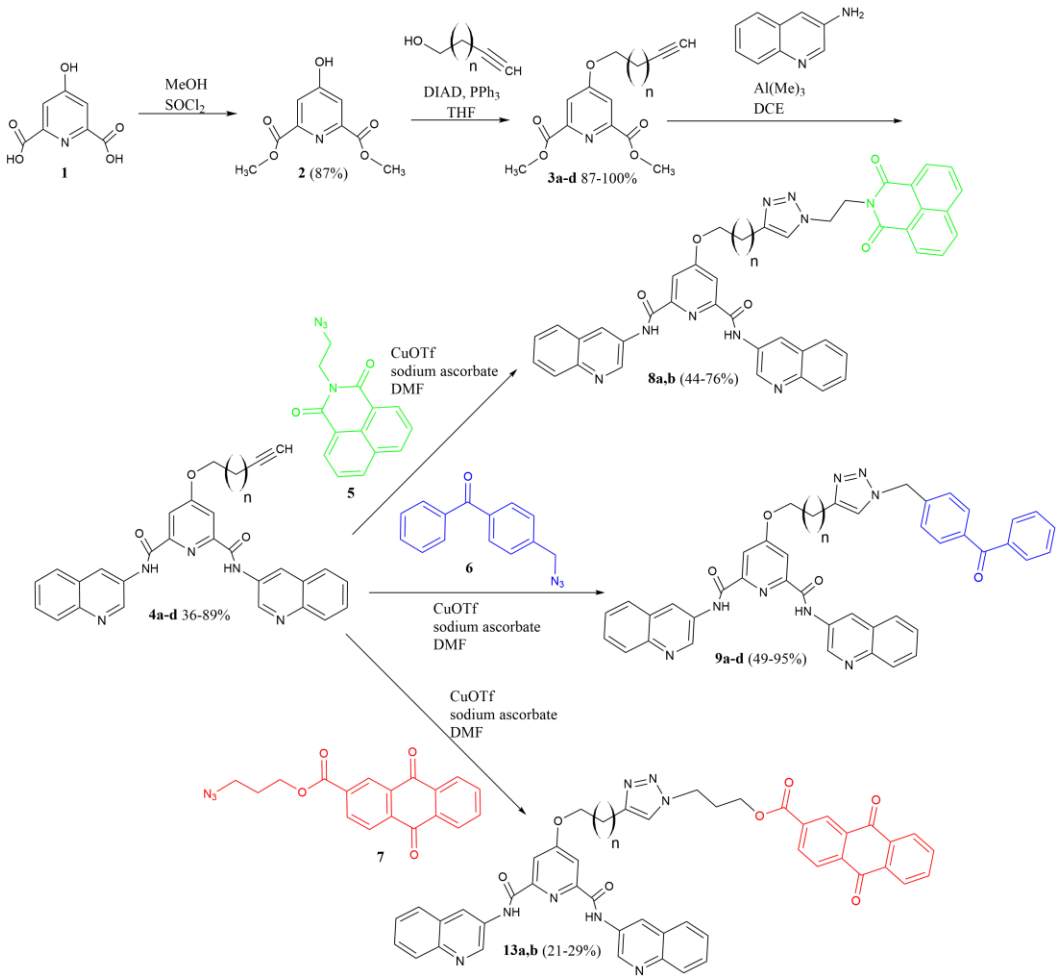


Figure 2.1: Synthesis of first generation library intermediate compounds. Photoactive moieties are colored (green for naphthalimide, blue for benzophenone, and red for anthraquinone functionalized compounds respectively) for reference with data associated with these groups in later chapters.

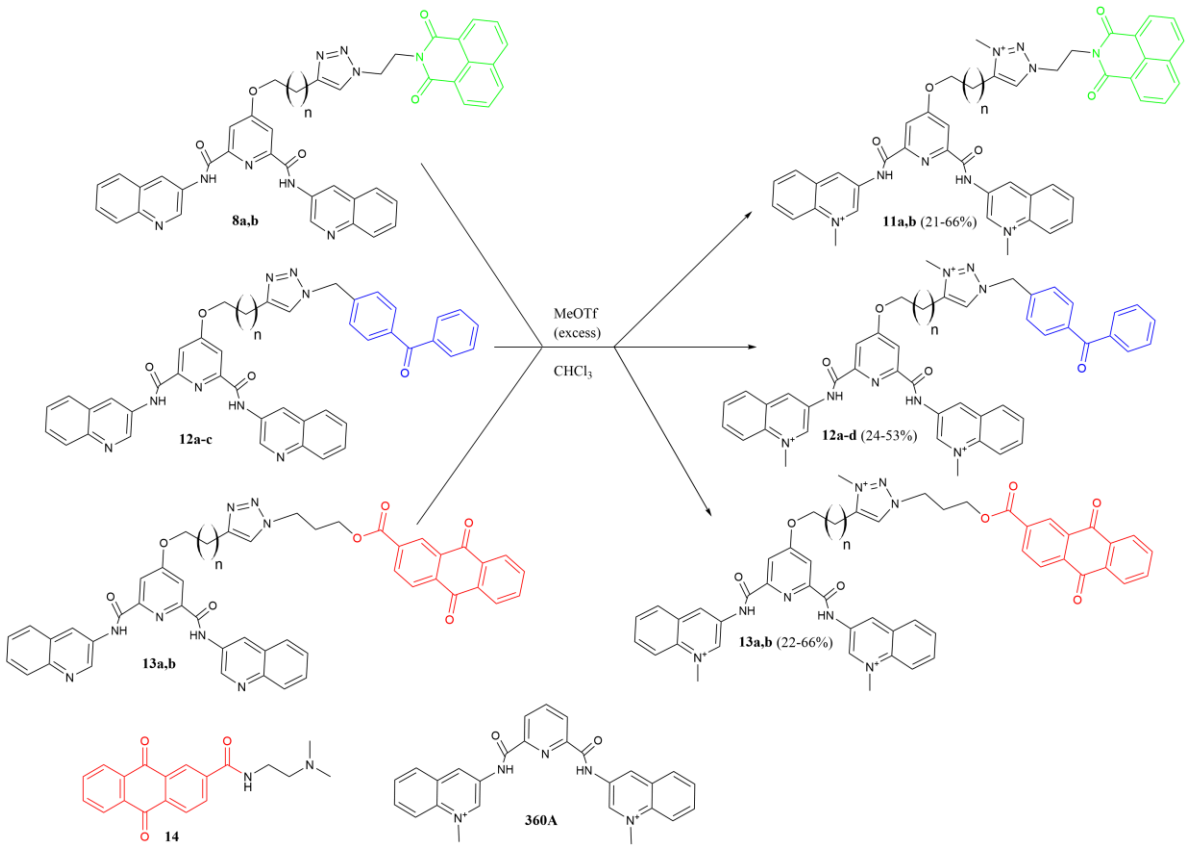


Figure 2.2: Synthesis of first generation library compounds: Photoactive moieties are colored for reference with data associated with these groups in later chapters.

The known azides **5** (17) and **6** (18) were prepared through reaction of the bromo-substituted photocleavage groups with sodium azide in DMF. Azide **7** was prepared by reaction of the acyl chloride of anthraquinone with **misc_3** in dichloromethane. The well-known azide-functionalized alcohol **misc_3** was prepared by reacting 3-chloro-1-hydroxy propane with sodium azide in DMF.

An additional compound, **14**, was synthesized to test the ability of the "warhead" moiety to independently conduct photocleavage (see Chapter 3). In this case, the acyl chloride of anthraquinone was reacted with N1,N1-dimethylethane-1,2-diamine to produce **14**.

SECOND GENERATION LIBRARY

The compounds in the second generation library were designed in an attempt to address some of the potential issues encountered during screening of the first generation library (Chapter 3). The second generation library avoided methylation. Instead, solubility and binding stabilization were achieved through functionalizing with basic (tertiary amine) arms extending from the quinoline groups. Screening showed improvement in the photocleavage activity compared with with the first generation compounds. However, the first generation library compounds were superior in binding as demonstrated by T_m increases in the G4-ligand complexes (Chapter 3).

The second generation library also began with chelidamic acid-monohydrate, and proceeded through the same diesters **3a-d** discussed above. However, instead of employing 3-aminoquinoline in the trimethylaluminum amidation reaction, the quinoline **17** was used. To prepare **17**, malononitrile was reacted with isatoic anhydride and trimethylamine in DMF, followed by treatment with hydrobromic acid to produce the hydroxyl-functionalized aminoquinoline **16**. The hydroxyl-functionalized aminoquinoline was then reacted with N,N-dimethylethanolamine in a Mitsunobu coupling reaction to give the alkoxylamine-functionalized quinoline **17**. Employing **17** in the trimethylaluminum amidation reaction gave compounds **18a-d**. The diamides **18a-d** were then reacted with different azide-functionalized photocleavage agents *via* "click" coupling to yield the naphthalimides (reaction with azide **5**) **19a-d**, the benzophenones (reaction with azide **6) 20a-c**, and the anthraquinones (reaction with azide **7**) **21a-c**. A control compound lacking a photocleavage moiety, **23**, was prepared by first reacting the diester **2** with methyl iodide in acetone to give **22**. The methoxy-functionalized diester **22** was then reacted with the alkoxylamine-functionalized quinoline **17** in the amidation reaction discussed above.

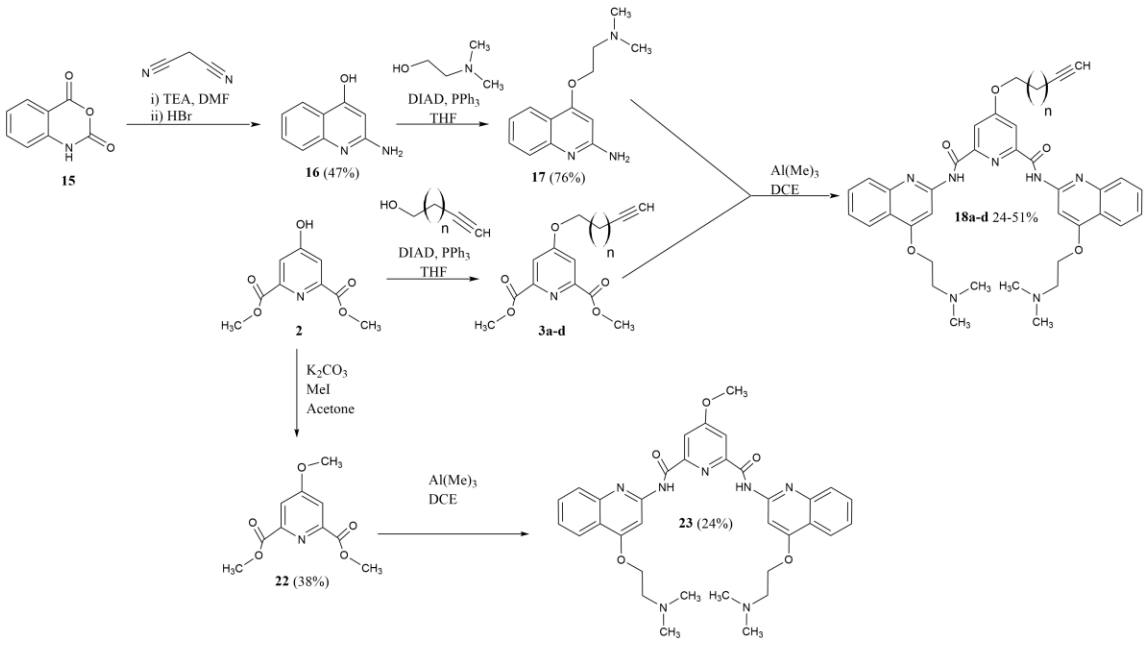


Figure 2.3: Synthesis of the second generation library intermediate compounds

The second generation compounds appeared to have an even stronger propensity for copper chelation than the first generation compounds. Their complexes proved resistant to disruption using the previous conditions. However, introduction of a competitive binding agent (see Appendix B) to the initial purification process enabled disruption of the complex and isolation of the free compound.

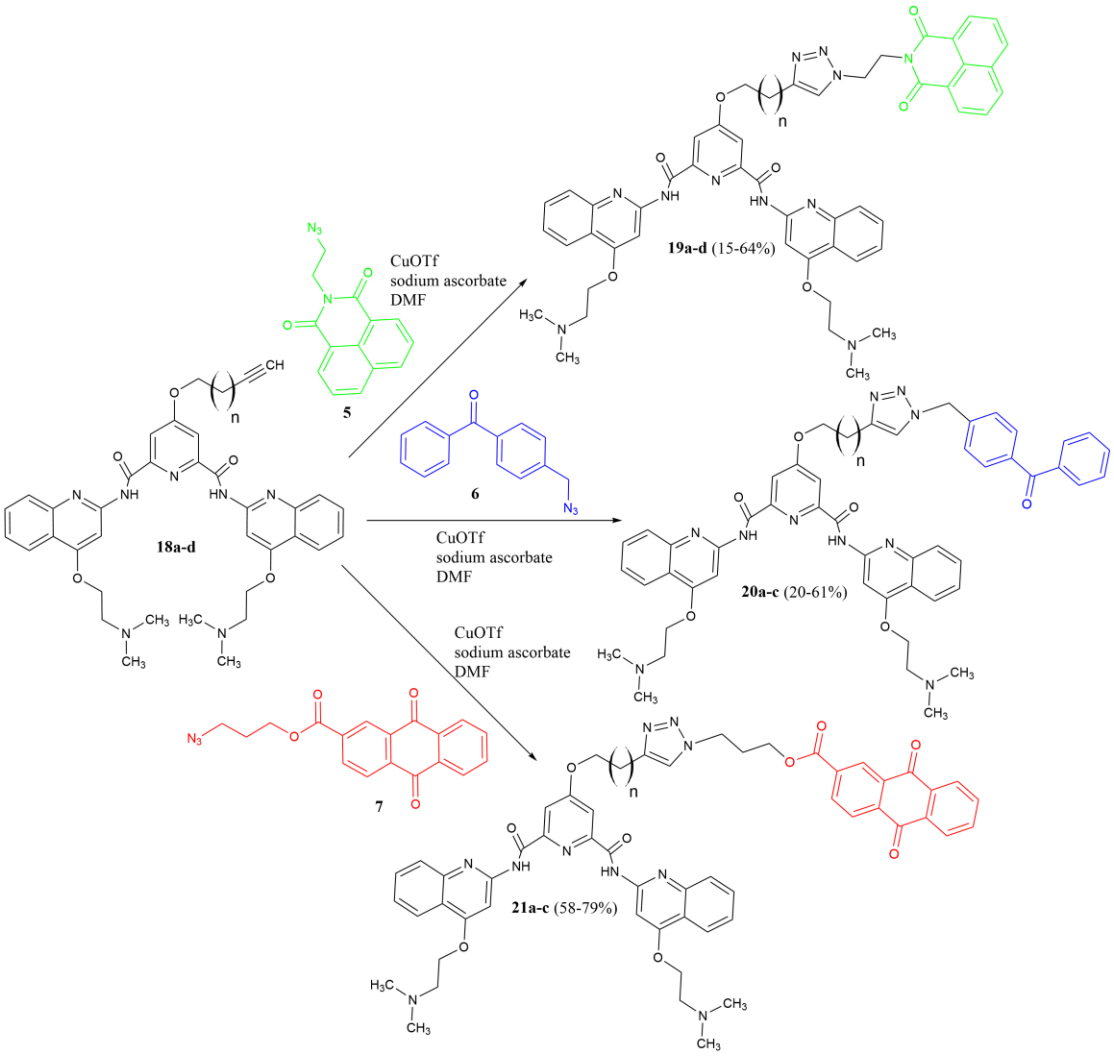


Figure 2.4: Synthesis of the second generation library compounds: Photoactive moieties are colored for reference (green for naphthalimide, blue for benzophenone, and red for anthraquinone functionalized compounds respectively) with data associated with these groups in later chapters.

OTHER 360A ANALOGUE SCAFFOLDS

When work first began on synthesizing the 360A-based first generation library, several approaches were pursued in parallel, and the first pathway to be successfully established was used to generate the final library. This section will discuss progress made in some of the competing approaches as well as some of the limitations that were encountered which slowed development of those pathways. The delays due to these

limitations prevented the use of these compounds in library generation for the studies discussed in this dissertation.

The amine arm pathway sought to increase solubility of the linker connecting the "scaffold" with the "warhead" using a 6-chloro-substituted 360A analogue **misc_4** as the scaffold intermediate to be functionalized. While **misc_4** was somewhat stable, it was sensitive to hydrolysis and suffered from relatively low solubility, limiting the solvents that could be used. Originally, this scaffold had also been considered as the primary scaffold for the pathways mentioned above. However, all attempts to alkoxylate at position 6 by reacting **misc_4** with an excess of reagent such as 4-pentyn-1-ol deprotonated by sodium hydride, failed. Keeping the alkoxide solution (which used DMF as solvent) completely anhydrous proved difficult and the presence of any water in the reaction would have introduced a competing reaction to form the undesired hydroxyl-substituted product. The same stability and solubility issues were also encountered when attempting the substitution reaction for the amine arm pathway products.

Another alternative scaffold that was briefly investigated used a 6-bromo 360A analogue **misc_16**. This particular approach would enable alternative approaches for functionalizing the scaffold at the 6 position. The 6-bromo compound was prepared by reacting chelidamic acid monohydrate with neat excess phosphorous pentabromide to yield the crude tribrominated intermediate which was reacted with 3-aminoquinoline to give the 6-bromo 360A analogue **misc_16**. The 6-bromo 360A analogue was subjected to a Sonogashira coupling with TMS-acetylene in THF to give the TMS-protected alkyne intermediate, which was deprotected by reaction with TBAF in THF to give the terminal alkyne **misc_17**. The terminal alkyne product was not reacted further because it was assumed the steric effects resulting from lack of linker would have made any "click" reaction unfavorable. However, this could become a viable scaffold with the introduction

of a linker using a dialkyne. This would require that conditions be found to reduce the competing polymerization reaction to tolerable levels, or that a monoprotected dialkyne be prepared. The 6-bromo 360A analogue **misc_16** was also methylated by reaction with excess methyl iodide in DMF/Acetone to give the dimethylated product **misc_17**. This was done with the initial intent of progressing with more polar solvents. However, the two full formal charges would likely limit the choice of solvents too severely at this stage of the synthesis. After these initial survey reactions, no further work was conducted with the 6-bromo analogues.

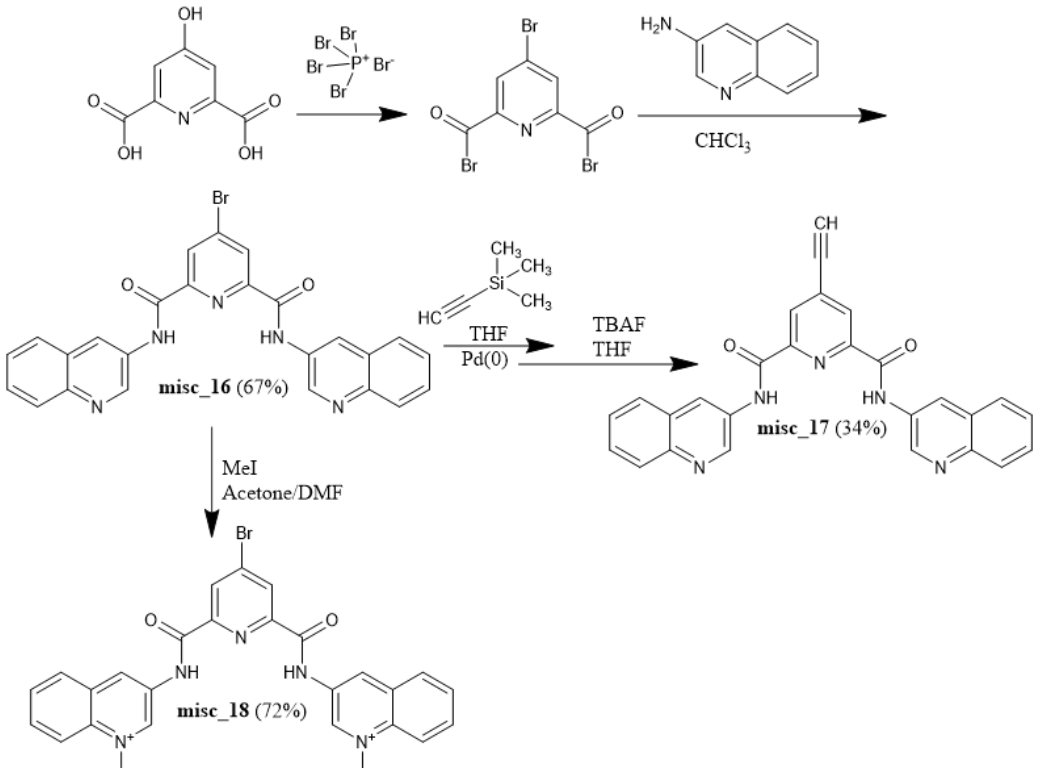


Figure 2.5: 6-Bromo scaffold based compounds

OTHER 360A ANALOGUE COMPOUNDS (AMINE ARM LINKER APPROACH)

The amine-arm-based synthesis sought to use a diamine-based linker for the 6-position functionalization, which was expected to improve the compounds' water

solubility dependent on the pH of the solution. This synthetic pathway required several extra steps due to the required amine protection/deprotection steps which slowed its development compared with the pathways used to generate the first and second generation libraries. The first protection pathway used a single phthalimide protecting group. This group was applied by reacting N1-ethylethane-1,2-diamine with phthalic anhydride in water (19) to give the amide intermediate. Ring closure to the imide was achieved by refluxing the amide in glacial acetic acid, giving the phthalimide-protected amine **misc_6**. Generally, **misc_6** was prepared shortly before use in other reactions.

The phthalimide-protected amine **misc_6** was reacted with chloroacetylchloride and trimethylamine in methylene chloride to give the alpha-chloro amide product **misc_7**. Reaction of the alpha-chloro amide product **misc_7** with pentyn-1-ol and sodium hydride in THF failed to form the desired ether linkage product **misc_8** (n=3), likely in part because of insufficient stability of the protecting group in the highly basic reaction conditions.

As an alternative to **misc_8**, the phthalimide protected amine **misc_6** was reacted with 6-heptanoyl chloride (generated in situ by reacting 6-heptanoic acid with oxalyl chloride in methylene chloride) and trimethylamine in methylene chloride to give the amide **misc_9**. The amide was then reacted with hydrazine in ethanol in order to remove the phthalimide protecting group to give the free primary amine **misc_10**. The reaction of the 6-chloro 360A-analogue **misc_4** with free primary amine **misc_10** and trimethylamine in DMSO was unsuccessful in generating the desired amine-arm functionalized product **misc_11**, likely because of the limited solubility of **misc_4.** Often efforts to dissolve **misc_4** only lead to hydrolysis and generation of the 6-hydroxyl 360A analogue. At this point in the amine-linker pathway development, the pathways for the

first and second generation libraries were successfully established, and this approach was not pursued further.

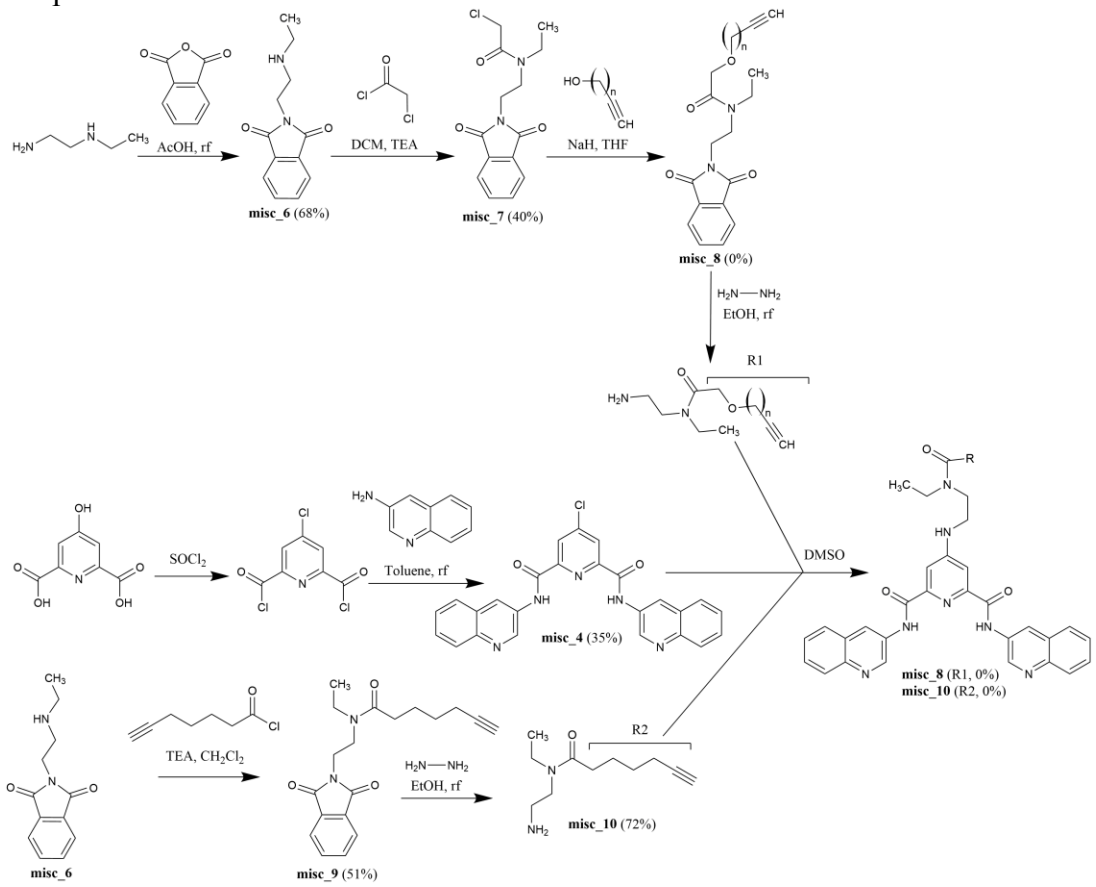


Figure 2.6: Synthesis of "Amine Arm Linker" compounds

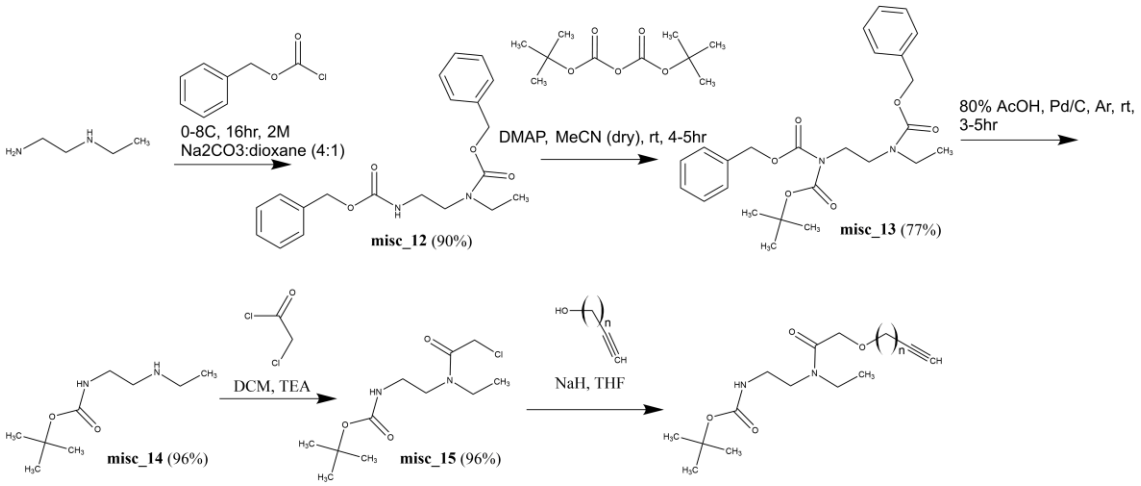


Figure 2.7: Alternative protection pathway for synthesis of the "Amine Arm Linker" compounds

A second protection pathway was considered for the amine-arm project which might be more amenable to the reaction conditions needed for the final functionalization at the 6 position. This involved multiple protection and deprotection steps in order to selectively mono-Boc protect the primary amine of N1-ethylethane-1,2-diamine. First, N1-ethylethane-1,2-diamine was reacted with benzyloxycarbonyl chloride (20) in dioxane to give the di-CBz protected product **misc_12**. Reaction of **misc_12** with DMAP and di-tert-butyl dicarbonate in acetone (20) gave the di-CBz-mono-Boc protected product **misc_13**. CBz deprotection of **misc_13** was accomplished by reduction under H_2 over Pd/C in aqueous acetic acid (20) to give the mono-Boc protected product **misc_14**. With mono-Boc-protection successful, **misc_14** was reacted with chloroacetylchloride and trimethylamine in methylene chloride to give the alpha-chloro amide product **misc_15**. The pathways used for the synthesis of the first and second generation libraries were complete at this point and no further work was done with this pathway.

The amine arm pathway suffered from two major obstacles. The first was the low solubility of the 6-chloro 360A analogue **misc_4**, which complicated the substitution

reaction to attach the amine arm to the quadruplex-binding scaffold. The second obstacle was the difficulties involved with the generation of the beta-ether linkage needed to attach the terminal alkyne functionality to the amine arm. The latter may have been partially addressed by the mono-Boc protection approach.

OTHER 360A ANALOGUE COMPOUNDS (QUINOLINE FUNCTIONALIZATION APPROACH)

As mentioned briefly above, functionalization of the quinolones on the 360A scaffold has the potential to generate a library of interesting compounds. The second generation library utilizes such functionalization to increase the compound solubility and give enhanced binding. However, attachment of the photoactive groups *via* a linker to the quinoline groups could also have interesting effects, especially if this could be accomplished asymmetrically. It would give a distinct structural alternative to the commonly used 6-position functionalization. Some of these possibilities were briefly explored while the library preparation pathways were being investigated.

Quinolines functionalized at positions 6, 7, or 8 proved to be expensive to purchase from commercial sources, however a small number were purchased for these survey reactions. 4-chloro-6-nitroquinoline was reacted with the alkoxide of 4-pentyn-1-ol (generated in situ with sodium hydride) in DMF to give the 4-alkoxy quinoline **misc_19**. This 4-alkoxy quinoline was reduced with sodium sulfide nonahydrate in aqueous ethanol after no product was isolated under other reducing conditions (H_2 over Pd/C), giving a low yield of 4-alkoxy-6-amino quinoline **misc_20**. It was not clear from the crude ^{1}H NMR that the reduction had been successful, so the crude **misc_20** was reacted with acetic anhydride and catalytic DMAP in methylene chloride to give the acetylated product. The acetylated product exhibited the expected acetyl methyl peak in its ^{1}H NMR, giving support that the 4-alkoxy-6-amino quinoline **misc_20** had been

successfully synthesized. However, repeated reduction attempts resulted in similarly ambiguous results despite efforts to fully purify the product. These results combined with relatively low apparent yields and exacerbated by the expense and low amount of remaining starting material discouraged further investigation.

A series of compounds incorporating 3-aminoquinoline functionalized with bromine at positions 6, 7, or 8 were also investigated as an alternative pathway towards the functionalization of the quinoline arms. 2,6-dicarboxypyridine was converted in situ to the diacyl chloride through treatment with thionyl chloride. The diacyl chloride was reacted with bromo-substituted quinolines in toluene to give diamides **misc_21** (6-bromo), **misc_22** (7-bromo), and **misc_23** (8-bromo). The diamides were reacted with methyl iodide in DMF in a failed attempt to produce the dimethylated products **misc_24** (6-bromo), **misc_25** (7-bromo), and **misc_26** (8-bromo). Lower solubility of the diamides likely contributed to the inability to isolate the desired material. The diamides were also subjected to a Sonogashira coupling reaction in TEA with octadiyne in a failed attempt to produce the dialkylated products **misc_27** (6-substituted), **misc_28** (7-substituted), **misc_29** (8-substituted). While lower solubility would have contributed to the inability to isolate the desired material, competing polymerization product would also have acted to lower the reaction yield. Limited starting material discouraged further investigation.

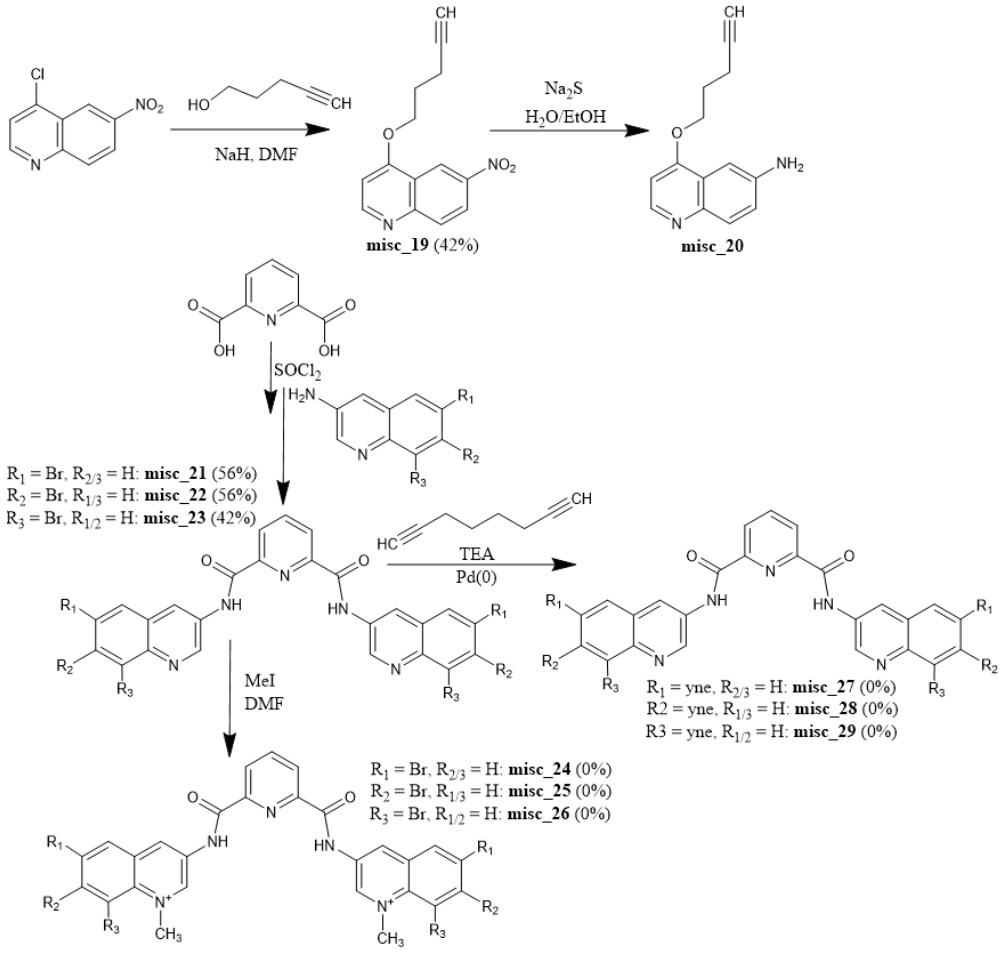


Figure 2.8: Quinoline functionalization survey compounds

PERYLENEDIIMIDE (PDI) ANALOGUE COMPOUNDS

The beginnings of several different pathways for the generation of a PDI-based library were surveyed. The intrinsic photocleavage activity of PDIs coupled with a reasonable affinity for G4 structures makes them good candidates for use as G4 photocleavage agents. The multiple available sites for functionalization also allow several opportunities for diversifying libraries based on these compounds. The primary difficulty was found in finding conditions that could reliably form the desired imide product (either symmetrically or asymmetrically substituted). The conditions (21–23)

attempted for imidation of perelene-3,4,9,10-tetracorboxylic acid dianhydride were insufficient to allow isolation of the desired material. However, the nature of the amine used for these imidation reactions has a large effect on the reaction yields (4). In an alternative approach, perelene-3,4,9,10-tetracorboxylic acid dianhydride was reacted with the "swallow tail" amine pentan-3-amine (24) with melted imidazole as solvent giving the swallow-tail-substituted PDI **misc_34**. The use of a "swallow tail" amine in this reaction allowed imidation to be readily accomplished (although not asymmetrically) and also gave a product that was soluble in several common solvents. Using this strategy also involved a shift towards functionalizing the PDI in the "bay" region while the amide functionalizations were used instead to increase solubility. The swallow-tail-substituted PDI **misc_34** was brominated (25) using elemental bromine in methylene chloride to give the mono-brominated product **misc_35**, with the site of bromination driven by steric effects. This final approach showed promise, as the monobromination in the bay region would allow for the potential generation of a library of asymmetrically functionalized PDIs. Possible future directions will be discussed further in Chapter 6.

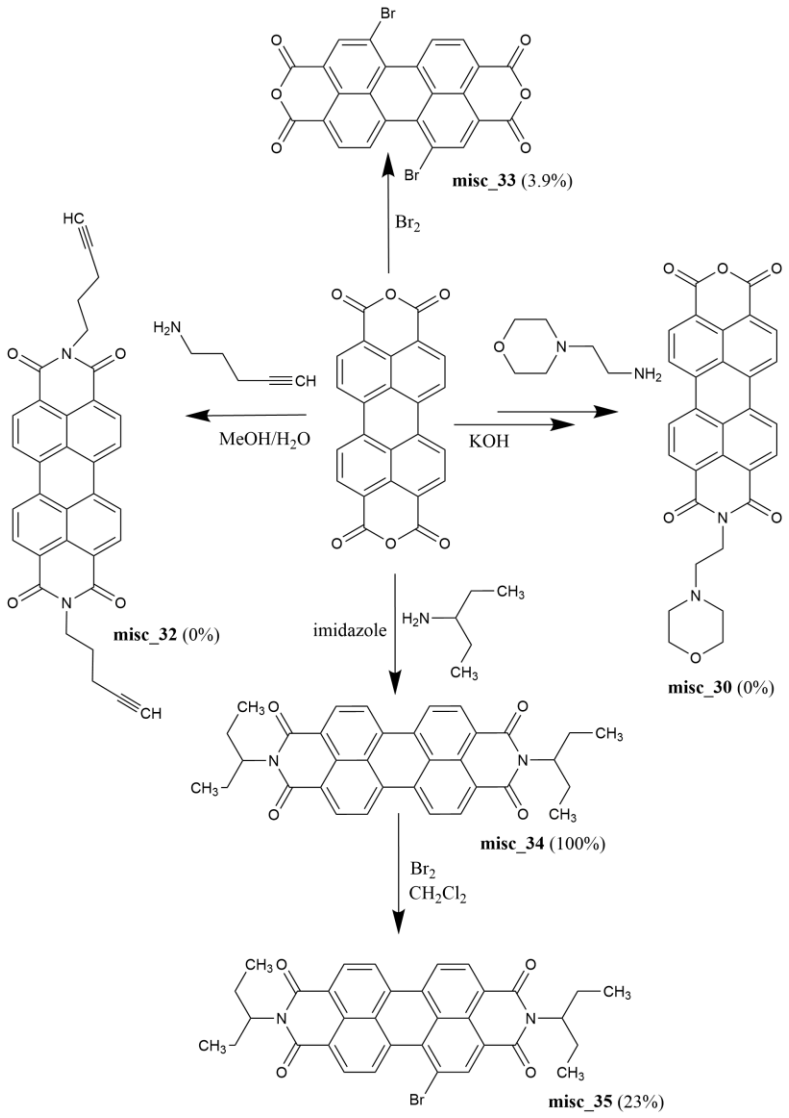


Figure 2.9: Compounds from brief survey of potential PDI functionalization pathways.

As is discussed in Chapter 3, photocleavage by the compound TEL011 (4), appeared to be largely dependent on an aggregated state. This prompted the investigation into PDI-based macrocycles, which could potentially act as "pre-aggregated" PDIs with superior concentration-dependent photocleavage activity. Unfortunately, such macrocycle reactions tend to have very low yields because of the competing formation of polymeric side product. The Ziegler-Rügli method (26) attempts to minimize the

polymeric product formed by using extremely dilute reaction conditions coupled with the slow addition of the reagents over the course of several weeks. Perelene-3,4,9,10-tetracorboxylic acid dianhydride was reacted with 1,10-diaminodecane in DMF *via* the Ziegler- Rügli method to give the alkyl chain linked PDI macrocycle **24**. Unsurprisingly, **24** exhibited very poor solubility in the aqueous conditions used for photocleavage screening (Chapter 3), where it displayed no discernable photocleavage activity. To address this issue, the linking group was altered to improve water solubility. Perelene-3,4,9,10-tetracorboxylic acid dianhydride was reacted with 3,3'-(piperazine-1,4-diyl)bis(propan-1-amine) in DMF *via* the Ziegler-Rügli method in an attempt to produce the more water soluble PDI macrocycle **misc_36**. However, none of the desired product could be isolated, even with the the addition of barium as a catalyst to aid in macrocycle formation.

Even using the Ziegler-Rügli method, the macrocycle yields were very low, due to the formation of comparatively large amounts of the competing polymerized product. In addition, these low yields can be partially attributed to the size of the reaction setup and the long reaction times which limited the number of different conditions that could be investigated simultaneously. The large size of the assembled apparatus also required more vigorous refluxing of DMF in order to ensure continued addition of the reagents. The long reaction times and high temperature may have led to partial decomposition of DMF, which in turn may have further exacerbated the already very low expected yields.

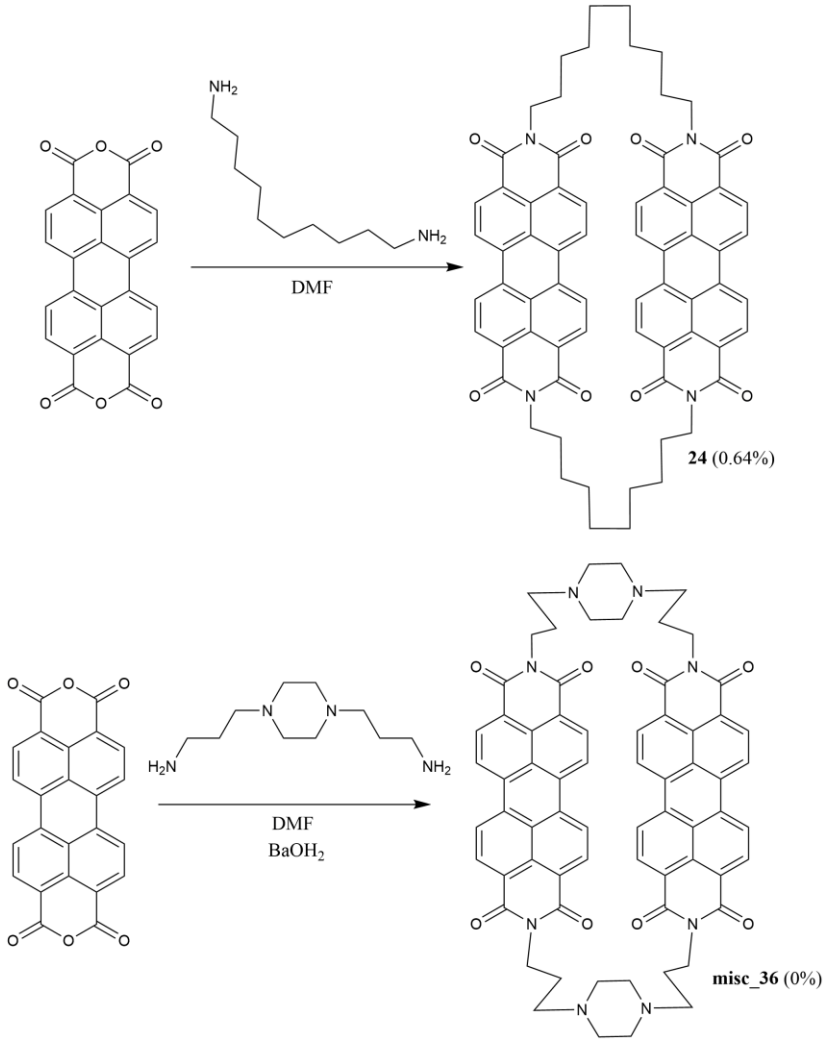


Figure 2.10: Additional structures synthesized for use in the studies conducted.

CONCLUSIONS

The use of pyridine-2,6-dicarboxamide as the core scaffold for the first and second generation libraries presented some difficulties due to the coordination of copper, which lowered the yield of isolated free compound. However, the high selectivity for G-quadruplexes shown by analogues incorporating this scaffold made it promising to investigate. While the copper complexes were not investigated in this work, the impact

of coordination with copper or other ions on G4 structure binding and perhaps DNA cleavage would be interesting to investigate further.

Several alternate approaches were considered when preparing the libraries investigated further in this dissertation. While the most expedient approach was used several of the others showed promise. The "amine-arm" approach was limited by difficulties encountered at two steps. The first being the formation of the ether that would allow for varying carbon chain lengths to be incorporated, and the second at the attachment of the arm to the scaffold itself. The alternate protection approach that was initiated may have already addressed some of the difficulties with the first obstacle. The greatest difficulty with the attachment of the arm seemed to be the solubility of the core scaffold structure. This could perhaps be improved by changing the scaffold's quinoline groups to incorporate a tertiary amine as was done in the second generation library. This should increase the number of solvents that could be considered while also improving solubility in those solvents already investigated.

The other approach that showed the most promise towards producing another library is that based on "swallow-tail" functionalized PDIs through substitution in the bay region. While the reaction will need to optimized, successful asymmetrical bromination was achieved, allowing future steps to be asymmetrical. This could prove to be advantageous when dealing with chiral G4 structures.

CHAPTER 2 REFERENCES

1. Granotier,C., Pennarun,G., Riou,L., Hoffschir,F., Gauthier,L.R., De Cian,A., Gomez,D., Mandine,E., Riou,J.-F., Mergny,J.-L., *et al.* (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.*, **33**, 4182–4190.

2. Rodriguez,R., Miller,K.M., Forment,J.V., Bradshaw,C.R., Nikan,M., Britton,S., Oelschlaegel,T., Xhemalce,B., Balasubramanian,S. and Jackson,S.P. (2012)

Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.*, **8**, 301–310.

3. Müller,S., Sanders,D.A., Antonio,M.D., Matsis,S., Riou,J.-F., Rodriguez,R. and Balasubramanian,S. (2012) Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org. Biomol. Chem.*, **10**, 6537–6546.

4. Pasaogullari,N., Icil,H. and Demuth,M. (2006) Symmetrical and unsymmetrical perylene diimides: Their synthesis, photophysical and electrochemical properties. *Dyes Pigments*, **69**, 118–127.

5. Schoonover,M. and Kerwin,S.M. G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*, 10.1016/j.bmc.2012.10.017.

6. Tuesuwan,B., Kern,J.T., Thomas,P.W., Rodriguez,M., Li,J., David,W.M. and Kerwin,S.M. (2008) Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents. *Biochemistry*, **47**, 1896–1909.

7. Kern,J.T., Thomas,P.W. and Kerwin,S.M. (2002) The Relationship between Ligand Aggregation and G-Quadruplex DNA Selectivity in a Series of 3,4,9,10-Perylenetetracarboxylic Acid Diimides. *Biochemistry*, **41**, 12568–12568.

8. Rostovtsev,V.V., Green,L.G., Fokin,V.V. and Sharpless,K.B. (2002) A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chem.-Int. Ed.*, **41**, 2596–+.

9. Cuquerella,M.C., Lhiaubet-Vallet,V., Cadet,J. and Miranda,M.A. (2012) Benzophenone Photosensitized DNA Damage. *Acc. Chem. Res.*, **45**, 1558–1570.

10. Breslin,D.T. and Schuster,G.B. (1996) Anthraquinone Photonucleases: Mechanisms for GG-Selective and Nonselective Cleavage of Double-Stranded DNA. *J. Am. Chem. Soc.*, **118**, 2311–2319.

11. Aveline,B.M., Matsugo,S. and Redmond,R.W. (1997) Photochemical Mechanisms Responsible for the Versatile Application of Naphthalimides and Naphthaldiimides in Biological Systems. *J. Am. Chem. Soc.*, **119**, 11785–11795.

12. Dorazco-González,A., Höpfl,H., Medrano,F. and Yatsimirsky,A.K. (2010) Recognition of Anions and Neutral Guests by Dicationic Pyridine-2,6-dicarboxamide Receptors. *J. Org. Chem.*, **75**, 2259–2273.

13. McPhee,M.M. and Kerwin,S.M. (1996) Synthesis and Metal Ion Binding Studies of Enediyne-Containing Crown Ethers. *J. Org. Chem.*, **61**, 9385–9393.

14. Triple-bond Isomerizations: 2- to 9-decyn-1-ol (1988) *Org. Synth.*, **66**, 127.

15. Rajput,A. and Mukherjee,R. (2013) Coordination chemistry with pyridine/pyrazine amide ligands. Some noteworthy results. *Coord. Chem. Rev.*, **257**, 350–368.

16. Shao,Y., Sheng,X., Li,Y., Jia,Z.-L., Zhang,J.-J., Liu,F. and Lu,G.-Y. (2008) DNA Binding and Cleaving Activity of the New Cleft Molecule N,N′-Bis(guanidinoethyl)-2,6-pyridinedicarboxamide in the Absence or in the Presence of Copper(II). *Bioconjug. Chem.*, **19**, 1840–1848.

17. Lee,L.V., Mitchell,M.L., Huang,S.-J., Fokin,V.V., Sharpless,K.B. and Wong,C.-H. (2003) A Potent and Highly Selective Inhibitor of Human α-1,3-Fucosyltransferase via Click Chemistry. *J. Am. Chem. Soc.*, **125**, 9588–9589.

18. Chenot,E.-D., Rodriguez-Dominguez,J.C., Hannewald,P., Comel,A. and Kirsch,G. (2008) Synthesis of potentially photoactivatable cournarin derivatives via a 1,3-dipolar cycloaddition. *J. Heterocycl. Chem.*, **45**, 1429–1435.

19. Li,H., Hao,M., Wang,L., Liang,W. and Chen,K. (2009) Preparation of Mono Boc-Protected Unsymmetrical Diamines. *Org. Prep. Proced. Int.*, **41**, 301–307.

20. Almeida,M., Grehn,L. and Ragnarsson,U. (1988) Selective Protection of Polyamines - Synthesis of Model Compounds and Spermidine Derivatives. *J. Chem. Soc.-Perkin Trans. 1*

21. Huang,H., Che,Y. and Zang,L. (2010) Direct synthesis of highly pure perylene tetracarboxylic monoimide. *Tetrahedron Lett.*, **51**, 6651–6653.

22. Amiralaei,S., Uzun,D. and Icil,H. (2008) Chiral substituent containing perylene monoanhydride monoimide and its highly soluble symmetrical diimide: synthesis, photophysics and electrochemistry from dilute solution to solid state. *Photochem. Photobiol. Sci.*, **7**, 936–947.

23. Asir,S., Demir,A.S. and Icil,H. (2010) The synthesis of novel, unsymmetrically substituted, chiral naphthalene and perylene diimides: Photophysical, electrochemical, chiroptical and intramolecular charge transfer properties. *Dyes Pigments*, **84**, 1–13.

24. Langhals,H., Krotz,O., Polborn,K. and Mayer,P. (2005) A Novel Fluorescent Dye with Strong, Anisotropic Solid-State Fluorescence, Small Stokes Shift, and High Photostability. *Angew. Chem. Int. Ed.*, **44**, 2427–2428.

25. Rajasingh,P., Cohen,R., Shirman,E., Shimon,L.J.W. and Rybtchinski,B. (2007) Selective bromination of perylene diimides under mild conditions. *J. Org. Chem.*, **72**, 5973–5979.

26. Langhals,H. and Ismael,R. (1998) Cyclophanes as Model Compounds for Permanent, Dynamic Aggregates – Induced Chirality with Strong CD Effects. *Eur. J. Org. Chem.*, **1998**, 1915–1917.

Chapter 3: Compound Screening

The compounds in each of the synthesized libraries were screened both for their G4 photocleavage activity and for their ability to stabilize G4 DNA structures as indicated through elevated melting temperatures of the complex compared with the DNA structure itself.

BACKGROUND

The design and preparation of the compounds discussed in this chapter was covered in Chapter 2. This chapter uses the same color and number identifiers for the compounds as were used in the previous chapter. In addition to the compounds discussed in Chapter 2, the perylene diimide TEL011 (1–3) and porphyrin NMM (4–8) were also studied because of previous evidence of selective binding and/or photocleavage of G4 structures.

Two different dual-labeled oligonucleotides, both established to form intramolecular G4 structures in vitro (3), were used for the screening of the compounds. On the 5' end, the oligos were labeled with fluorescein (FAM), while on the 3' end, they were labeled with 5-carboxytetramethylrhodamine (TAMRA). These two fluorescent dyes can undergo Förster resonance energy transfer (FRET) when the two dyes are in close proximity to each other, during which the fluorescence of FAM is quenched by TAMRA. When the DNA is folded into a G4 structure, the dyes are brought close enough to effectively quench FAM fluorescence, giving a concentration-dependent means of measuring unfolding of G4 structures in solution by measuring the FAM fluorescence. Because of this useful characteristic, FRET quenching is used as the basis for most of the assays discussed in this chapter.

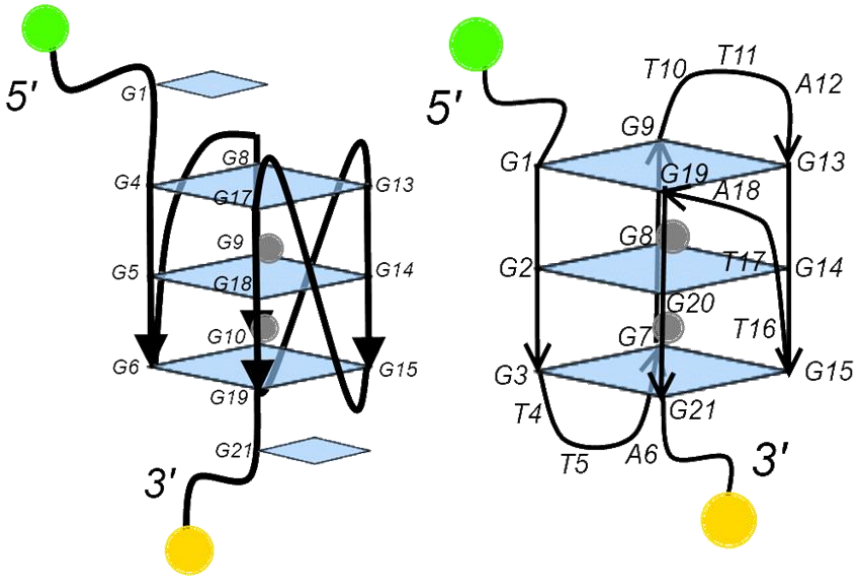


Figure 3.1: Diagrams of the folded forms of the oligonucleotides used in this study. The green circles represent FAM, the yellow circles represent TAMRA, and blue squares represent guanines (large squares represent hydrogen-bonded tetrads incorporated four guanines). Arrows indicate the "flow" of the strand folding from the 5' end to the 3' end. Gray circles represent the locations of the stabilizing monovalent ions. cMYC, left, adopts the structure depicted in both sodium and potassium buffers. F21T, right, adopts the mixed anti-parallel basket structure shown in potassium buffer. This conformation changes in sodium buffers.

The first of the oligonucleotides used was a twenty-one nucleotide human telomeric sequence known as F21T (5'-FAM-dGGG(TTAGGG)$_3$-TAMRA-3'). The second, referred to as cMYC (5'-FAM-dTGAGGGTGGGTAGGGTGGGTAG-TAMRA-3') represents a sequence modified from that found within the NHEIII$_1$ region upstream of the human cMYC proto-oncogene promoter that has been shown to control 90% of all transcription by that gene (9). The modification is a single base mutation that was shown to remove the polymorphism of the G4 structures and force the sequence into a parallel-stranded G4 structure (10).

The photocleavage screening and DNA melt assays discussed in this chapter have been published previously (11). Details on the methods used can be found in Appendix A.

In the competition melts that were conducted, several different concentrations were used for the different oligonucleotides. The concentrations used are important because the structures adopted by the different oligonucleotides vary and thus the concentration of the potential binding sites is not necessarily equivalent to the molecular concentration. Therefore, it is common practice to report concentrations in terms of major substructures (e.g. in terms of tetrads in the case of G-quadruplexes). This simplifies comparisons between intramolecular G4 structures (F21T and cMYC) and intermolecular structures ($T_2G_{20}T_2$, see Table 2.1 below). But G4 structures could just as easily be reported in terms of ends (the "top" and "bottom" of a formed G4 structure), strands, or nucleotides. The concentration used can become important when interpreting the results for experiments. Table 2.1 demonstrates how DNA having the same concentration in one context can vary drastically if concentration is considered in another context. For the photocleavage experiments and melts in this chapter where only intramolecular G4 structures are considered, G4 DNA concentration is reported in terms of strands. For melts involving unlabeled competitor DNA (Chapter 4), G4 DNA concentration is reported in terms of tetrads, duplex DNA concentrations are reported in terms of base pairs (bp), and triplex DNA is reported in terms of triplets.

Table 3.1: DNA Concentration Conversion Table (all units in nM)

nX	**Tetrads F21T/cMYC**	**Ends F21T/cMYC**	**Strands F21T/cMYC**	**G4 F21T/cMYC**	
1	600	400	200	200	
nX	**Tetrad T2G20T2**	**Ends T2G20T2**	**Strand T2G20T2**	**G4 T2G20T2**	**bp duplex (CT DNA)**
0.5	300	30	60	15	300
1	600	60	120	30	600
2	1200	120	240	60	1200
5	3000	300	600	150	3000
10	6000	600	1200	300	6000
50	30000	3000	6000	1500	30000
100	60000	6000	12000	3000	60000

PHOTOCLEAVAGE ASSAYS

TMPyP4, which is a relatively potent DNA photocleavage compound, was used as a control to show the viability of the photocleavage screening assay. These experiments also illustrate some of the considerations that must be made when interpreting the data. For example, the concentration range used can have important impacts on the observed signals. Figure 3.2 shows the change in apparent percent cleavage observed for F21T after treatment with different concentrations of TMPyP4 with increasing irradiation times. The apparent percent cleavage increases with increasing concentration as expected, but increasing irradiation times lead to a maximum amount of cleavage where nearly all cleavage sites appear to have been cut. Above this concentration/irradiation time combination, the apparent photocleavage no longer increases, and because of photobleaching, may actually give a decreasing signal at very high concentrations of the photocleavage compound and long irradiation times. Comparing these results with the gel results (Figure 3.2 B), it can be seen that after reaching the saturated signal point, the assay is underestimating cleavage by about 10-20%. Below this saturation point, the gel and solution data are in agreement.

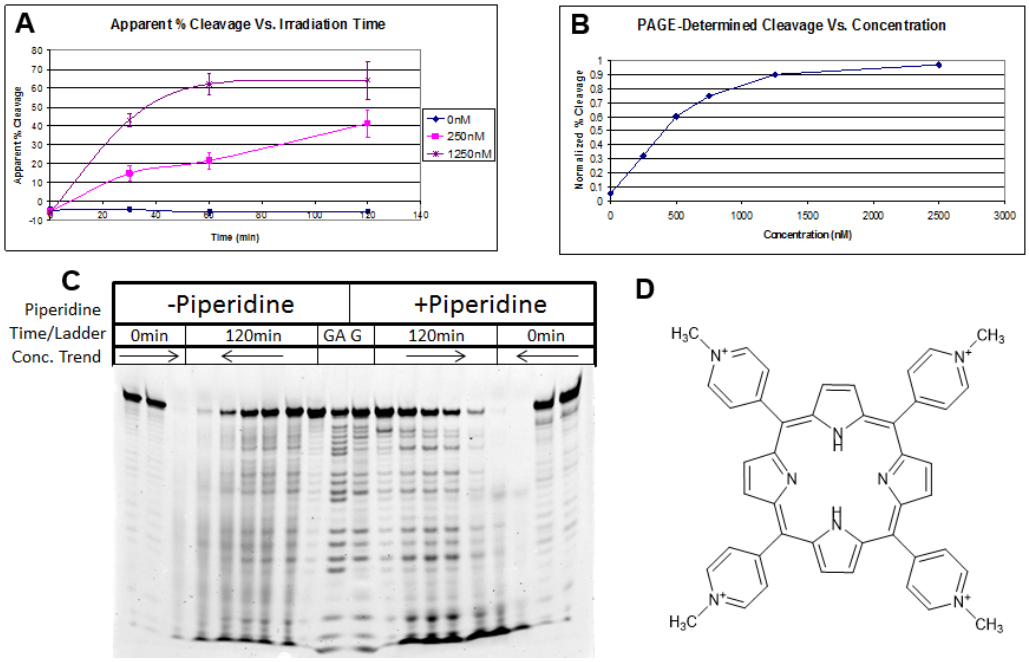


Figure 3.2: Photocleavage of cMYC by TMPyP4. (A) Apparent percent cleavage of cMYC by TMPyP4 as detected by solution fluorescence in microwell plates (B) Normalized photocleavage of cMYC by TMPyP4 as determined from the gel (C) Gel of photocleavage products from cMYC treated with TMPyP4 and irradiated with 420nm-centered lamps (D) Structure of TMPyP4.

The first generation library was analyzed primarily by PAGE. As seen by Figure 3.3, the library exhibited low to moderate photocleavage activity under the conditions of the assay and cMYC was cleaved more readily than was F21T. Increasing concentration had little effect on the amount of photocleavage observed. Compounds which used benzophenone as the "warhead" appeared to cleave slightly better than the other functional groups and the highest levels of cleavage were seen with those compounds incorporating an intermediately sized linker length (Figure 3.4). The first generation library compounds cleaved better when irradiated under the UVA-centered lamps rather than the UVB-centered lamps (their absorption spectra showed a maximum near the interface between UVB and UVA (Appendix C)).

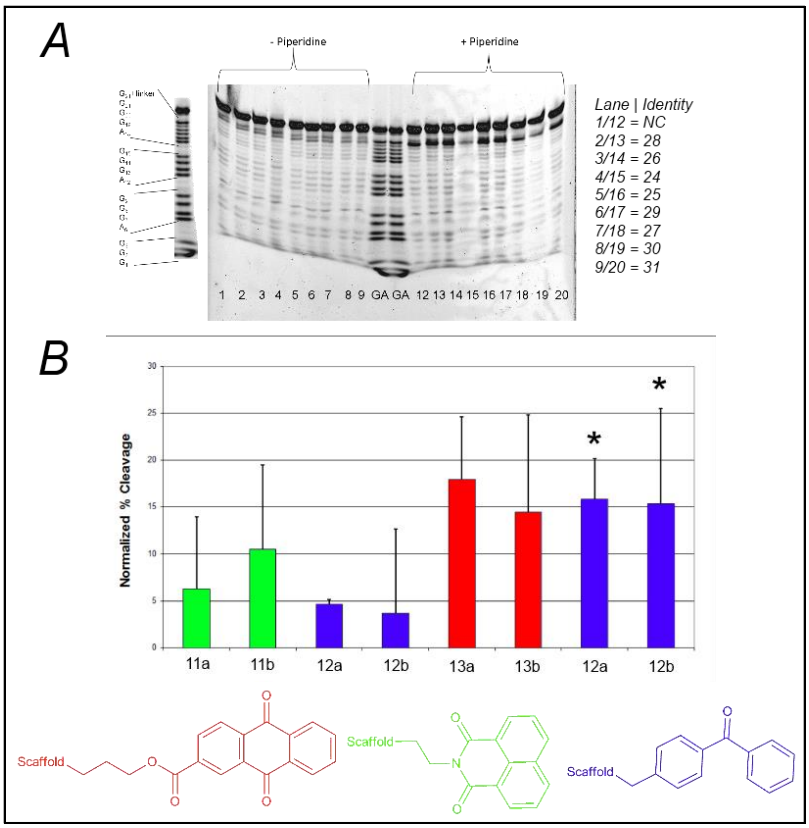


Figure 3.3: F21T UVA first generation photocleavage data. (A) Example photocleavage gel of F21T after 30min UV irradiation of DNA incubated with 500nM compound, (B) F21T cleavage data as determined from gels. Red, green, and blue bars correspond to compounds incorporating anthraquinone, naphthalimide, and benzophenone respectively. NC stands for "No Compound". *cMYC photocleavage data for comparison.

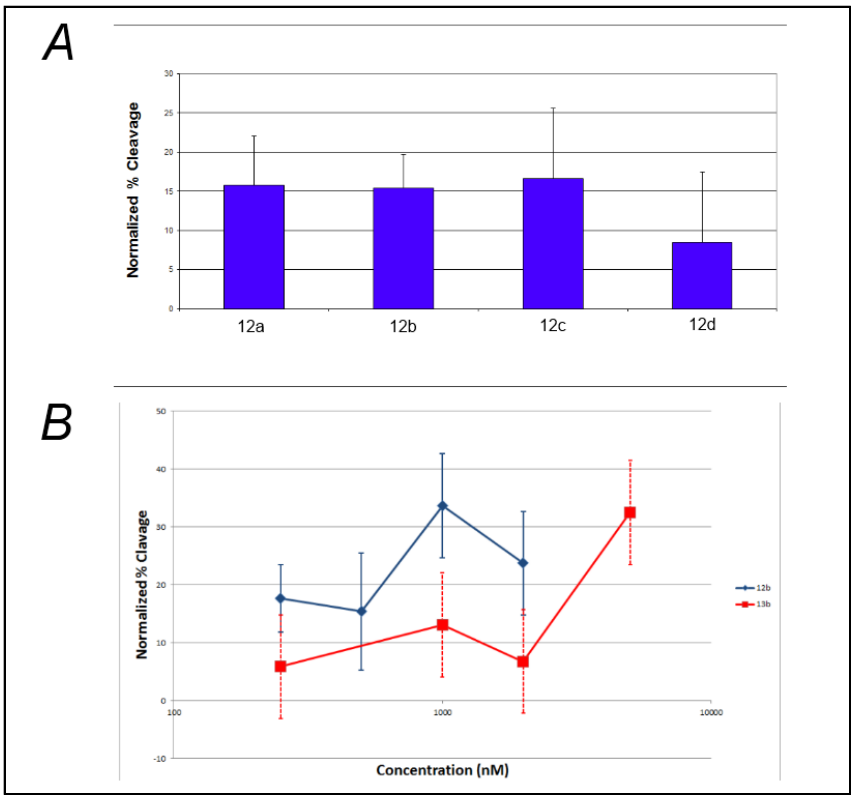


Figure 3.4 cMYC UVA First Generation Photocleavage Data. (A) Effect on photocleavage of increasing the linker length for 500nM treatments of the benzophenone-incorporated compounds, (B) Photocleavage dependence on concentration for benzophenone or anthraquinone-incorporated compounds.

The presence of negative apparent percent cleavage for many of the compounds in the solution-derived data suggest that many of the first generation compounds interact with the fluors used for detection, causing quenching of the fluorescence and the appearance that cleavage with the compound is below that of background. This effect's prevalence among the first generation compounds required analysis to be primarily conducted via PAGE. To verify that cleavage could be detected in solution, an amine-functionalized analogue of anthraquinone, **14**, was synthesized (Chapter 2). Experiments with **14** showed that cleavage could indeed be detected, although it was present at relatively low levels. The apparent negative percent cleavages observed were a feature of

the first generation library compounds themselves and not of the photoactive moieties (Figure 3.5).

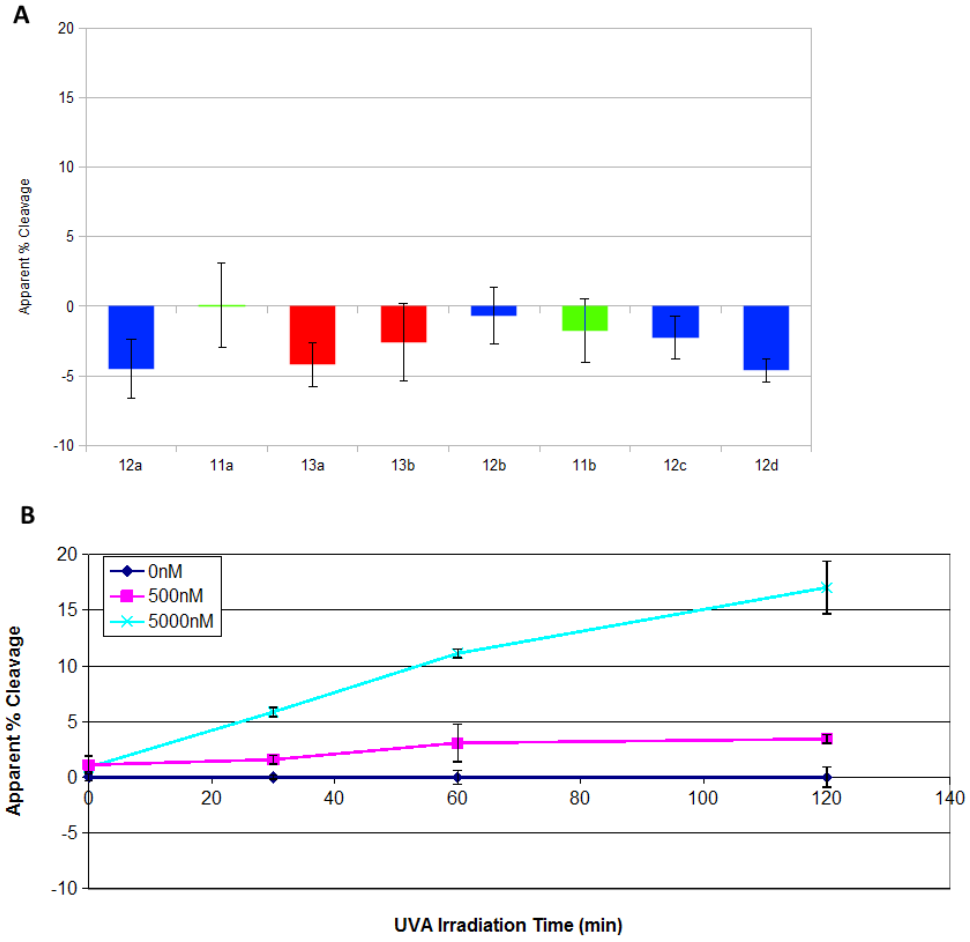


Figure 3.5: First generation solution photocleavage results. (A) Negative apparent percent cleavage of F21T observed for 500nM treatments of first generation compounds irradiated with UVA-centered lamps for 30min. (B) Cleavage of F21T by **14** at several concentrations over the course of 120min irradiation with UVA-centered lamps.

The second generation library compounds gave apparent percent cleavage values that were higher than the first generation compounds, although the levels of cleavage were still small when compared with TMPyP4. The second generation compounds showed a preference for F21T as a photocleavage substrate compared with cMYC

(Figure 3.6). This is the first case of such selectivity seen by our group. Prior to this, the cMYC sequence had always shown higher levels of cleavage when treated with photocleavage compounds. This selectivity is interesting, as it suggests that these compounds, to some extent, can discriminate between G4 structures with different conformations. This ability to discriminate is valuable given the high polymorphism that most G4-forming sequences exhibit. Similar to the first generation compounds, the second generation compounds' photocleavage had unexpected concentration dependence. This could be due to variations in active concentrations arising from self-association of the compounds.

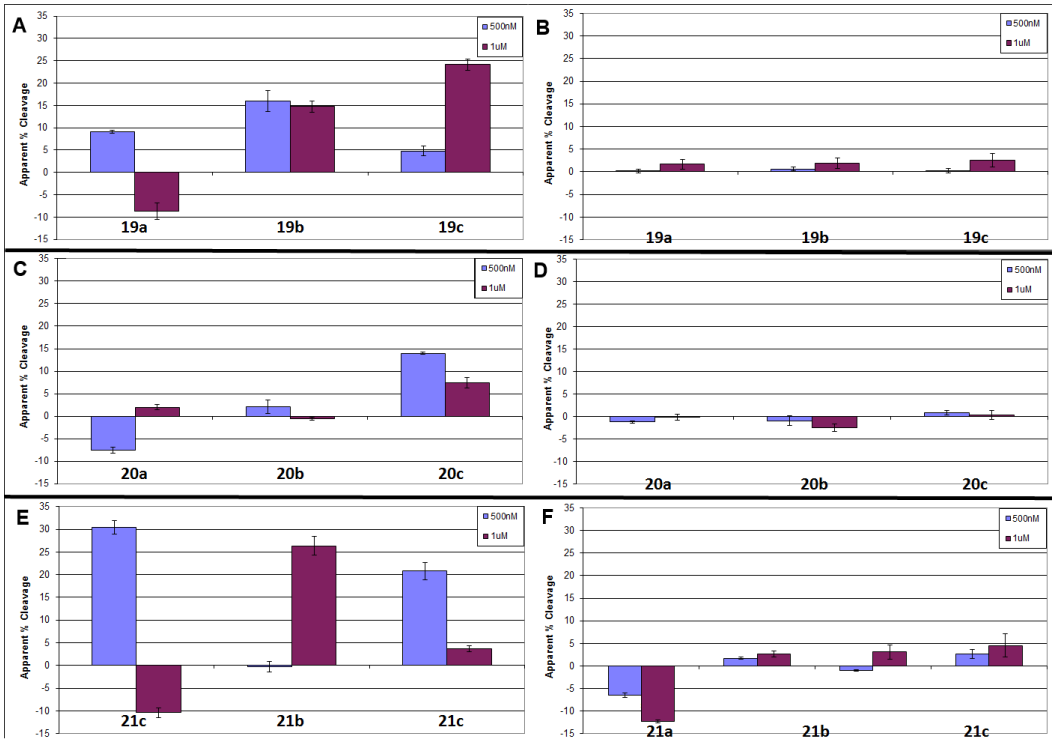


Figure 3.6: Apparent percent DNA cleavage solution data from second generation library. (A, C, E) Photocleavage of F21T by naphthalimide, benzophenone, and anthraquinone-functionalized compounds respectively. (B, D, F) Photocleavage of cMYC by naphthalimide, benzophenone, and anthraquinone-functionalized compounds respectively.

Second generation compounds with shorter linkers showed more association with the fluors, giving negative apparent percent cleavage readings similar to those observed with the first generation compounds. Overall, the second generation compounds suffered much less from this characteristic than did the first generation compounds and gave useful solution results.

As is discussed below, the second generation compounds show relatively little change in T_m when forming a complex with either F21T or cMYC. This suggests that there is little stabilization of the DNA structure through interaction with the compounds as a ligand. However, the second generation photocleavage levels were improved over those of the first generation library compounds, despite superior stabilization of G4 structures by some of the first generation library compounds. In order to better understand the interaction between cleavage and binding, a photocleavage-inactive compound, **23** (Chapter 2), was synthesized. Photocleavage experiments with this inactive compound present as a competitive photocleavage inhibitor showed concentration-dependent inhibition of photocleavage (Figure 3.7, Appendix C). This suggests that photocleavage of DNA by the second generation compounds is dependent on binding, but the small change in T_m suggests that the binding is likely transitory. Interestingly, increasing concentrations of **23** could lead to negative apparent percent cleavage values in solution, suggesting association with and quenching of the fluors either by **23** or by displaced second generation compounds (Appendix C). Another interesting observation from the gel data (Figure 3.7), is that **23** was better at inhibiting frank strand cleavage than it was at inhibiting base damage that required treatment with base to resolve into full strand cleavage.

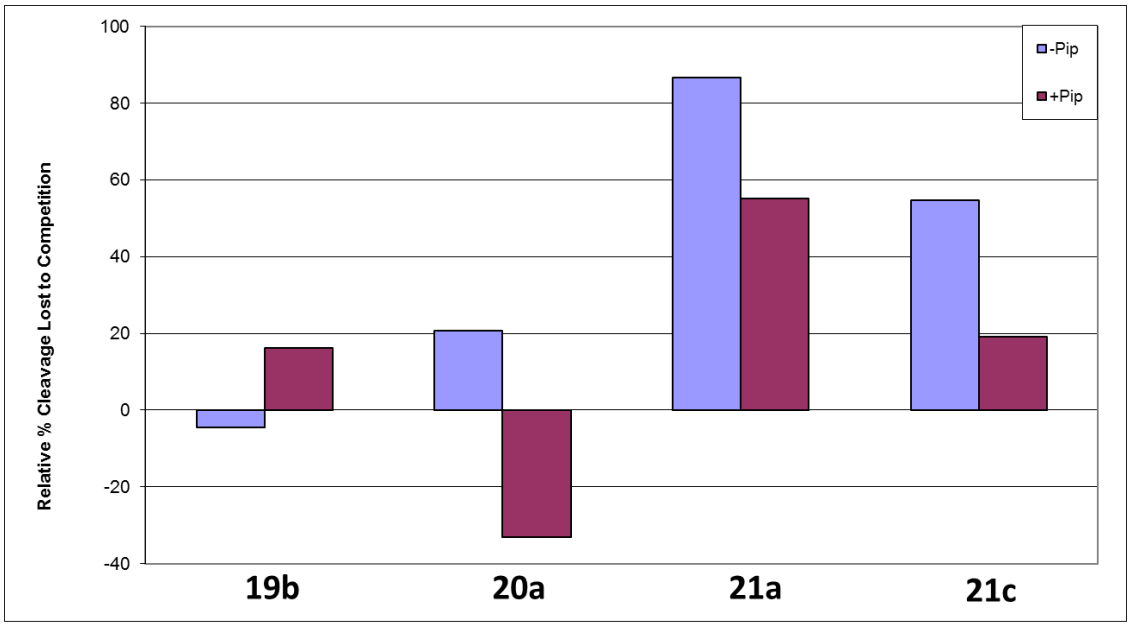


Figure 3.7: Inhibition of second generation compounds resulting from the presence of an inactive inhibitor as detected by PAGE. The displayed compounds were incubated with a 2-fold excess of photocleavage-inactive compound **23** and inhibition compared with an untreated control of each. Samples were split and half were treated with piperidine (+Pip) and heat while the other half was not treated with piperidine or heat (-Pip).

Photocleavage by TEL011 was followed using gel results because it too gave negative apparent percent cleavage values from the solution data. This is unsurprising given its strong absorbance that competes with that of the FRET pair used, creating a strong quenching effect. The photocleavage by TEL011 showed an interesting trend in concentration dependence. Only when a certain concentration threshold (4μM in the screen conditions) was reached would photocleavage occur (Figure 3.8). Further increases in concentration had only a small effect on increasing photocleavage. This concentration dependence coupled with the tendency of PDIs to self-aggregate (12, 13) strongly suggests that photocleavage was being achieved by an aggregation state of TEL011. As further verification of this, the study was repeated in buffer spiked with potassium chloride to increase the ionic strength (increasing the favorability of

aggregation). These conditions resulted in a lowering of the threshold concentration required for photocleavage as would be expected for aggregation-dependent cleavage (Figure 3.9).

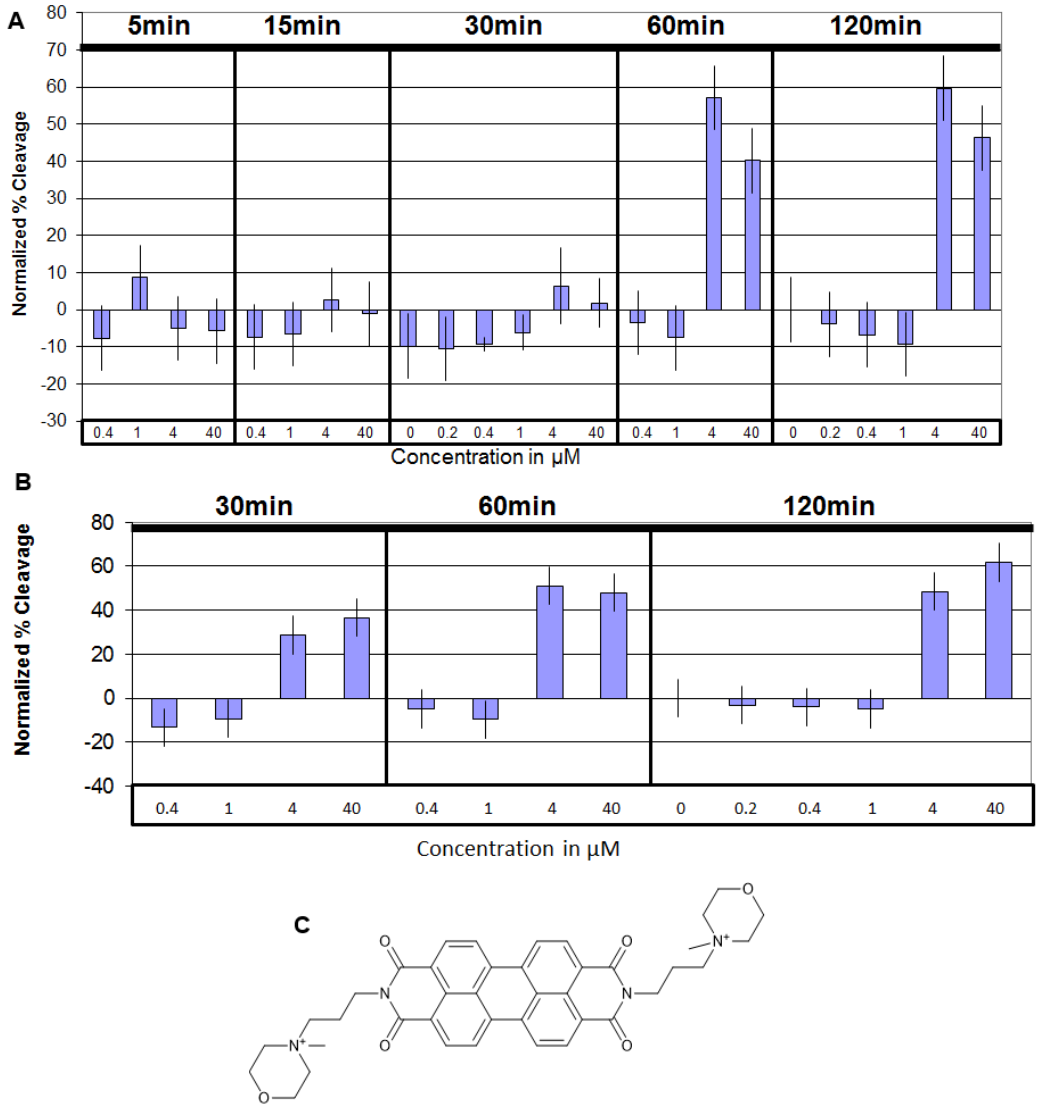


Figure 3.8 TEL011 photocleavage results. (A) Cleavage of cMYC by TEL011 after irradiation with white-light lamps. Photocleavage appeared to be strongly dependent on TEL011 concentration, requiring the attainment of a threshold level before photocleavage occurred. (B) Photocleavage of F21T by TEL011 after irradiation with white light lamps. A concentration threshold was required for photocleavage to occur. (C) Structure of TEL011

TEL011 also showed some preference for F21T as a photocleavage substrate over cMYC, although not to the extent that the second generation library did. cMYC required at least 60min of irradiation in order for cleavage to be detected whereas F21T showed cleavage after 30min of irradiation. At 60min irradiation and beyond, photocleavage of F21T and cMYC were comparable (Figure 3.8).

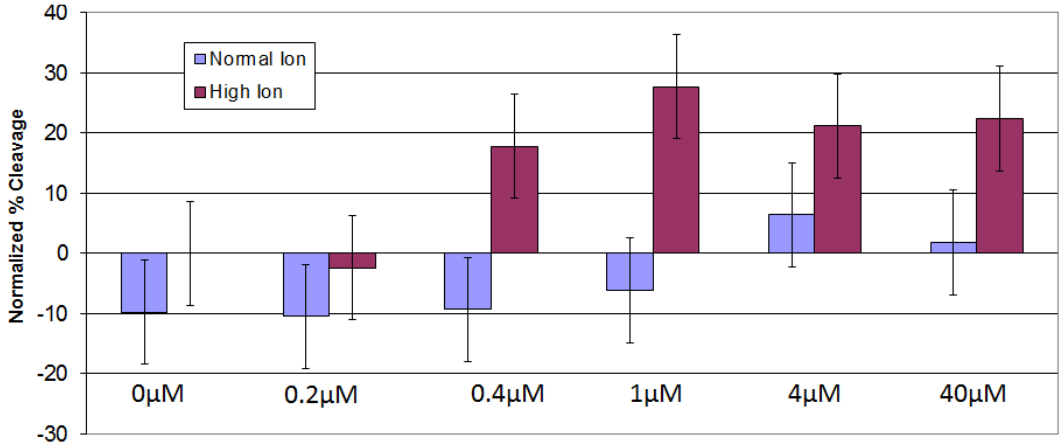


Figure 3.9: cMYC photocleavage by TEL011 in high ionic strength buffer. Conditions of high ionic strength lead to lowering of the TEL011 concentration threshold required for photocleavage to occur. Samples were irradiated for 30min with white light lamps

The conditions used for the NMM photocleavage screen (Figure 3.11) were chosen based on DNA complex melt results (Figure 3.14) in order to investigate if the photocleavage activity of NMM changed after induction of its preferred G4 conformation. The impact of sodium or potassium buffers on cleavage was investigated in cases where the DNA was either annealed before treatment (preannealed) with NMM or with DNA that was annealed in the presence of NMM (co-annealed). It was also verified that lamps centered at 420nm gave increased photocleavage relative to white light lamps.

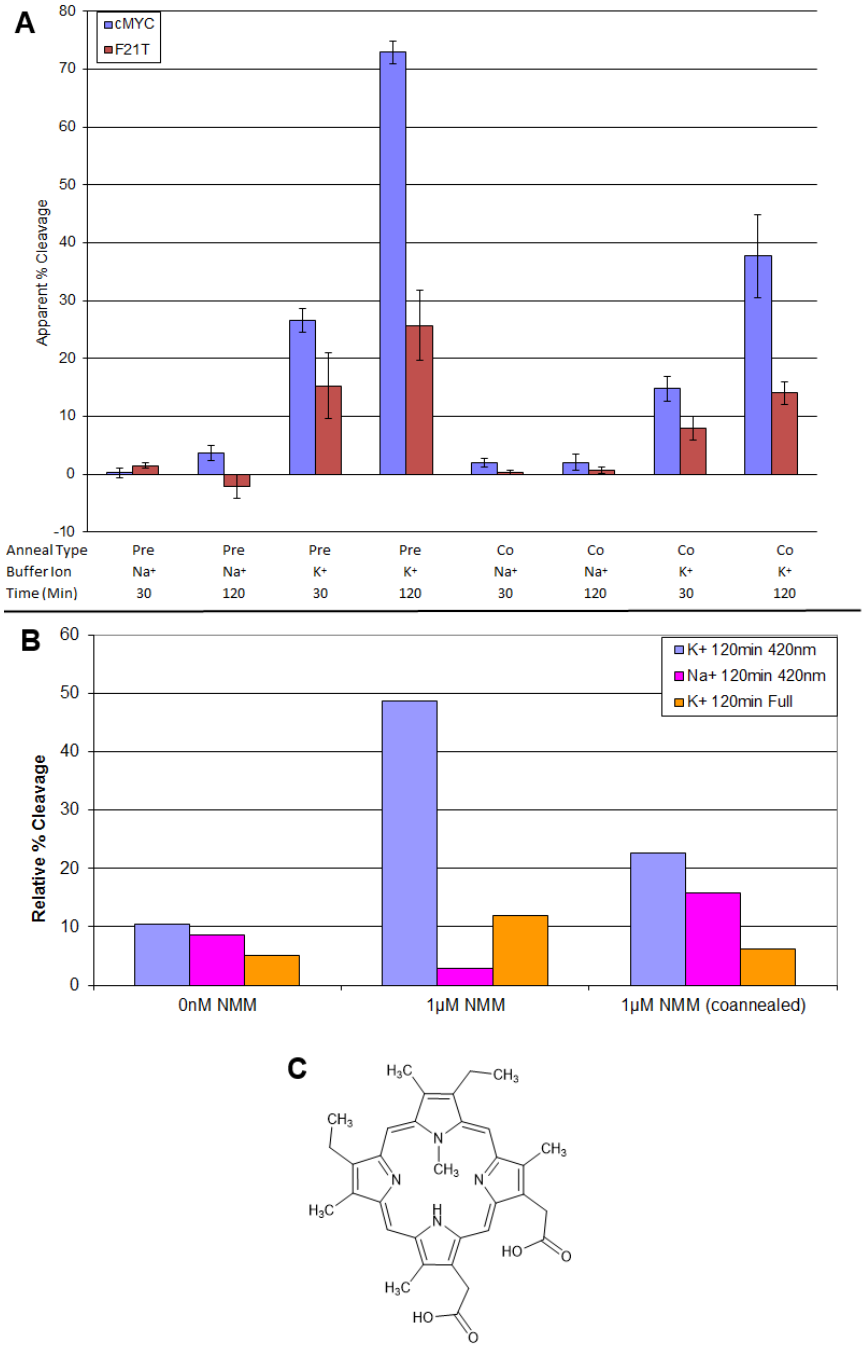


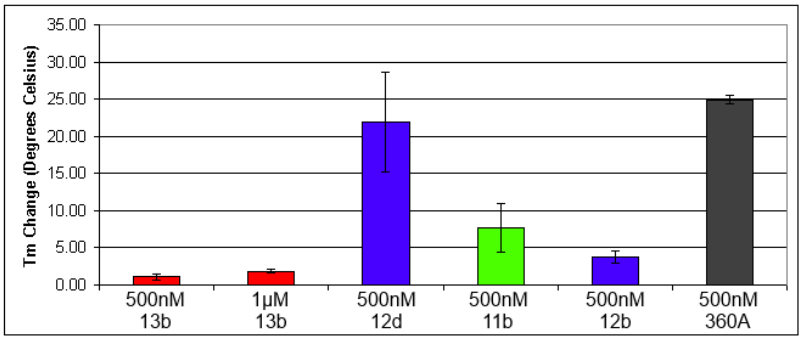
Figure 3.10: Photocleavage of dual-labeled oligonucleotides by NMM. (A) Apparent percent cleavage from solution data under different anneal and buffer conditions with irradiation at 420nm for 30min or 120min (B) Percent cleavage of cMYC determined from PAGE analysis with differing anneal (pre vs. co), buffer (K^+ vs. Na^+), and irradiation wavelength (420nm vs. "full" visible light) conditions. (C) Chemical structure of NMM.

Interestingly, NMM's photocleavage preferences are inverted from its binding preference (as indicated by T_m increases for the DNA-NMM complex). But, in all cases, photocleavage was much higher in potassium buffers than in sodium buffers. The former suggests that NMM likely has different binding modes with different G4 conformations and that these modes impact its ability to photocleave the DNA. The latter suggests that the G4 conformation most conducive to photocleavage is likely favored by potassium rather than sodium and likely involves improved binding and or placement of NMM for photocleavage. Like most of the photocleavage compounds studied by our group, NMM is more proficient at cleaving cMYC than F21T under comparable conditions.

DNA-LIGAND MELTING ASSAYS

DNA melts of F21T and cMYC conducted with first generation compounds showed a linker-length dependent increase in T_m. Those compounds with short linkers showed very small increases in T_m. However, compounds with the longest linker nearly "rescued" stabilization of the DNA-ligand complex as compared to the stabilization seen when F21T or cMYC were treated with 360A. The first generation compounds also increased the T_m of F21T much more than they did cMYC, suggesting selective binding to the conformation adopted by F21T over that adopted by cMYC.

A



B

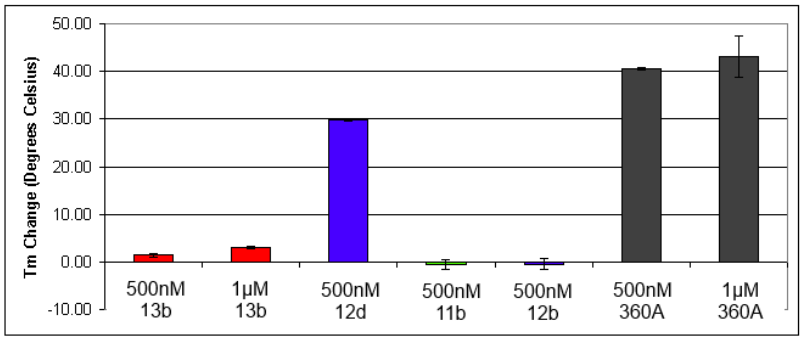


Figure 3.11: Changes in T_m upon formation of the DNA-compound complex. (A) Average melt data for representative compounds incubated with cMYC, (B) Average melt data for representative compounds incubated with F21T, Red, blue, green, and gray bars represent anthraquinone-incorporated, benzophenone-incorporated, naphthalimide-incorporated, and positive control compounds respectively.

This data suggests that the positioning of the triazolium group (Chapter 2) and/or the photocleavage moiety can strongly influence the ability of the compound to bind to and stabilize the G4 structure. Longer linkers would give greater opportunities for the aromatic groups of the photocleavage moieties to base stack with the tetrads at either end of the G-quadruplex. Likewise, longer linkers would ease association of the positive charge of the triazolium group with the negatively charged DNA backbone, allowing more optimal arrangement of the targeting moiety with the DNA. The results from the second generation compounds suggest that the latter effect was likely dominant.

The second generation library, which lacked the positively charged triazolium group, showed decreased binding to both F21T and cMYC when compared with the first generation compounds. A second generation compound analogue, **23** (mentioned above), which lacked the linker and photocleavage moieties, was able to significantly increase the T_m for F21T, but not for cMYC. Although it did not significantly increase the T_m for cMYC, it still increased the T_m more than any of the second generation compounds (Figure 3.13). Longer linker lengths did not have the same effect for the second generation library as they did for the first generation library. This lack of increased Tm for the second generation library compounds with longer linkers is likely due to the absence of the positive charge on the triazol group present in the first generation library

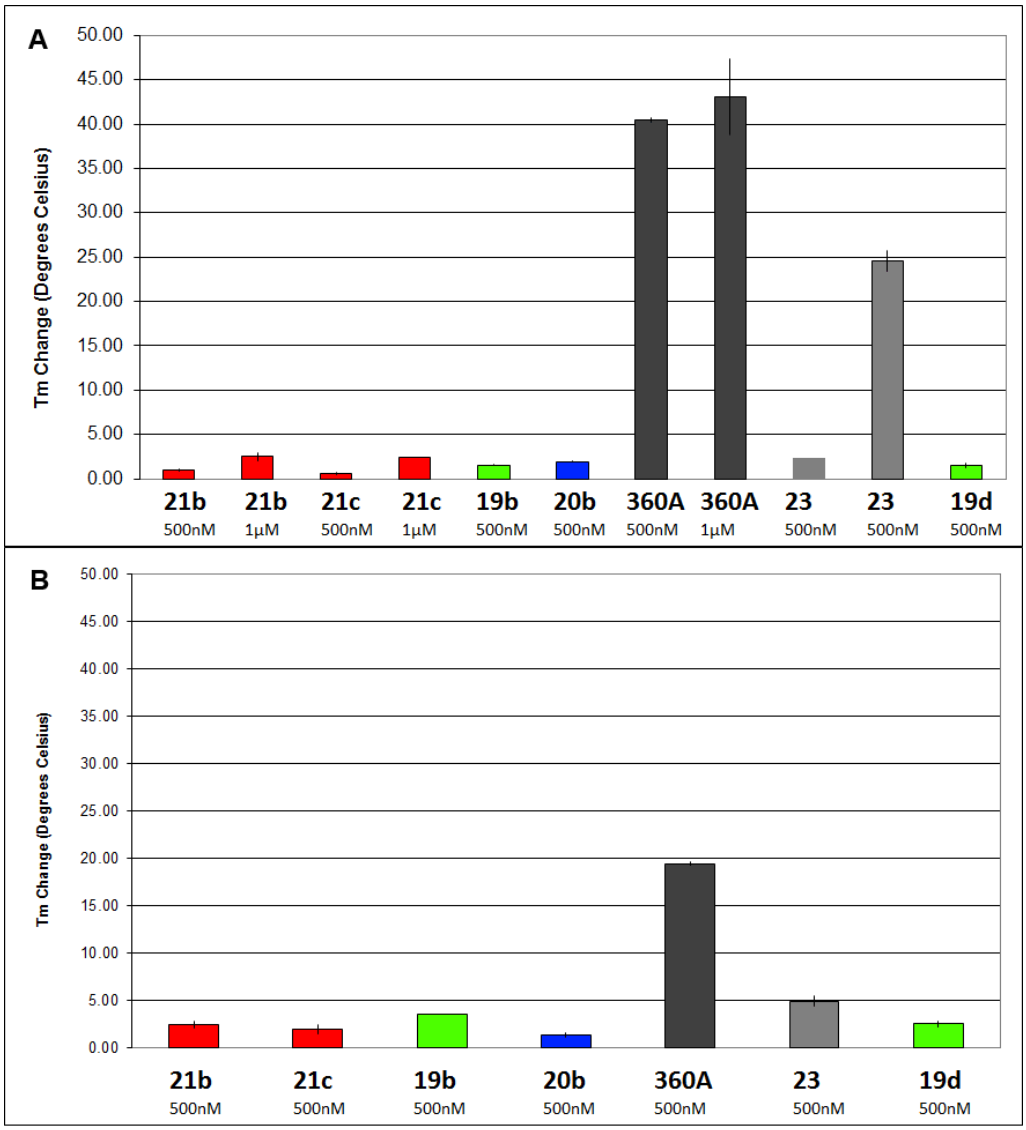


Figure 3.12: DNA Melts with Representatives of Second Generation Library. (A) Melts with F21T - increases in T_m were generally below 5 degrees, except for the positive control 360A and the unfunctionalized control **23**. **23** showed a biphasic melting curve with two melting temperatures, (B) Melts with cMYC - all compounds tested showed increases in T_m below 5°C except for the positive control.

That **23** was able to significantly increase the T_m of F21T while none of the second generation compounds could, suggests that either the linker and/or the photocleavage moieties are responsible for the apparent loss of binding. However, as discussed above, photocleavage activity was shown to be dependent on binding (because

competition experiments with **23** showed concentration dependent inhibition of photocleavage). Thus binding is occurring, but it is likely transitory. In comparison, the first generation compounds had several members that showed significant T_m increases, but all of the first generation compounds showed low photocleavage activity. The triazolium group of the first generation compounds likely is responsible for the apparent improvement in binding, and may also be responsible for the low photocleavage activity. It is possible that the orientation that is adopted limits the extent to which the photocleavage moiety can interact with the quadruplex. To address this, it would be necessary to add an additional linker region on the other side of the triazolium group in the case of the first generation compounds. The second generation's binding stability could be improved by incorporation of a positive charge (perhaps through incorporation of an amine-arm linker strategy such as that discussed in Chapter 2) midway between the binding moiety and the photocleavage moiety. However, a transient binder with high photocleavage activity and high selectivity may serve as a better probe than a more stable binder. It would allow biological processes to be followed with minimal impact on the process by the probe itself since it would be less likely to act as a barrier to binding by proteins such as helicases or polymerases.

TEL011 showed small increases in T_m, when mixed with F21T solutions (approximately 10°C at 1μM), that were only slightly higher than those seen in the second generation compounds, suggesting that it may be a transient binder as well. However, this interpretation is complicated by the photocleavage results discussed above. Because TEL011's photocleavage activity appears to be dependent on an aggregated form, it is possible that at much higher concentrations of TEL011, when an aggregated form is present, that a nonlinear spike in T_m increase could be observed for F21T. The concentrations required however, might also lead to precipitation, further complicating

analysis. TEL011 also appeared to inhibit the ability of F21T to reform the G-quadruplex structure after melting. When cMYC was treated with TEL011, the melting curve was not logistic, preventing the determination of a T_m value. This effect may be related to the photocleavage preference of TEL011 for F21T over cMYC

NMM showed an interesting pattern in T_m changes (Figure 3.14). When F21T was first treated with NMM, it showed a depressed T_m, suggesting that NMM destabilized the G4 structure assumed by F21T. However, if F21T was melted a second time, after reannealing in the presence of NMM, it showed an increase in T_m relative to the oligonucleotide on its own. This suggests that NMM is able to induce F21T from its original conformation to a different conformation that is more favorable for binding by NMM. To verify this, the experiment was redone in sodium buffer, which gives F21T a different preliminary conformation. Under these conditions, NMM increased the T_m immediately without need of co-annealing, and co-annealing now showed no effect. In the case of cMYC, which had the adjusted sequence which forces it into a single conformation, NMM is initially destabilizing in both sodium and potassium buffers. This ability of NMM to induce a different G4 conformation is in agreement with what others have reported (8). However, even after induction of a new conformation, the T_m increase resulting from addition from NMM is very low for both F21T and cMYC.

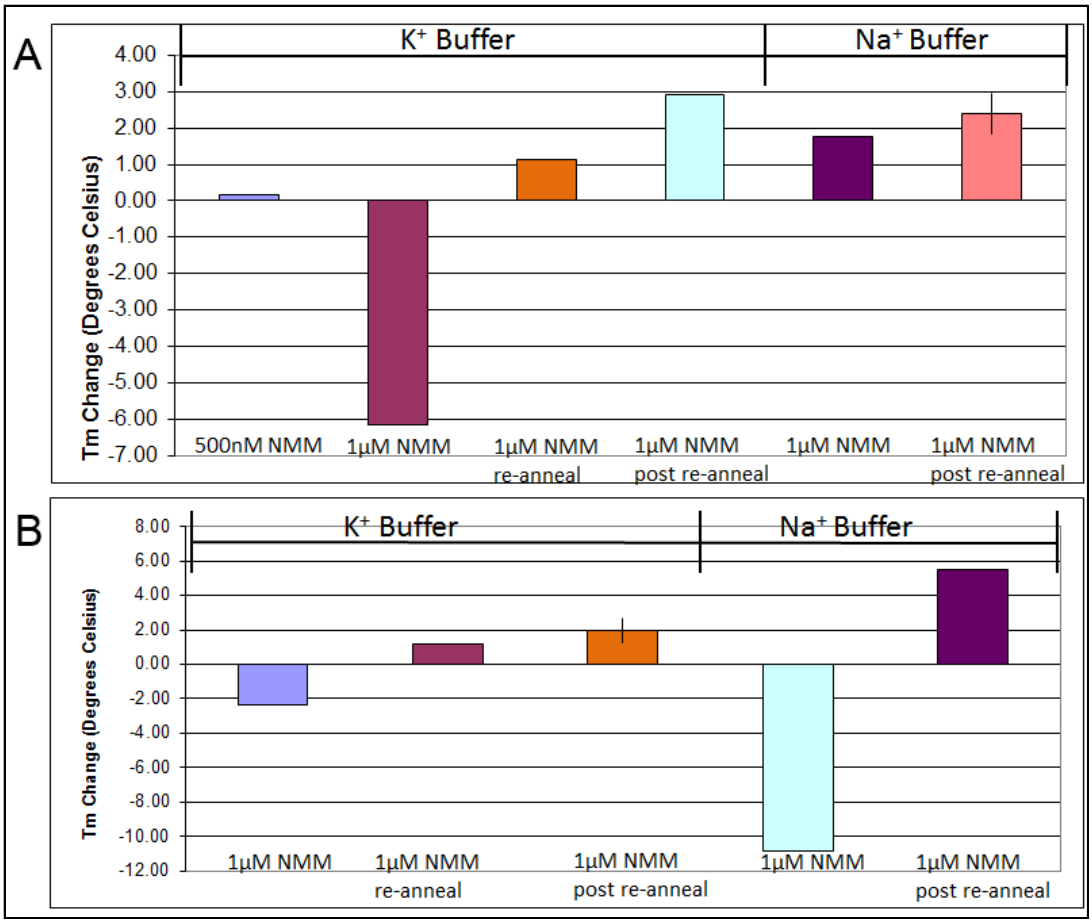


Figure 3.13: DNA melts with NMM. A) F21T incubated with NMM in potassium or sodium buffer B) cMYC incubated with NMM in potassium or sodium buffer

CANDIDATE SELECTION

The low photocleavage activity of the first generation compounds precluded selection of any of them for further investigation. Although a few compounds from the first library (such as **13a**) occasionally approached the photocleavage activity of the second generation compounds, this level of cleavage was less reproducible than it was for the second generation compounds. The second generation library also had relatively low photocleavage activity, but did show an overall improvement compared to the first generation library compounds. In the second generation library, the compound with the

highest and most consistent photocleavage activity, **21c**, was selected for further investigation. TEL011 and NMM were also selected for further studies because both showed very promising levels of photocleavage for at least one of the oligonucleotides utilized. These selections also cover a range of structural families with different optimal excitation wavelengths.

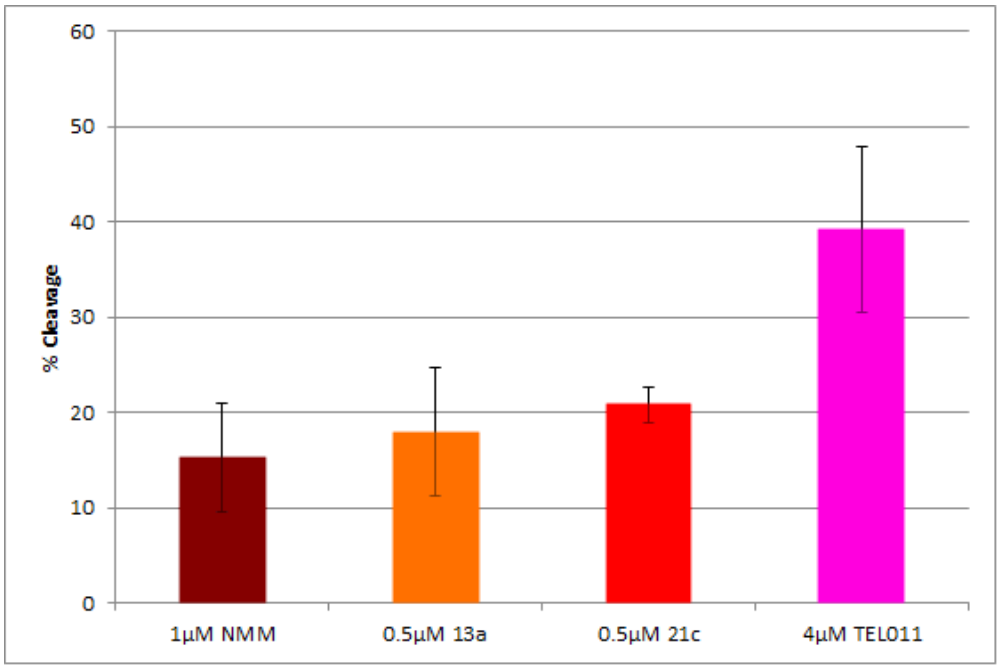


Figure 3.14: Comparison of photocleavage of F21T by top compounds. NMM (420nm lamps), a top cleaver in the first (**13a**) and second generation libraries (**21c**) (both using UVA lamps), and TEL011 (white light) each irradiated for 30min.

Out of the compounds selected, only TEL011 showed a significant increase in T_m when forming a complex with F21T or cMYC. However, as mentioned above, if the small increase in T_m is indicative of transient binding, this may indicate that the other compounds would be more advantageous choices for a G4 probe. The selected compounds also represent a spectrum of different binding mechanisms. The

photocleavage data suggests that TEL011 is most active in an aggregated form. NMM shows strong G4 structure induction effects and **21c** may be exhibiting transient binding.

The photocleavage screens gave some insights into the structure activity relationship involved in both binding and photocleavage for the 360A analogues and also allowed comparison with representatives from other structural families. The choice of **21c**, NMM, and TEL011 as probe candidates allows for this comparison to be extended into the experiments that are discussed in chapter 4.

CHAPTER 3 REFERENCES

1. Kern,J.T. and Kerwin,S.M. (2002) The aggregation and G-quadruplex DNA selectivity of charged 3,4,9,10-perylenetetracarboxylic acid diimides. *Bioorg. Med. Chem. Lett.*, **12**, 3395–3398.

2. Tuesuwan,B., Kern,J.T., Thomas,P.W., Rodriguez,M., Li,J., David,W.M. and Kerwin,S.M. (2008) Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents. *Biochemistry*, **47**, 1896–1909.

3. Schoonover,M. and Kerwin,S.M. G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*, 10.1016/j.bmc.2012.10.017.

4. Ren,J.S. and Chaires,J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.

5. Kerwin,S. (2000) G-quadruplex DNA as a target for drug design. *Curr. Pharm. Des.*, **6**, 441–471.

6. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.

7. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.

8. Nicoludis,J.M., Miller,S.T., Jeffrey,P.D., Barrett,S.P., Rablen,P.R., Lawton,T.J. and Yatsunyk,L.A. (2012) Optimized end-stacking provides specificity of N-methyl

mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.,* **134**, 20446–20456.

9. González,V. and Hurley,L.H. (2010) The c-MYC NHE III1: Function and Regulation. *Annu. Rev. Pharmacol. Toxicol.*, **50**, 111–129.

10. Ambrus,A., Chen,D., Dai,J.X., Jones,R.A. and Yang,D.Z. (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.

11. Schoonover,M. and Kerwin,S.M. (2011) A Fluorescence-Based G-Quadruplex DNA Cleavage Assay. In Sheardy,R.D., Winkle,S.A. (eds), *Frontiers in Nucleic Acids*. Amer Chemical Soc, Washington, Vol. 1082, pp. 13–32.

12. Kern,J.T., Thomas,P.W. and Kerwin,S.M. (2002) The Relationship between Ligand Aggregation and G-Quadruplex DNA Selectivity in a Series of 3,4,9,10-Perylenetetracarboxylic Acid Diimides. *Biochemistry*, **41**, 11379–11389.

13. Aoki,K., Ishikawa,Y., Oyama,M., Tomisugi,Y. and Uno,T. (2003) Self-Aggregation Inhibits the Photonuclease Activity of Porphyrins. *Chem. Pharm. Bull. (Tokyo)*, **51**, 899–903.

Chapter 4: Probe Candidate Assessment and Characterization

Three compounds were chosen as probe candidates based on the results of the screening experiments. These compounds each represent a different compound family. N-methyl mesoporphyrin IX (NMM) (1–5) is a porphyrin, TEL011 (6–8) is a perylene diimide (PDI), and **21c** is a 360A analogue from the second generation library (discussed in Chapters 2 and 3). Assessment and characterization of viability as a G4 probe involved using selectivity tests for both photocleavage and binding.

BACKGROUND

Assessing photocleavage selectivity was achieved through several plasmid-based photocleavage experiments. First, DNA nicking assays were conducted on a "generic" duplex plasmid, ϕX174 RF-1. Nicking assays are quite sensitive. Any cleavage will lead to nicking with concomittant relaxation of supercoiling in the plasmid, giving a new, lower-mobility band. By comparing the relative intensities of the supercoiled and nicked bands and subtracting out any background nicking present in an untreated control, it is possible to determine the relative extent of nicking due to compound-induced photocleavage. In these first experiments, low nicking is indicative of low photocleavage activity for duplex DNA. Lower duplex photocleavage activity is desirable for a G4 specific photocleavage probe.

The use of a plasmid containing the insert 5'-dAGCTCAATGGGGTTGGGGTGGGGTTGGGGCGC-3' capable of forming a G4 structure (pSP189G4) allows photocleavage selectivity of G4 structures in the context of surrounding duplex DNA to be evaluated. A nicking assay very similar to that mentioned

above was utilized, except that in this case, the relative amount of nicking is compared with the empty vector, pSP189 (9), which lacks the G4-forming insert. Here, increased nicking of the plasmid with the G4 insert relative to the empty vector would suggest targeting of the insert for photocleavage.

In addition to the photocleavage selectivity experiments, binding selectivity was also considered. This involved experiments very similar to the melt assays discussed in Chapter 3, except that an unlabled DNA competitor is added. By monitoring the changes in the Tm change due to treatment with the compounds at different competitor DNA concentrations, an assessment of the relative selectivity of the compounds for the G4 structure over the structure adopted by the competitor DNA sequence can be determined. For competition melt experiments discussed in this chapter, the competitiors were: $T_2G_{20}T_2$ (an intermolecular G4-forming sequence), calf thymus DNA (duplex DNA that was processed (see appendix A) to give duplex strands with a distribution centered at 200bp), and dG_1dC_2 (a triplex DNA-forming sequence).

PLASMID NICKING ASSAYS

21c showed no signs of duplex plasmid (ϕX174 RF-1) nicking after 2 hours of irradiation under UVA-centered lamps. It actually appeared to confer some protection from background cleavage (as seen from a concentration-dependent increase in negative photocleavage). However, at higher concentrations of **21c**, precipitated DNA-complexes were observed (represented by staining in the wells of the agarose gel). This suggests that while **21c** appears to have low photocleavage activity on duplex DNA, it does appear to interact with it and it may suffer from concentration limitations due to its lower solubility.

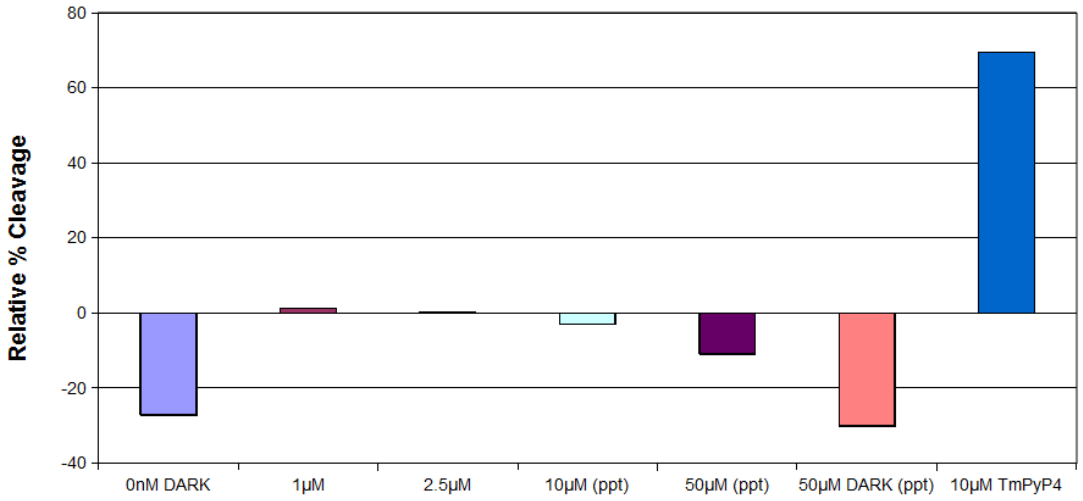


Figure 4.1: Relative nicking of ϕX174 RF-1 by 21c. A range of 21c concentrations were tested All were irradiated with UVA light for 2hrs.

TEL011 showed very little cleavage of ϕX174 RF-1. Only at the highest concentration tested was a significant level of nicking observed even with two hours irradiation under white light. This is much less than the positive controls that used TMPyP4 (10, 11). TEL011 is at least 50-200X less active than TMPyP4. However, similar to **21c**, TEL011-treated samples suffered from precipitation at the higher concentrations. A control experiment where the samples were irradiated only for 15 minutes showed no nicking at any concentration of TEL011 although TMPyP4 photocleavage was still readily observable.

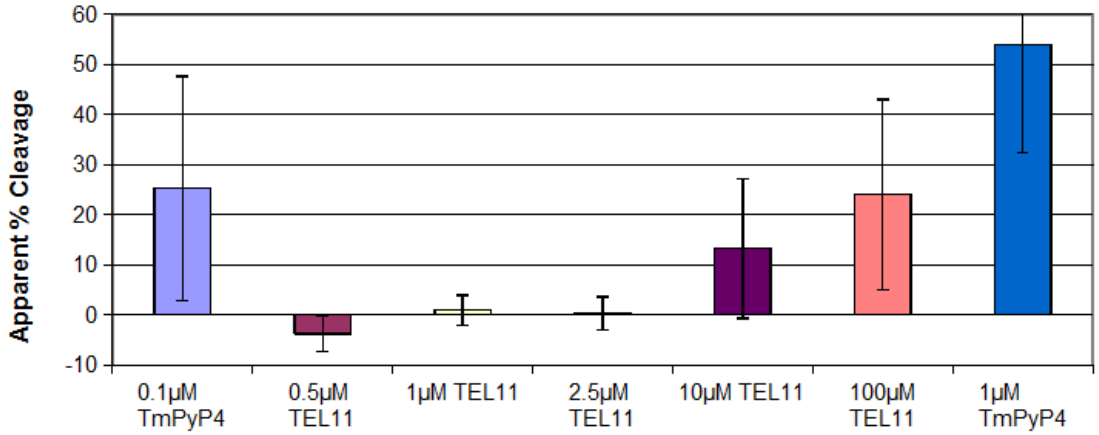


Figure 4.2: Relative nicking of ϕX174 RF-1 by TEL011. A range of TEL011 concentrations of were investigated. Each was irradiated under white light lamps for 2hrs.

Similar to TEL011, NMM showed negligable cleavage of ϕX174 RF-1 under all experimental conditions except for the highest concentration treatment (irradiating with UVA-centered lamps for two hours), where the level of nicking was similar to that of the 1μM TMPyP4 positive control, suggesting it is at least 50-400X less active than is TMPyP4 at photocleaving duplex DNA. Unlike **21c** and TEL011, NMM showed little evidence of precipitation, which would allow its use over a much wider concentration range than the other two compounds investigated.

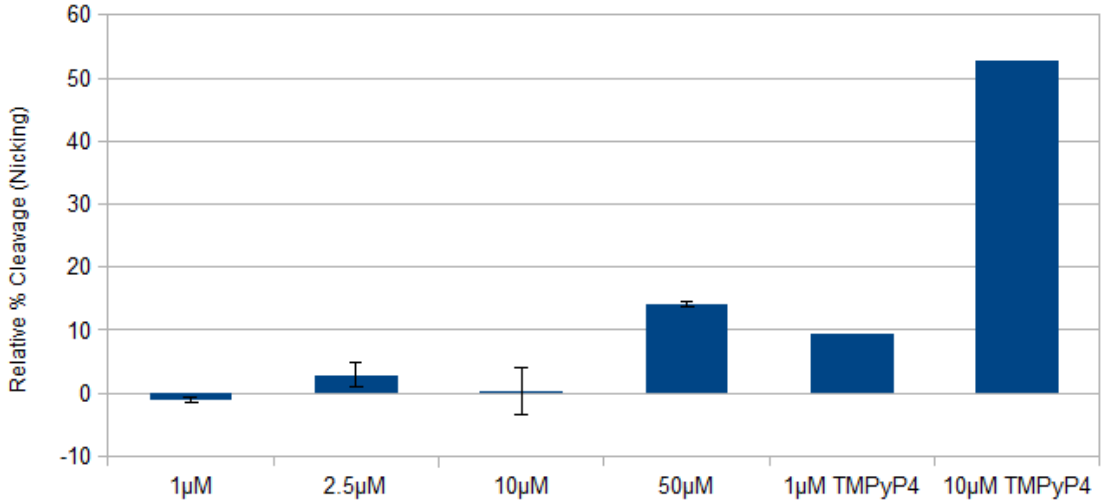


Figure 4.3: Nicking of ϕX174 RF-1 by NMM. Several NMM concentrations were investigated. Each was irradiated with 420nm-centered lamps for 2hrs.

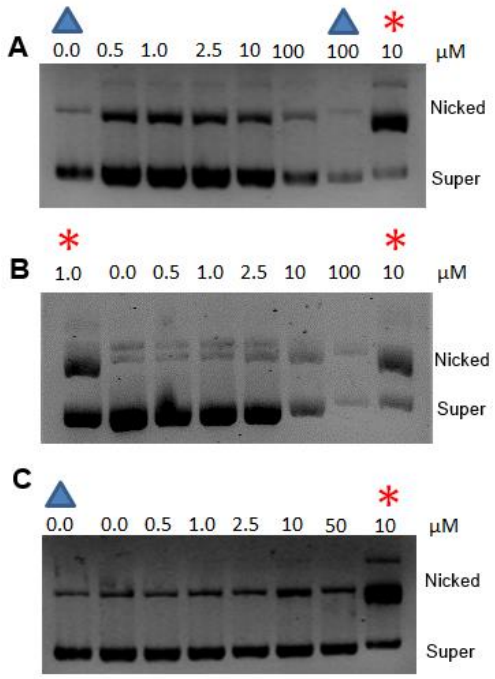


Figure 4.4: Example ϕX174 RF-1 Nicking Gels. Agarose Gels (1%) in TBE buffer were run with (A) **21b** irradiated with UVA for 2 hours (B) TEL011 irradiated with white light for 2 hours, and (C) NMM irradiated with 420nm for 2 hours. Blue triangles indicate nonirradiated controls. Red Asterisks indicate TMPyP4 controls which were irradiated with 420nm for 2 hours.

PLASMID NICKING ASSAYS WITH G4 INSERT

Plasmid nicking assays sites were conducted on pSP189G4 plasmid treated with TEL011, NMM, or **21c** and compared with pSP189 plasmid that had been similarly treated (Figure 4.5).

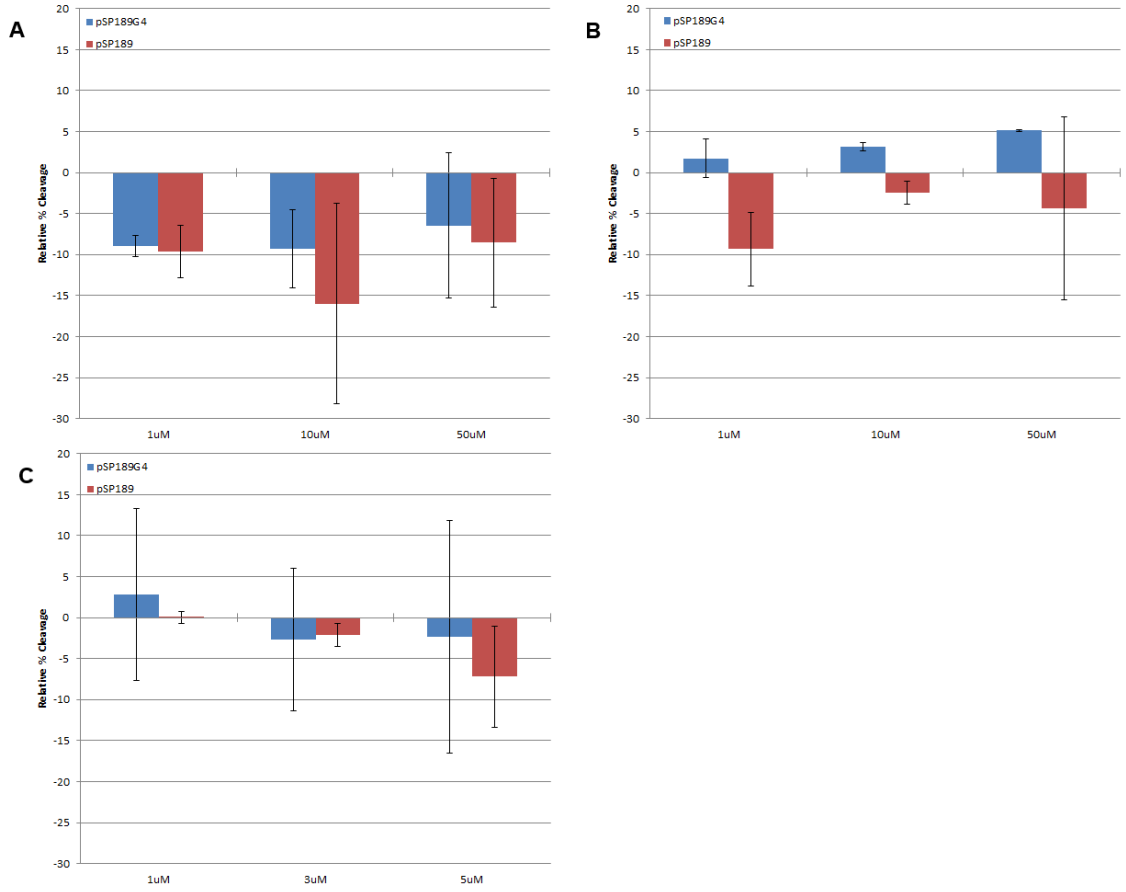


Figure 4.5: pSP189 nicking assay results. pSP189G4 and pSP189 were treated with (A) **21c** and irradiated with UVA for 2 hours, (B) NMM and irradiated at 420nm for 2 hours, and (C) TEL011 and irradiated with full spectrum light for 2 hours. Cleavage is normalized to pSP189G4 and pSP189 treated irradiated without treatment with compound.

Of the three compounds, NMM showed the highest level of G4-specific cleavage (Figure 4.5B). Treatment with 50µM led to cleavage of approximately 5% after 2 hours. In contrast, pSP189 showed negative relative cleavage when treated with NMM. Negative relative cleavage suggests that treatment with the compound led to *less* cleavage

than was seen with the plasmid irradiated without treatment with compound. This suggests that NMM binds to duplex DNA but does not efficiently cleave it. Rather, it appears to protect the DNA from damage due to irradiation.

Treatment with **21c** also suggests protection from the irradiation source. This effect is somewhat more prounounced than seen with NMM, but this is not surprising given the higher energy wavelengths (UVA) used for irradiation. In the case of **21c**, the level of G4-specific photocleavage is ambiguous. On average, **21c** appears to protect the DNA slightly less from damage in the case of pSP189G4 than for pSP189. While this could be attributed to photocleavage of pSP189G4 by **21c**, this photocleavage is below the background levels of cleavage present due to irradiation alone.

TEL011 was surprisingly active in cleavage of both pSP189G4 and pSP189. Since it showed very little photocleavage activity when incubated with ϕX174 RF-1, this cleavage may be partially due to differences in the intrinsic features of pSP189 and ϕX174 RF-1. ϕX174 RF-1 is a phage plasmid and some phages have a high prevalence of modified DNA bases (12–14), while these base modifications do not appear to be as pronounced in ϕX174 RF-1 as in other phages, base modifications could be one explanation for a difference in photocleavage activity. Treatment with 10μM or greater TEL011 completely removed the band corresponding to supercoiled plasmid. As such, lower concentrations of TEL011 were used for these nicking assays than were used for the other two compounds. At these lower concentrations, nicking was associated with high error for pSP189G4. In general, it appears that TEL011 may have nicked pSP189G4 more than pSP189, but did not do so unambiguously. This contrast with the cleavage at the higher concentrations may be related to the aggregation-dependent cleavage discussed in Chapter 3.

Control experiments with TMPyP4 showed near complete destruction of both plasmids when treated with 5μM and irradiating 2 hours with 420nm light. These results were expected and demonstrate that the compounds retain low duplex photocleavage activity even in the new plasmids.

The errors on all of these nicking assays are fairly large. A major contributor to this is that in general, the levels of nicking were relatively low and there is fairly high background from irradiation-induced nicking. Distinguishing between pSP189G4 cleavage and pSP189 can be difficult, because in the supercoiled plasmid, only a relatively small fraction (approximately 10-20%) are expected to form G4 structures. Another possible source is the presence of potential G4 structures. Both pSP189 and ϕX174 RF-1 have several sequences capable of forming G4 structures with a tract size of two nucleotides. G4 structures from such small tract sizes are less stable and presumably less likely to form. The presence of additional G4 structures would complicate the interpretation as pSP189G4, even though it contains a sequence capable of forming a much more stable G4 (with a tract size of four nucleotides). The presence or absence of folded G4 structures could also potentially explain the differences in photocleavage activity of the compounds when acting on pSP189 and ϕX174 RF-1.

COMPETITION MELT ASSAYS

Because a T_m could not be determined for the cMYC-TEL011 complex (see Chapter 3), only F21T was used for the competition melt experiments. When $T_2G_{20}T_2$ was used as the competitor, only a 2X excess was required to half the delt T_m assodiated with the F21T-TEL011 complex while a 5X excess all but abolishes the stabilization by TEL011. This suggests that TEL011 is not particularly selective for the largely anti-parallel conformation of F21T over the parallel-stranded G4 structure of $T_2G_{20}T_2$.

Interestingly, at 5X excess and larger amounts of $T_2G_{20}T_2$, the ability of F21T to reanneal was recovered, however, now the delta Tm value was even larger than with F21T-TEL011 complex alone. This suggests that the competitor strands are likely interacting the the F21T strand, perhaps forming a different G4 structure altogether upon reannealing. This makes interpretation of such reanneal curves difficult. TEL011 is much more selective for the F21T G4 structure than it is for duplex DNA. A 100X excess of duplex bp was required to approach the level of compeition observed with 2X excess of a G4 competitor. This binding (as interpreted by T_m changes) behavior supports the photocleavage selectivity observed in the plasmid nicking assays. Competition melts with triplex DNA showed levels of compeition more similar to those of a G4 structure than duplex, with 10X triplex not quite reaching the level of compeition of 5X G4.

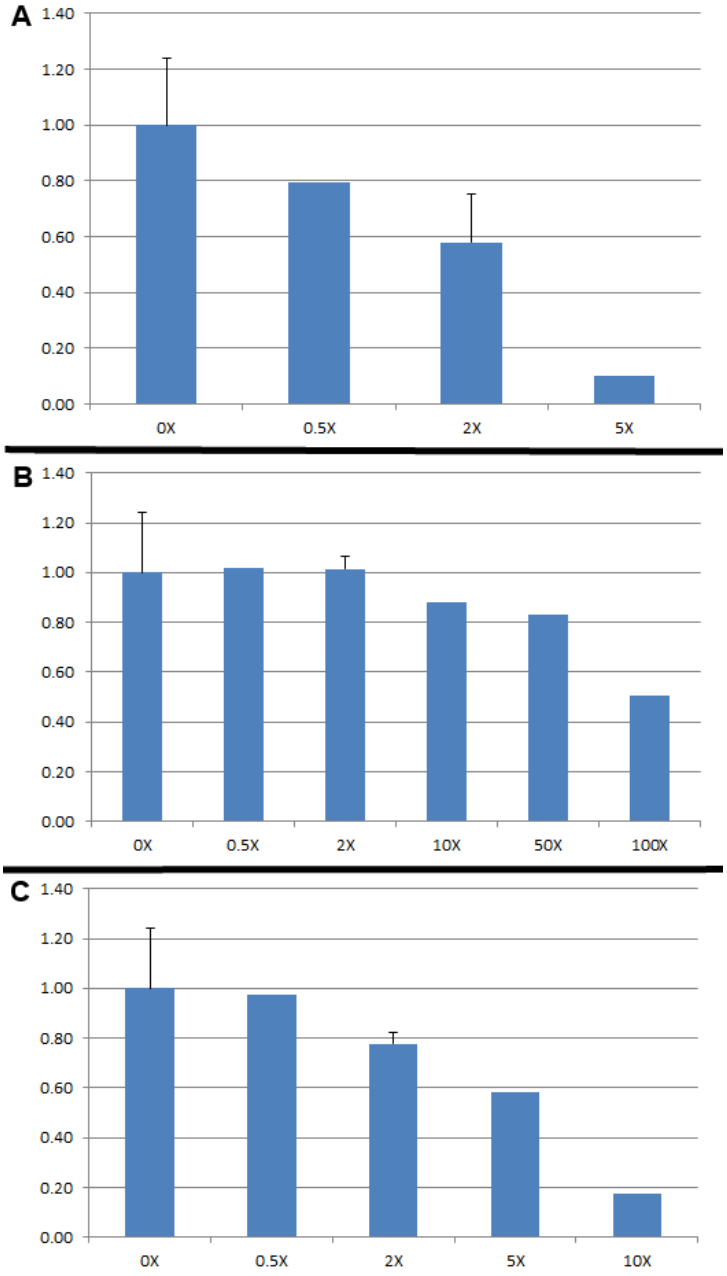


Figure 4.6: Normalized TEL011 Melt Competition Data for F21T. (A) $T_2G_{20}T_2$ (B) duplex DNA (C) poly dG_1dC_2. Competitor concentrations are relative to 600nM tetrad F21T and are in tetrad, bp, or triplets as appropriate. Data is normalized to treatment with 1µM TEL011 without competitor present.

It was not possible to conduct NMM competition melts because of the limited increase in melting temperature upon formation of the DNA-NMM complex. The T_m

could not be increased by increasing the NMM concentration, suggesting a high specificity of NMM for the G4 structure since this binding-linked signal reached saturation at comparitively low concentrations of ligand. This specificity may be related to the ability of NMM to induce different G4 conformations. These induction effects are discussed further in the CD experiments below.

In order for compeition melt experiments to be reliable, it is necessary that the baseline results obtained without competitior are reliable and produce a signal large enough that decreases can be measured. Because NMM is initially destabilizing, competition melt experiments were first attempted with co-annealed samples of NMM with either F21T or cMYC. As mentioned above, a sufficiently high T_m value could not be obtained for the complex to accurately detect its loss through the effects of competition. Because NMM is initially destabilizing, it was investigated if a "rescue" of the original melting temperature could be used as a way of evaluating the effects of competition. Unsurprisingly, this initial destabilization also appears to be insensitive to increased concentrations of NMM.

As with NMM, **21c** reached a threshold increase in T_m thta was too low to accurately follow the effects of competition. As mentioned above, this suggests that **21c** is reasonably specific for the F21T quadruplex.

CD EXPERIMENTS

To further investigate the ability of NMM to induce different G4 conformations, CD experiments were conducted (15–18). These experiments used an unlabeled 21-mer with the same sequence as F21T (see Appendix A for details), although one spectrum of F21T was acquired to ensure that they were indeed similar (which, except for a difference in peak intensities at 210nm, they were). The untreated 21-mer

had a major peak at 290nm, corresponding to an anti-parallel structure. After being co-annealed with NMM, however, a distinctive peak at 260nm formed while the peak at 290nm declined in intensity (although it did not disappear altogether). The peak at 260nm is associated with a parallel structure. Longer incubation with NMM failed to lead to any further changes in the spectrum. This suggests that the conformation adopted is a mixed structure, in disagreement with the parallel structure seen in the crystal structure (5), which could be accounted for in the conditions required to obtain the crystals. Alternatively, it is possible that at the concentration of NMM used, there is a mixed population of G4 structures, some with induced structures, and some still with their native structure. In either case, it is quite clear that NMM is capable of inducing G4 structures to adopt a parallel-stranded structure, which is more favorable for NMM binding.

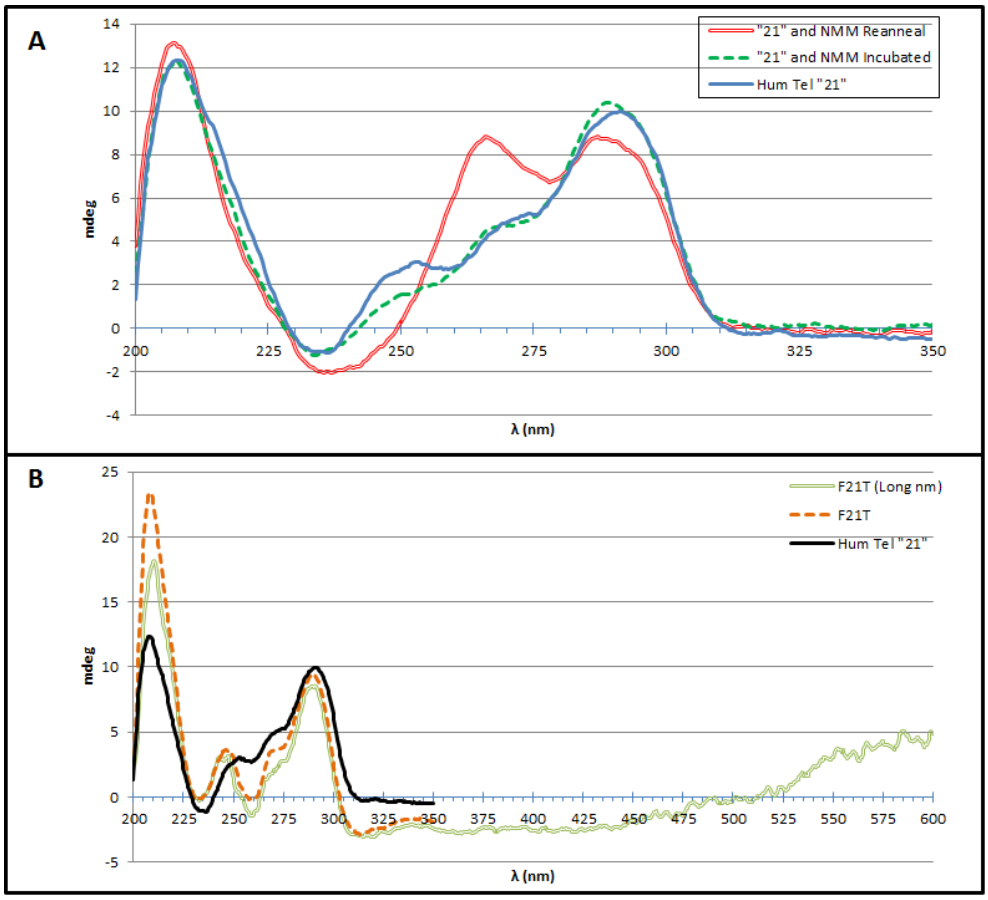


Figure 4.7: CD Data Obtained with 30µM human telomeric 21-mer or 28µM F21T. A) Overlays of 21-mer before treatment, after 24hr incubation, and after co-annealing with 5X excess NMM. B) Overlays of F21T with the 21-mer, including a long λ trace.

DMS EXPERIMENTS

The conditions used for the DMS protection allowed for multiple cleavage events to take place. Even under these high cleavage conditions, there was significant protection observed at the guanines expected to be involved in the formation of the G4 structure. Less expected was that there was also a lesser degree of protection protection seen for some of the adenines, suggesting the presence of protective interactions involving not just the guannines but to a lesser extent, some of the adenines. Under the conditions used, the

presence of NMM had little effect on the degree of methylation protection. This could be a further indication of relatively transient binding of the G4 structure by NMM.

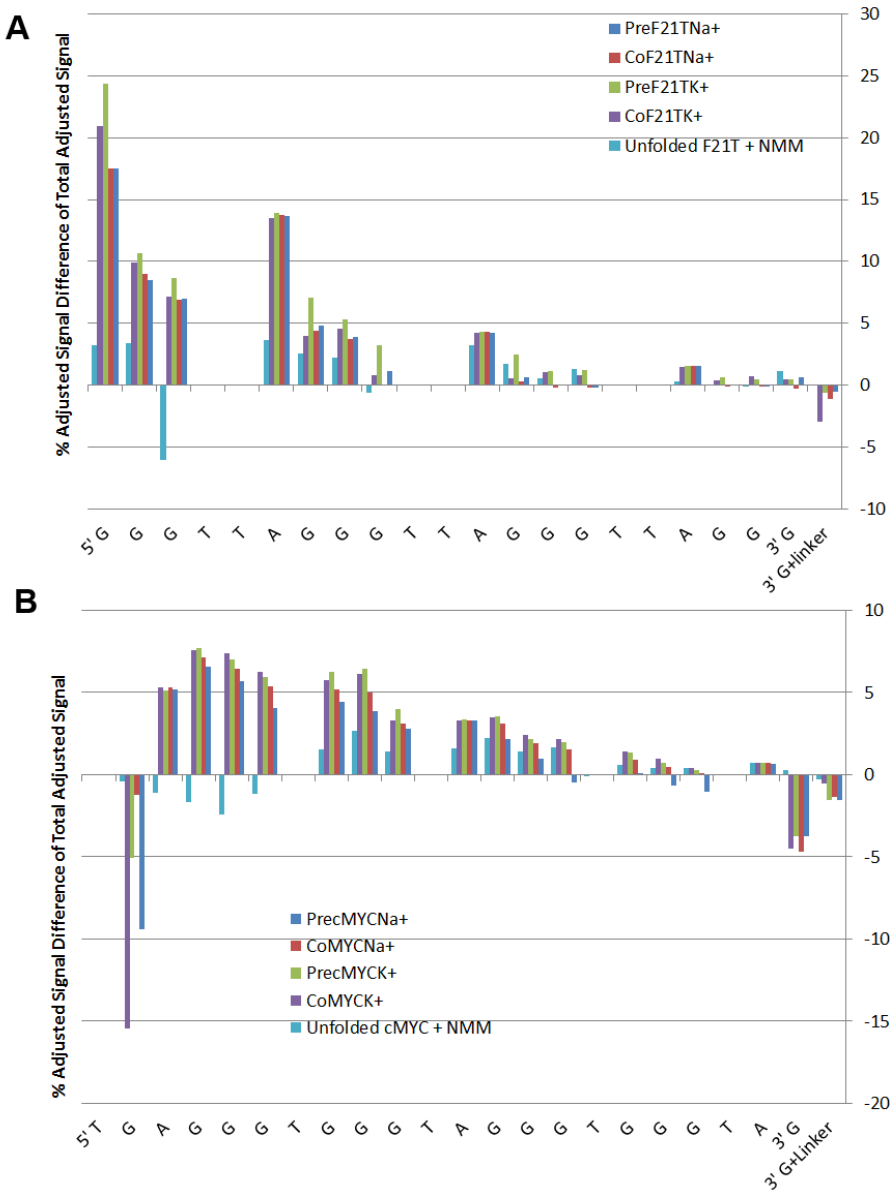


Figure 4.8: Relative DMS protection of F21T and cMYC. (A) F21T (B) cMYC. The apparent increasing loss of protection from the 5' FAM label to the 3' TAMRA label is indicative of "multi-hit" methylation/cleavage conditions. Protection is primarily observed at guanines with some adenines showing some protection as well.

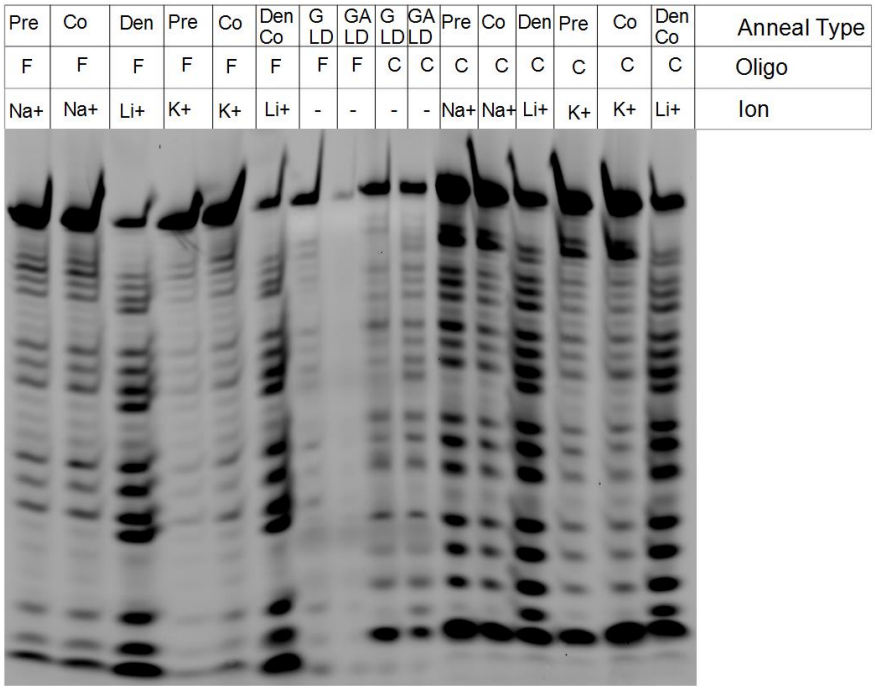


Figure 4.9: Example DMS PAGE results. Anneal types include preannealed (pre), coannealed (co), denatured (den), Denatured treated with NMM (den co) and G/GA ladders (G/GA LD). Oligos used were F21T (F) and cMYC (C).

CONCLUSIONS

The experiments discussed in this chapter extend those from Chapter 3, and have shown that all three of the compounds tested showed a significant preference for cleavage of G4 structures over duplex DNA. TEL011 was shown to also have binding selectivity for G4 structures over duplex DNA, and to a lesser extent triplex DNA structures. NMM was shown to have strong G4 induction effects, where it could force a change in G4 conformation. DMS experiments verified that a G4 structure was indeed being formed and also indicated that, under the conditions used, NMM does not protect the guanines of the G4 structure from methylation by DMS.

Although the photocleavage selectivity is promising, the activity could still use improvement. The shorter the irradiation time needed to accomplish photocleavage, the

greater resolution that the probes will have when studying biologically relevant processes such as helicase unwinding or G4-dependent replication/transcription. Ideally, probes would be able to resolve G4 structures on the scale of milliseconds to seconds rather than minutes to hours. None of the leading probe candidates approach this level of photocleavage activity currently. The time resolution minimum for all three is around 15-30min. This limits the level of detail that these compounds can be used to detect when following a biological process. This time could potentially be lowered by increasing the concentration of the compound, but may be limited by solubility, and are unlikely to reach millisecond to second time scales.

In terms of photocleavage efficiency, **21b** appears to be the most active (when cleaving F21T), accomplishing the same level of cleavage as NMM at half the concentration. The aggregation requirement of TEL011 makes it the least efficient cleavage agent due to the need to for higher concentrations to achieve effective levels of cleavage. Although **21b** appears to be the most efficient, there were indications of limited solubility in the nicking assays. Solubility issues could limit the use of TEL011 and **21b** as probes if attempting to increase the time resolution. NMM showed a comparatively high solubility and may be more conducive to increasing photocleavage through an increase in concentration.

The next logical step in assessing these compounds is to map the photocleavage sites to see if they are indeed selectively cleaving at the site or sites of G4 structures. Initially, pSP189 and pSP189G4 were considered for this. However, as discussed above, pSP189 does contain endogenous sequences potentially capable of adopting G4 structures. The formation of these G4 structures greatly complicates analysis, especially if more sensitive methods are used to make the determination. Preliminary experiments with pSP189 and pSP189G4 suggest that this complication may indeed be an issue with

these plasmids. Thus, before proceeding, an alternative plasmid needs to be considered which eliminates or at least minimizes the potential for signals arising from other areas of the plasmid than at the G4 sequence insert.

Once a plasmid that more rigorously controls for G4 formation is obtained, there are several methods that can be applied in order to map the site of cleavage. An alkaline agarose gel could be used to identify if the photocleaved plasmid generates fragments of the predicted size. A more sensitive method is a PCR "run-off" assay where a single primer (designed to use the strand containing the predicted G4 structure as template) is used and the nicking site identified by after amplification of the region between the priming site and the nicking site using conventional or alkaline agarose electrophoresis.

CHAPTER 4 REFERENCES

1. Ren,J.S. and Chaires,J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.

2. Kerwin,S. (2000) G-quadruplex DNA as a target for drug design. *Curr. Pharm. Des.*, **6**, 441–471.

3. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.

4. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.

5. Nicoludis,J.M., Miller,S.T., Jeffrey,P.D., Barrett,S.P., Rablen,P.R., Lawton,T.J. and Yatsunyk,L.A. (2012) Optimized End-Stacking Provides Specificity of N-Methyl Mesoporphyrin IX for Human Telomeric G-Quadruplex DNA. *J. Am. Chem. Soc.*, **134**, 20446–20456.

6. Kern,J.T. and Kerwin,S.M. (2002) The aggregation and G-quadruplex DNA selectivity of charged 3,4,9,10-perylenetetracarboxylic acid diimides. *Bioorg. Med. Chem. Lett.*, **12**, 3395–3398.

7. Tuesuwan,B., Kern,J.T., Thomas,P.W., Rodriguez,M., Li,J., David,W.M. and Kerwin,S.M. (2008) Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents. *Biochemistry*, **47**, 1896–1909.

8. Schoonover,M. and Kerwin,S.M. (2012) G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*, **20**, 6904–6918.

9. Parris,C.N. and Seidman,M.M. (1992) A signature element distinguishes sibling and independent mutations in a shuttle vector plasmid. *Gene*, **117**, 1–5.

10. Han,F., Wheelhouse,R. and Hurley,L. (1999) Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. *Jourrnal Am. Chem. Socieity*, **121**, 3561–3570.

11. Han,H., Langley,D.R., Rangan,A. and Hurley,L.H. (2001) Selective Interactions of Cationic Porphyrins with G-Quadruplex Structures. *J. Am. Chem. Soc.*, **123**, 8902–8913.

12. Razin,A., Sedat,J.W. and Sinsheimer,R.L. (1970) Structure of the DNA of bacteriophage φX174: VII. Methylation. *J. Mol. Biol.*, **53**, 251–259.

13. Warren,R.A.J. (1980) Modified Bases in Bacteriophage DNAs. *Annu. Rev. Microbiol.*, **34**, 137–158.

14. Gommers-Ampt,J.H. and Borst,P. (1995) Hypermodified bases in DNA. *FASEB J.*, **9**, 1034–1042.

15. Paramasivan,S., Rujan,I. and Bolton,P.H. (2007) Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods*, **43**, 324–331.

16. Kypr,J., Kejnovská,I., Renčiuk,D. and Vorlíčková,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.

17. Vorlíčková,M., Kejnovská,I., Sagi,J., Renčiuk,D., Bednářová,K., Motlová,J. and Kypr,J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64–75.

18. Randazzo,A., Spada,G.P. and da Silva,M.W. (2013) Circular Dichroism of Quadruplex Structures. *Quadruplex Nucleic Acids*, **330**, 67–86.

Chapter 5: G-Triplex Search Program

As discussed briefly in Chapter 1, G-triplexes (G3) are secondary structures that are very similar to G-quadruplexes (G4), and have been hypothesized to potentially be folding intermediates to G4 structures (1–3). While generally less stable than G4 structures, G3s show sufficient stability that it is possible such structures may form independently within living systems (4). Preliminary work done by our group has supported this by showing measurable melting temperatures for a single stranded nucleotide containing three G-tracts in several buffers. However, because these melting temperatures are significantly lower than are typically seen for G4 structures, the conditions in which G3s independently form are likely restricted to those that minimize the possibility of competing folds than are G4 structures.

G-Trip is a program I wrote in Python which can be used to aid in the analysis of genomes in order to investigate how likely G3 structures are to form based on the occurrence of potential G3-forming sequences. The program attempts to take the potential for the formation of competing G4 structures into account as a penalty score when reporting potential G3 hits. This chapter will discuss the rationale behind the approach that G-Trip takes and discuss the core features of the G-Trip program. Some preliminary analysis of search results from several genomes will also be discussed. The source code, a draft user manual/tutorial, and additional function documentation can be made available upon request.

BACKGROUND

Using programs to do a sequence-based search for locations where certain structures may form has been used several times to aid in the study of G4 structures (5, 6). G-Trip uses a similar sequence-based approach to search for G3 structures. In

particular, it uses regular expressions (REGEX) to rapidly search DNA sequences for matches to a general potential G3 template that can be modified by the user. As with G4 structures, the ability to analyze the distribution of potential G3 sites may allow determinations of whether or not a selective bias may be present (in the case of G4 structures, it was found that there were far fewer locations capable of forming them than would be predicted to be present purely by chance in mRNA sequences (6)). This can give some evidence for or against the likelihood of G3 structures serving some function that has been modulated by selection. The ability to identify potential sites in genomes also provides starting points for wet-lab based investigations.

Several structural terms are used when discussing G-Trip. The first of these is a G-tract (6). In G-Trip, a G-tract is defined as a region in a sequence containing a contiguous set of Gs of a given size (the definition of a G-tract's size can be changed by the user, see below). For instance, AGGGA would contain one G-tract that was three nucleotides and/or two mutually exclusive G-tracts that were two nucleotides in size (each utilizing the central G). Loops are non-G-tract sequences which join G-tracts to each other. With this notation, a G4 structure can be described as being composed of four alternating G-tracts and three alternating loops whereas a G3 structure is composed of three G-tracts and two loops. G-Trip also distinguishes between two types of G4 structures in relation to a potential G3 structural hit. An “internal” G4 is one that is capable of forming within the G3 hit itself. While rare, certain combinations and size definitions of G-tracts make such occurrences possible. The second form of G4 structure is the “contextual” G4. This is a G4 structure that can form using a G-tract located within the G3 hit and other G-tracts located in the surrounding sequence outside of the G3 hit itself. What determines the surrounding sequence is defined by the user as discussed below. Internal and contextual G4s each have their own associated penalty sub-score.

While G-Trip supports the use of less specific nucleotide notation in the loops, only clearly defined G-tracts are allowed. For instance, AGGNA would be considered to have a G-tract of size two even though N could potentially be a G.

G-Trip uses several modules from the third party package BioPython (7) as well as a module called regex (https://pypi.python.org/pypi/regex) which is in development to replace the current re module in the standard python 3 distribution (G-Trip requires python version 3.3 or later to run correctly). BioPython is used to read gene bank and FASTA files such that included annotations can be assigned to appropriate hits as discussed below. The regex module is used in the detection of potential G4s because it easily supports overlaps between hits, which allows for more exhaustive detection of potential G4 structures.

GENERAL PROGRAM FEATURES

There are two general functions that G-Trip performs. The first of these is a scan, which uses user-specified definitions for G-tract and loop lengths to generate a REGEX recognition pattern. The sequences within the indicated source file(s) are then scanned using this recognition pattern to generate a list of hits. As each sequence is completed, the hits from that sequence are then further processed, applying internal and external G4 scores. Then, a hit's scores are checked against cut-offs that can be set by the user, allowing removal of raw hits that score poorly during processing, effectively applying a user-determined quality filter for G3 hits. If indicated by the user, and if a given hit scores well enough, overlapping genome annotations (if present in the source file) will be associated with the hit. Hits that scored high enough in both the internal and contextual G4 sub-scores are then appended to a hit summary file. A hit summary file is generated for each source file for each scan direction indicated by the user. The program will

continue to scan until all sequences have been scanned from every direction indicated by the user.

The variables that a user can change in the scan mostly impact the recognition pattern. The user may set the minimum size for a G-tract (by definition it must be at least 2), the maximum size of the G-tract, the min loop size (by definition at least 1), the max loop size, and which scan directions are desired. There are four scan directions: Normal (positive strand 5'-->3'), Reverse (positive strand 3'-->5'), Complement (negative strand, 3'-->5'), and Reverse Complement (negative strand, 5'-->3'). For most users, the Normal and Reverse Complement scans are sufficient. The other two scan directions are included primarily because of a bias that is introduced when using REGEX to scan the sequences. REGEX uses a "lazy" search approach, where it matches the first sequence of characters that match its definition template. This bias weights results towards whichever end the scan begins at. This is because hits are not allowed to overlap (if they were, a very large number of hits with very low scores would be included in every search, much as is seen in bi-directional strand search), and as a result, there can be a "frame shift" of sorts introduced when a strand is scanned starting from the 5' end vs. the 3' end. Scanning from both directions allows a search to be more thorough, but then requires removal of the numerous duplicates that arise (increasing processing time substantially). Generally, I have found that the number of additional (non-duplicate) hits found from scanning both directions is small enough that there is little justification for scanning both directions. However, depending on how restrictive the search parameters are, the difference may become significant and there could then be benefits to being able to scan in both directions with removal of duplicates.

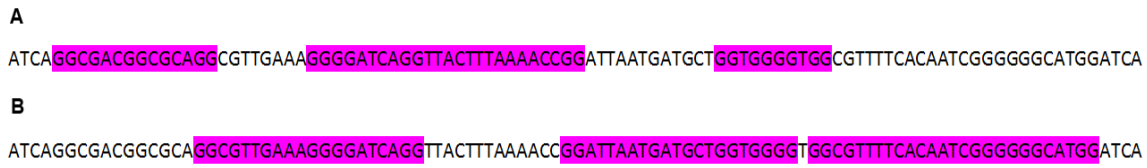


Figure 5.1: Example of G-Trip Detection. (A) Scan results of a short example sequence using a Normal (5'->3') scan. (B) Scan results of the same short example sequence using Reverse (3'->5') scan. Overlap detection leads the program to only accept hits one and two (from 5' end) from (A) and hit one (from 3' end) from (B) for compilation based on their internal and context scores. This example illustrates the "frame shift" effect that results from scans weighted towards the end they begin scanning from. Note that all hits shown in this example score very poorly in context and are likely to be rejected by most sub-score filter settings.

The second function that G-Trip performs is a combination of final scoring and compilation. This allows associated hits from multiple sources to be combined. For instance, it allows compilation of all the results of scanning the human genome using individual chromosome source files. During compilation, the hits are filtered for duplicates by searching for overlaps (if both search options for a given strand were used in the preliminary searches). Any overlapping sequences are considered to be duplicates (the highest scoring of two overlapping hits is kept while the other is discarded). The total score is then calculated. The total score incorporates scores for the internal and contextual scores based on the weighting set by the user. If the user is using total score filtering, the hits are then filtered again by comparing the total score to the cut-off value set by the user. Finally, the hits are assembled, and each unique hit sequence is optionally assigned a unique integer ID number (to aid in identification of potential G3 motifs). These compiled results are then saved to a final master hit summary file in a format intended to be easy to filter and/or sort in spreadsheet programs such as Microsoft Excel or Libre Office Calc.

The program can conduct scans on individual files or on a batch file created by the user. A batch file is a comma-separated-value (.csv) file of the following format for each line after the first line in the file: source_filename.extension**,**output_filename**,**filetype. The first line of the batch file is ignored by the program and can be used for category headers or other notes. The source file is placed in the main G-Trip directory and its name must include its extension in the batch file definition. The output file name does not use an extension because G-Trip will save all output files as a tab-delimited .csv file. Three file types are acceptable for source files (each corresponding to one of the three source file types supported by G-Trip). The first is "gb", which indicates that the source file is a gene bank file. The second is "fasta" which indicates that the source file is a FASTA formatted file. The last file type is "txt" indicating a plain text file. Plain text files should be formatted with each sequence occupying its own line in the file. This last file format does not support any annotation or other labeling.

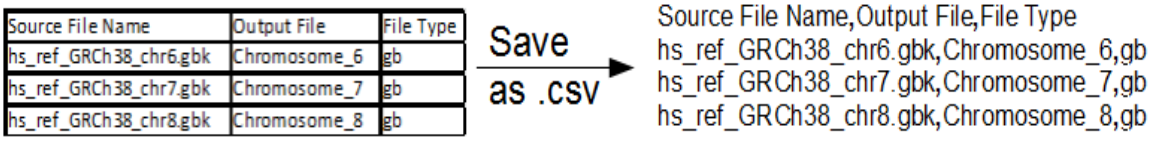


Figure 5.2: Example process for preparing a batch file. Batch files can be prepared in a spreadsheet program (left) and saved as a .csv file which can be read by G-Trip (right). Note that the first line of the batch file is reserved for headers/notes and is not processed by G-Trip.

Another less important feature that the program offers is the ability to generate a quick summary of hit results from multiple sources. The utility of this summary is limited, and this feature will likely be phased out in later versions of the program.

HIT SCORING IN DETAIL

As discussed above, G-Trip attempts to apply score penalties to G3 hits if there are potential G4 structures associated with those hits. Because research into G3 structures is still at an early stage, there is limited information available by which to develop these scores. As a result, they are arbitrary, but can be modified by the user. Because of the arbitrary nature of this scoring, it is vital that all results from G-Trip include the search and scoring parameters used to generate the results so that they can be judged in context. G-Trip automatically includes this information in the hit result and compilation files, but does not include this information for hit density analysis (see below).

The strategy that G-Trip uses to penalize G3 hits that contain or are near G4 structures is as follows. Before running a search, the user specifies a "slice" size. A slice in this case is the base sequence extension used for context score determination (in nucleotides). The choice of slice size is arbitrary, relying on the requirements of the user. The larger the slice size, the greater the restraint against contextual G4 structures. To determine the maximum extension to either side of the G3 hit, the slice size should be multiplied by 3. For illustrative purposes, this discussion uses a slice size of 18nt. When a hit is being scored, the hit and the surrounding sequence are analyzed. The sequence is extended by 3*slice size (54nt when using a slice size of 18) on either end of the hit sequence. The extended sequence is then scanned for potential G4 structures. This scan has two levels of sensitivity. The first is an exhaustive G4 search which will count up all possible ways to assemble a G4 structure within the extended sequence. The second, less sensitive detection method only counts possible G4 structures that constitute a unique sequence (however, these unique sequences can overlap with each other)

To detect G4 structures, the program first identifies all G-tracts in the sequence, starting first with G-tracts of size 2 and continuing up to G-tracts of size one less than that defined in the user's scan specifications (it does this using REGEX recognition with overlap permitted). It then combines G-tracts in sets of four to predict G4 structures. However, combinations are limited by a set of rules. A given G-tract can only be used once in a given G4 structure, all G-tracts must be separated by at least one nucleotide to be counted in the assembly, and G-tracts may only interact with G-tracts "down sequence" of themselves (this last is to prevent duplication during the iterative process). Finally, for a G4 prediction to be counted, it must include at least one G-tract that was contained by the original G3 hit (e.g. it must overlap with the original G3 hit). For G4 prediction, overlaps between different G4 structures are allowed. A loop limit can also be imposed by the user in the settings.

Once all G4s have been detected within the extended sequence they are binned based on the location within the extended sequence. If a G4 hit resides entirely within the original G3 hit, it is binned as an internal G4 structure. If it forms within the first slice extension (using a slice of 18, this would be within the G3 sequence including 18nt extensions to either side), it is considered a "small slice" hit. "Medium slice" hits involve extensions of 2*slice to either side of the original G3 hit sequence, and "Large slice" hits are any that remain. By doing this segmented binning procedure, it is possible to vary the penalty of a context score based on how close it is to the original hit (the premise being that G4 structures with smaller loops are more likely to form and/or be more stable than ones with very large loops (5, 6, 8)). The internal, small, medium, and large G4 penalty scores are then generated by using a base penalty and raising it to the number of G4 structures predicted (a separate base penalty can be set for internal G4 structures. Small, medium, and large G4 predictions share a base penalty but can vary in weighting). For

example, if 5 internal G4 hits were detected and the base penalty was 0.5, the final internal score would be 0.5^5 = 0.0125. The final internal score is retained and later used in calculating the total score. The small, medium, and large scores are combined into a single context score using weightings set by the user. For example, a user might weight small scores as 100, medium scores as 10 (10X less contribution than small scores) for medium scores, and 1 for large scores. The single context score is an aggregation of the small, medium, and large scores. It is used later in conjunction with the internal score to calculate the total score. Additionally, the user can assign different weightings for the internal and contextual scores.

There are some caveats with the current scoring method used. One is that by weighting the different categories separately, a minimum score is reached no matter how badly a hit may score in a given category since the contribution of that category would effectively be reduced to zero if it is the only consideration. This may also be an advantage if a certain category is desired to have a certain maximum impact. For example, if using slice sizes that are many nucleotides in length, the large slice G4s predicted would increase in number while at the same time becoming more unlikely to form. In this case, it would make sense to limit the contribution of the large slice G4s that are detected. This makes the choice of weighting an important one. However, more work is needed to determine the weightings that are optimal for screening for potential G-triplex forming sequences. The option to filter by sub-score was chosen as a means of partially addressing this issue as it gives a cutoff that is independent of the total score weighting.

5’CCCGGGTTCCCAAAGCAGAGGGCGTGGGGGAAAAGAAAAAAGATCCTCTCTCGCTAATCTCCGCCCACCGGCCCTTTATAATGCGAGGGTCTGG
ACGGCTGAGGACCCCCGAGCTGTGCTGCTCGCGGCCGCCACCGCCGGGCCCCGGCCGTCCCTGGCTCCCCTCCTGCCTCGAGAAGGGCAGGGCTT
CTCAGAGGCTTGGCGGGAAAAAGAACGGAGGGAGGGATCGCGCTGAGTATAAAAGCCGGTTTTCGGGGCTTTATCTAACTCGCTGTAGTAATTCC
AGCGAGAGGCAGAGGGAGCGAGCGGGCGGCCGGCTAGGGTGGAAGAGCCGGGCGAGCAGAGCTGCGCTGCGGGCGTCCTGGGAAGGGAGAT
CCGGAGCGAATAGGGGGCTTCGCCTCTGGCCCAGACCTCCCGCTGATGCCCCAGCCAGCGGTCCGCAACCCTTGCCGCATCCACGAAACTTTGCCC
ATAGCAGCGGGCGGGCTCTTTGCACTGGAACTTACAACACCCGAGTAAGGACGCGACTCTCCCGACGCGGGGAGGCTATTCTGCCCATTTGGGGA
CACTTCCCCGCCGCTGCCAGGACCCGCTTCTCTGAAAGGCTCTCCTTGCACCTGCTTAGACGCTGCATTTTTTTCGGGTAGTGGAAAACCAGGTAAG
CACCGAAGTCCACTTGCCTTTTTATTTTTTTTTTTTTCTCTTTCATGCTGAGATGAGTCGAATGCCTAAATAGGGTGTCTTTTCTCCCATTCCTGCACTA
TTGACACTTTTCTCAGAGTAGTTATGGTAACTGGGGCTGGGGTGGGGGGTAATCCAGAACTGGATCGGGGTAAAGTGACTTGTCAAGATGGGAGA
GGAGAAGGCAGAGGGAAAACGGGAATGGTTTTTAAGACTACCCTTTCGAGATTTGTGCCTTATGAATATATTCACACTGACTCCCGGCCGGTCGGA
CATTCCTGCTTTATTGTGTTAATTGTTTTCTGGGTTTTGGGGGGCTGGGGGTTGCTTTGCGGTGGGCAGAAAGCCCCTTGCATCCTGAGCTCCTTGG
AGTAGGGACCGCATATCGCCTGTGTGAGCCAGATCGCTCCGCAGCCGCTGACTTGTCCCCGTCTCCGGGAGGGCATTTAAATTTCGGCTCACCGCA
TTTCTGACAGCCGGAGACGGACACTGCGGCGCGTCCCGCCCGCCTGTCCCCGCGGCGATTCCAACCCGCCCTAATCCTTTTAAGAAGTTGGCATTTG
GCTTTTTAAAAAGCAATAATACAATTTAAAACCTGGGTCTCTAGA3’

Figure 5.3: Example of G-Trip Scoring. The sequence for the cMYC gene is shown. The underlined sequence indicates a triplex hit. Red, yellow, and green highlights indicate the large slice, medium slice, and small slice context extensions respectively. G-tracts within the limits of the large slice are bolded and are referred to by number from 5’ to 3’ below. There are 11 G-tracts of size 2nt, 3 G-tracts of size 3nt, and 1 G-tract of size 4nt. In this case, G4s can only be constructed from tracts of size 2 nt. Tracts 2, 3, 4, 5, and 6 are part of the original hit, however tracts 2 and 3 as well as 4 and 5 overlap with each other. Overlapping tracts may not both be included in a predicted G4. Since there are no G-tracts present in the medium slice extension, there are no medium G4s possible. There are many large G4s possible including 1,2,4,5; 1,2,5,6; 1,2,6,7; 1,2,6,8; 1,2,6,9; 1,3,4,5; etc. when exhaustive G4 detection is permitted. These G4s are numerous enough to effectively reduce the large slice score to 0. Because of tract 7 in the medium slice extension, there are four possible medium slice G4s: 2,4,6,7; 2,5,6,7; 3,4,6,7; and 3,5,6,7. There are no internal G4s possible. If 0.75 is used as the base context score, the temporary context scores become 1.0, 0.75^4 = 0.3164, and 0 for the small, medium, and large slices respectively. If the weighting is 10:5:1 for small:medium:large, the final context score becomes (10*1.0+5*0.3164+1*0)/(10+5+1) = 0.724. If the weighting is 5:1 for internal: context, the total score becomes (5*1.0+1*0.724)/(5+1) = 0.954

HIT ANALYSIS

Once the hits have been collected and scored. G-Trip allows the analysis of hit distributions within the sequences searched. This is done in several steps. The first step bins the hits into bins of a size defined by the user (in nucleotides). Binning in this way gives information about the densities of hits within the sequences scanned. In G-Trip, the user can define a total score cut-off so that only hits with sufficiently high scores are included in the binning process. The output file includes a list of the bins and the number of hits each contains. A summary of all sub-sequences and their lengths is also included at the bottom of the file.

Once the hit densities have been obtained, G-Trip determines peaks and troughs within the hit densities. This is accomplished in a second binning process based on runs of like bins. In G-Trip, a run is defined as a contiguous set of bins of the same type (either bins containing no hits, or bins containing greater than zero hits). The first type of run is composed of empty bins. From these G-Trip will determine troughs. From the second type, bins with greater than zero hits, G-Trip determines peaks. For example, if there are 10 bins with bin 1 containing zero hits, bins 2 through 3 containing at least one hit, and 4 through 10 containing zero hits, then there would be three runs detected, the first containing only bin 1 (potentially a trough), the second containing bins 2 and 3 (potentially a peak), and the third containing bins 4 through 10 (potentially a trough).

The average length and standard deviation for all runs of like types is calculated and is used to assign runs as being either peaks or troughs. In order for a run to be considered a trough or a peak, it must satisfy the following condition: length >= average_length+2*stdev, where length is the length of the considered run, average_length is the average length of the appropriate run category of the considered run, and stdev is the standard deviation associated with the appropriate run category. Satisfying this condition corresponds to a run being outside of the 95% confidence interval and thus a statistical outlier from the other runs of its same type. While all runs are exported, only those which pass the trough/peak test are labeled as peaks/troughs. The reason that empty bins and nonzero bins are treated separately is because in some cases empty bins occur disproportionately more often than do nonzero ones and would preclude the selection of any run containing hits if their lengths were factored into the same average and standard deviation calculation. Alternative criteria for defining troughs and peaks are still being considered, which might better address the issue.

After peaks and troughs have been detected, G-Trip has the option to annotate them with any gene features or notes that were supplied with the original source sequence used to generate them. This works similarly to the annotation option given for individual hits earlier, but provides more general information, because a given run generally represents a much larger portion of the source sequence than the hits. However, this annotation is comparatively faster, and can provide some preliminary assessment of the hit density context.

As part of the generation of the master summary file, G-Trip also calculates the proportion of sequence space that is occupied by the combined hits. It does this simply through summing the lengths of all of the hits and comparing this with the combined lengths of the source sequences searched. This value may prove useful in comparisons to statistical predictions of the amount of sequence space expected given the search pattern used for the scan that produced the list of hits.

PRELIMINARY SEARCH RESULTS

G-Trip has been used to scan several genomes. Data was collected using NCBI reference genome sequences for *H. sapiens*, *S. cerevisiae*, and several strains (B, C, K12, and W) and their substrains of *E. coli*. These scans were done using fairly permissive search criteria. G-tracts were defined as being sets of contiguous Gs between two and eight nucleotides in length while loops were permitted to be between one and fifteen nucleotides. Scans were done from 5'→3' on the positive and negative strands (using Normal and Reverse Complement search modes). The base slice increment was defined as 18 nucleotides. The base penalties for internal and context scores were set to 0.5 and 0.75 respectively. A sub-score cutoff of 0.001 and a total score cutoff of 0.5 was used. In cases where the data volume was very large (particularly for the case of *H. sapiens*),

only hits with a total score of 1.0 were used for hit density analyses. For detecting G4s, the exhaustive G4 detection option was used with no loop length limit. When binning to calculate hit densities, a bin size of 500nt was used.

The distributions of the scores obtained for *E. coli* (averaged between the substrains) are shown in Figure 5.4. The relative numbers of hits in each category is about the same for all strains tested. This is expected because all of the strains came from the same species. Strain B had slightly fewer G3 hits than the other strains. Otherwise, the occurrence of hits was similar amongst the strains investigated. The variability among the substrains within a given strain was also very low. The one exception to this was strain K12, which showed deviations in hit counts as great as 9%.

The hit densities (generated using a bin size of 500nt and using all hits with a total score >= 0.5) for all of the *E. coli* strains investigated indicate a relatively uniform distribution of hits across their genomes. This even distribution doesn't give any clear indication of the presence of selection for or against G3 structures in *E. coli*. When the binned data is further divided into runs so that peaks and troughs are detected, it is found that peaks outnumber troughs. Each run of peaks covers a large sequence space, suggesting that there are regions that are relatively G3-rich, but very few that are devoid of G3s. In these peak G3 density regions, the G3 density comes to approximately 4/bin or 1 out of every 125nt.

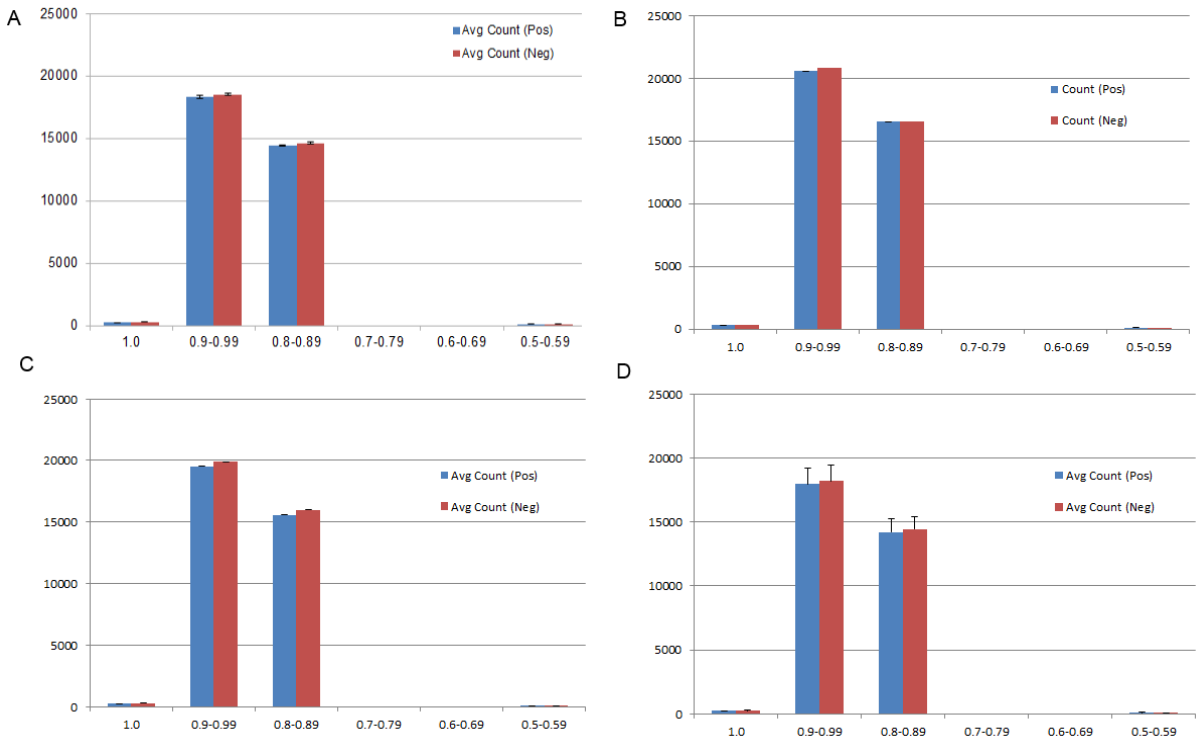


Figure 5.4: E. coli Score Distributions of Four Common Lab Strains. The y-axis denotes the number of hits found, while the x-axis shows several total score ranges. (A) Average results from 4 substrains of strain B (B) Results from strain C (C) Average results from 2 substrains of strain W (D) Average results from 4 substrains of strain K12. The base penalties used prevent total scores between 0.6 and 0.7. In general the negative strand had slightly more hits. Perfect scores (value of 1.0) were a small fraction of all hits with scores >= 0.5, with most hits having total scores between 0.8 and 0.99. The E.coli strains showed very similar results, although strain B had slightly fewer hits than the others and K12 showed a much larger amount of variation between its substrains.

It should be noted that there are several regions where localized hit densities have exceeded those of the peaks chosen, however, the peak regions track reasonably well with areas showing the high peak densities. Figure 5.6 demonstrates this for the positive strand data from the *E. coli* strain C. This suggests that although the G3 hits appear to be relatively uniformly distributed, there are some regions of sustained enrichment. The run peak analysis plots also support generally low variation among the substrains of *E. coli* investigated. However, there were two distinct pattern types among the four substrains of strain B that were investigated (See appendix C for additional figures).

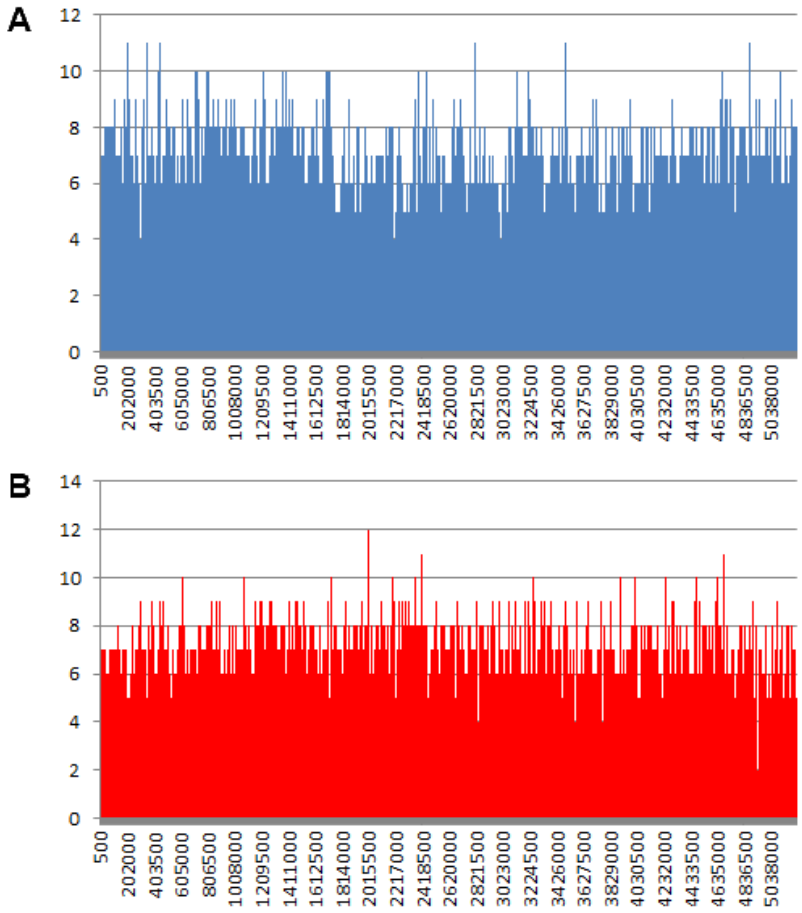


Figure 5.5: Representative E. coli Hit Distributions. (A) Positive strand hit distribution for strain C (B) Negative strand hit distribution for strain C. The y-axis displays bin hit counts while the x-axis shows the location in the genome in nucleotides. The bin size was 500nt.

The score distributions (Figure 5.7) for *S. cerevisiae*, which used the same parameters as for the *E. coli* analysis, show an enrichment of higher scoring hits, with the 0.9-0.99 category having approximately twice as many hits as the second highest category (0.8-0.89). This contrasts with the *E. coli* results where the two categories contained nearly the same number of hits. This suggests that compared to *E. coli*, *S. cerevisiae* tends to have higher scoring hits. However, like the *E. coli* results, the *S. cerevisiae* hit distributions (Figure 5.8) show a relatively uniform distribution of hits with total scores greater than or equal to 0.5.

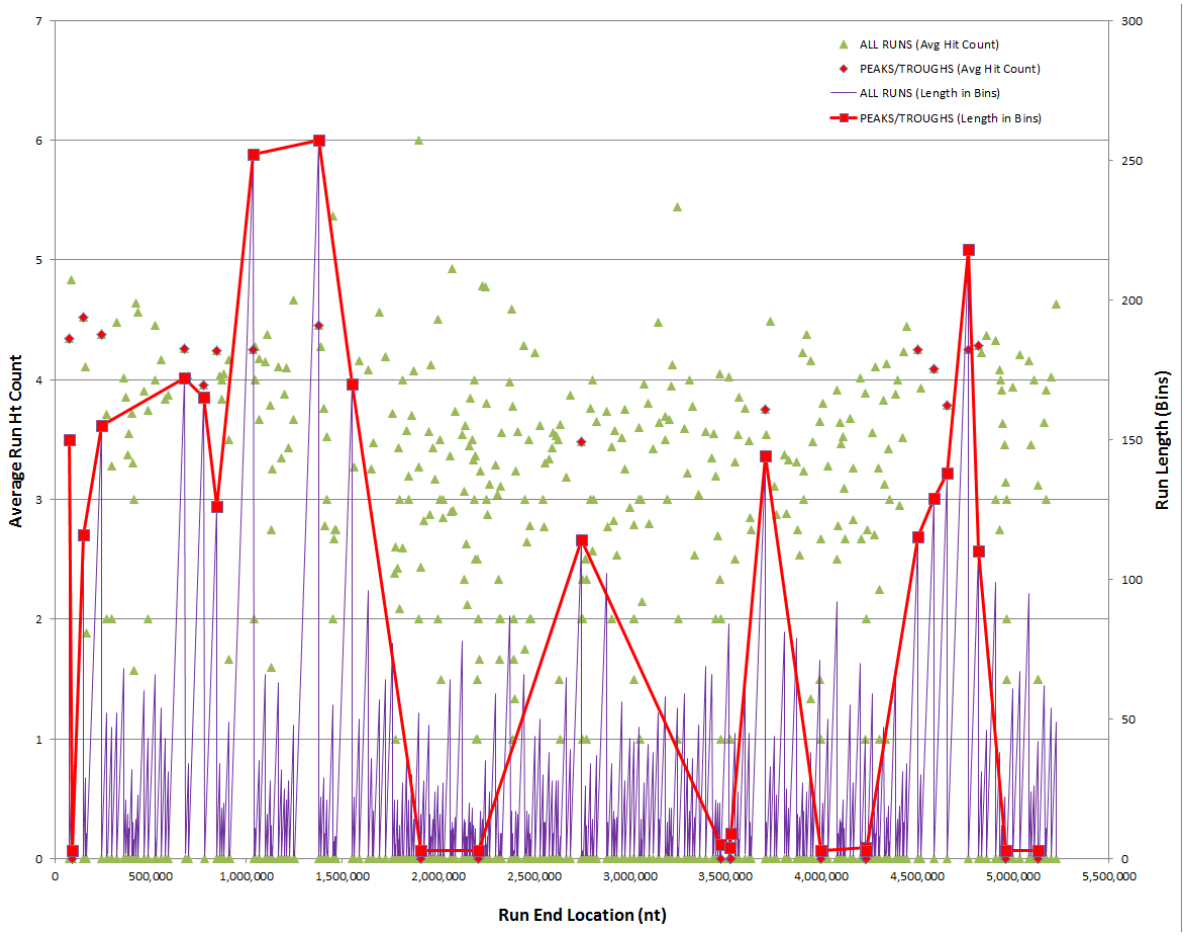


Figure 5.6: Example E. coli Run Peak Analysis. Data corresponds to that taken from scanning the positive strand (Normal5to3 scan) of *E. coli* strain C. Green triangles mark the locations of run ends with the indicated average hits/bin. Small red diamonds mark those runs that were selected as either peaks or troughs. Purple lines show how the runs varied in length. Red lines with red boxes outline the peaks and troughs that were selected by G-Trip based on run length.

Peak detection reveals that, similar to the case with *E. coli*, the *S. cerevisiae* genome also favors longer peak run lengths than troughs. However, the average number of hits per bin in the runs dropped to about 2.4, or approximately one hit every 208 nt. The peak runs were also generally shorter in length, and were more numerous in *S. cerevisiae* than in *E. coli*. Another interesting occurrence was that the hit distributions were less scattered in *S. cerevisiae*. With clear "bands" present at hit densities of 2, 3, and 4 hits per bin (Figure 5.9).

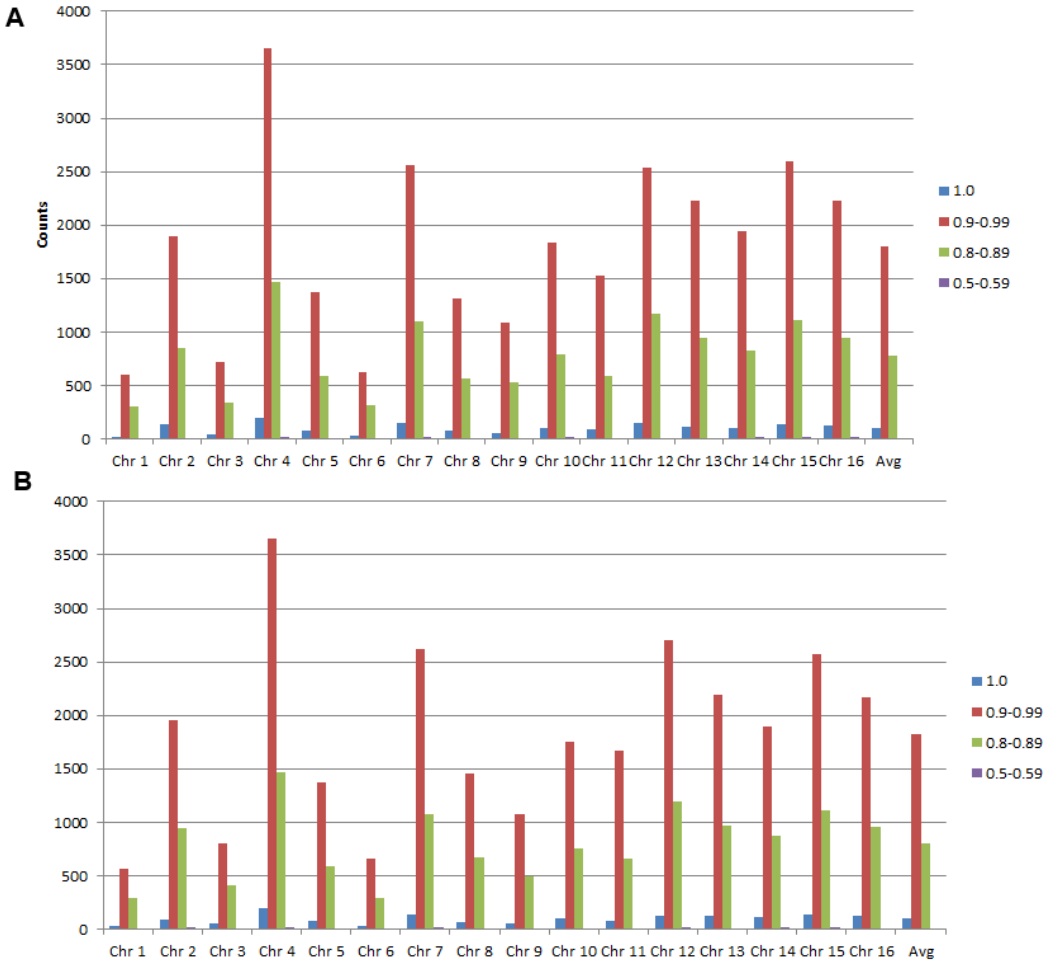


Figure 5.7: S. cerevisiae Score Distributions. The data is shown for each of the 16 chromosomes present in *S. cerevisiae* as well as for an average of the chromosomes for (A) the positive strand and (B) the negative strand. Compared to the *E. coli* results, scores of 0.9-0.99 are much more enriched relative to the other score categories (0.6-0.69 and 0.7-0.79 are omitted because the scoring settings used prevent values from falling in these ranges).

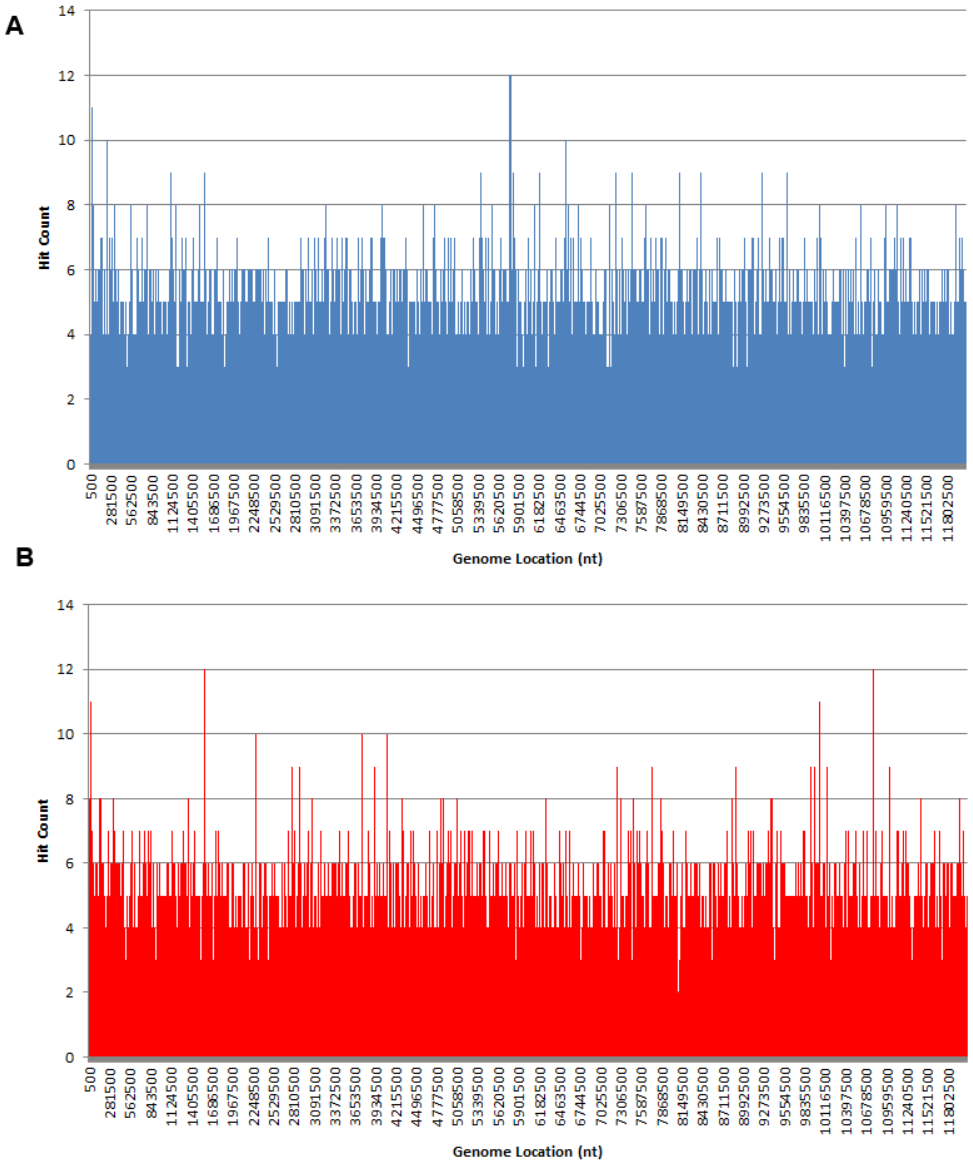


Figure 5.8: S. cerevisiae Hit Distributions. (A) Positive strand hit distribution (B) Negative strand hit distribution. The y-axis displays bin hit counts while the x-axis shows the location in the genome in nucleotides. The bin size was 500nt.

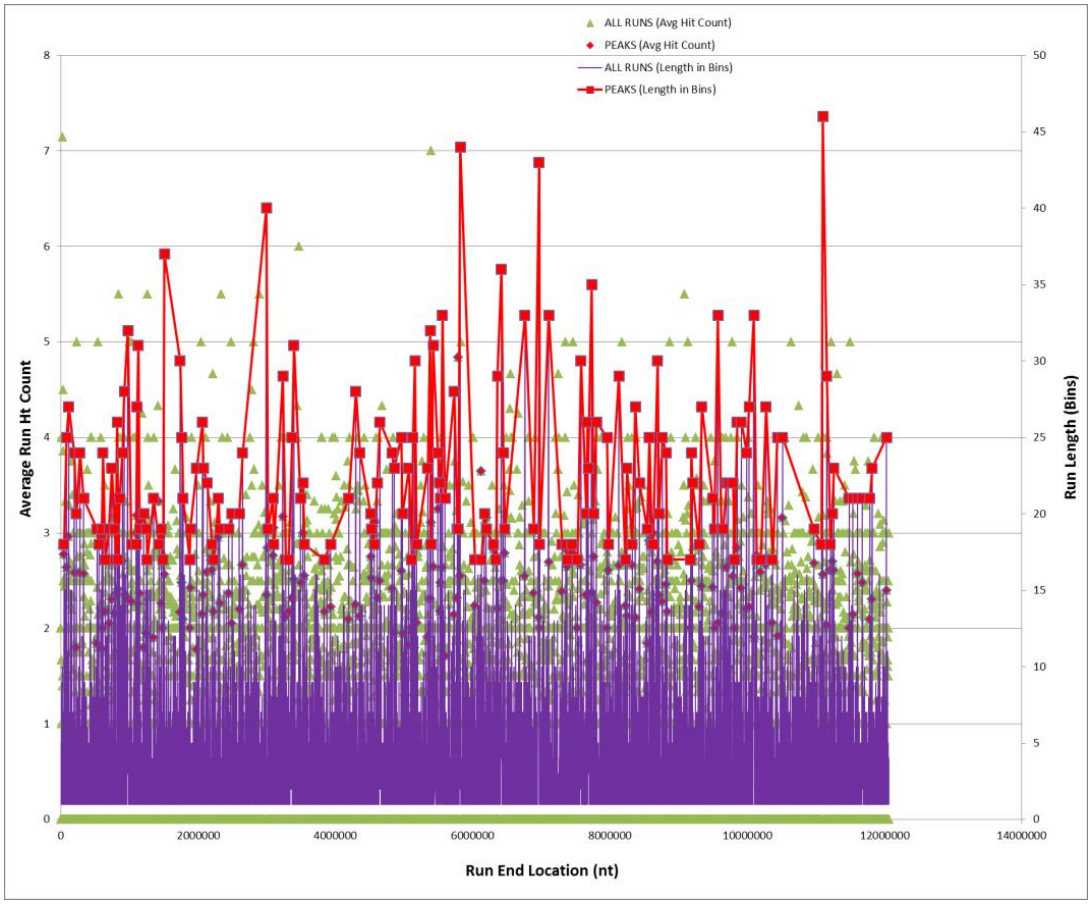


Figure 5.9: Example S. cerevisiae Run Peak Analysis. Data corresponds to that taken from scanning the positive strand (Normal5to3 scan) of *S. cerevisiae*. Green triangles mark the locations of run ends with the indicated average hits/bin. Small red diamonds mark those runs which were selected as either peaks or troughs. Purple lines show how the runs varied in length. Red lines with red boxes outline the peaks (troughs are omitted to ease interpretation) that were selected by G-Trip based on run length.

Due to large size of the *H. sapiens* genome and the enormous number of hits detected (in the tens of millions), analysis with conventional spreadsheet programs proved impractical. Thus, for *H. sapiens,* only the top scoring hits (hits with a "perfect" score of 1.0) were used for the preliminary analysis. Another difference is that a bin size of 1000nt was used for the generation of the hit density plot. The peak analysis still uses a 500nt bin size. The way that the *H. sapiens* data was analyzed limits direct comparison with the other two species studied, especially given the low percentage of "perfect"

scoring hits in both *E. coli* and *S. cerevisiae*. However, like the data for *S. cerevisiae*, the hits found for *H. sapiens* formed bands of hit densities, except that for *H. sapiens*, the effect was much more pronounced. For the first chromosome, bands were observed at 1, 1.3, 2, and 3 hits per bin. G3 hits with a score of 1.0 are relatively rare in *H. sapiens*. While the number of peaks still outnumbers that of the troughs, the trough run lengths are many times larger than those of the peaks (by several hundred fold in some cases).

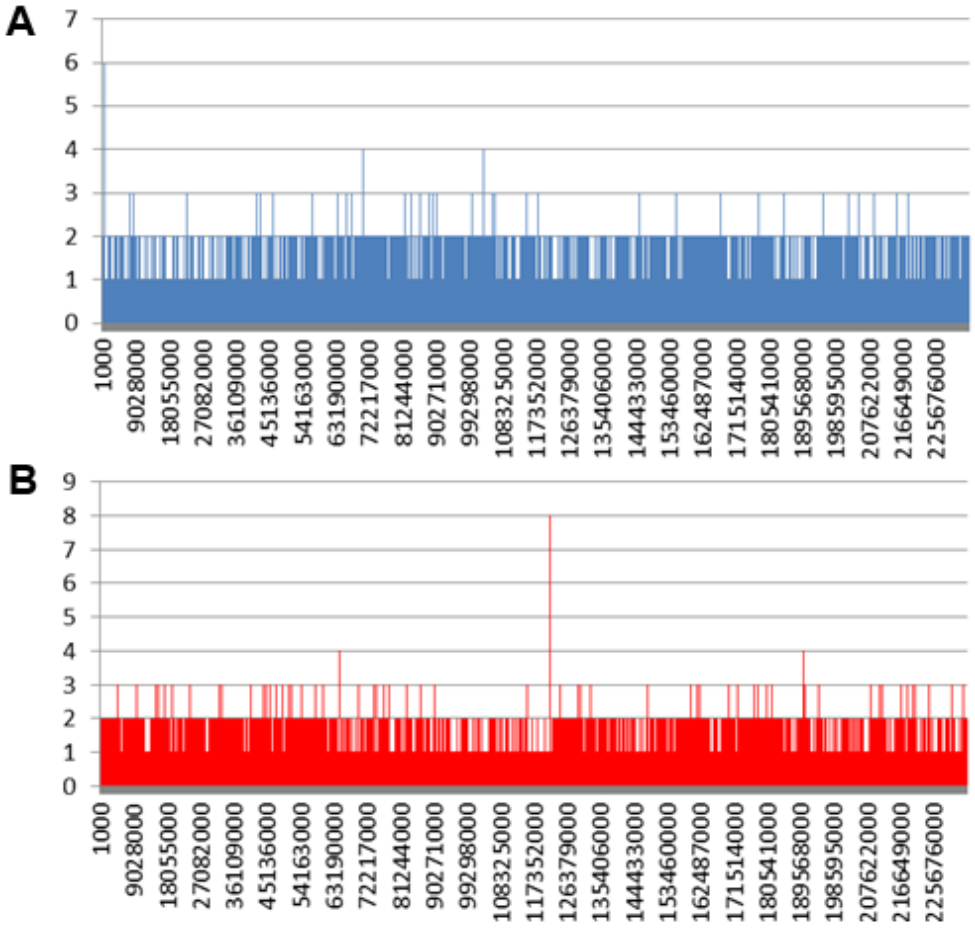


Figure 5.10: H. sapiens Chromosome 1 Hit Distributions. (A) Positive strand hit distribution (B) Negative strand hit distribution for chromosome 1. The y-axis displays bin hit counts while the x-axis shows the location in the genome in nucleotides. The bin size was 1000nt.

On average, the peaks obtained for the hits with total scores equal to 1 had 1.07 hits per bin or approximately 1 hit every 468nt. The average peak run length was about 3 bins. As shown in Figure 5.11 below, the troughs, with runs up into the hundreds of bins are the dominant feature. Especially interesting is that most of the chromosomes contained at least one very large hit trough. This occurred more often on the positive strand, but was often found in the negative strand as well (see Appendix C for plots corresponding to each of the chromosomes of *H. sapiens*).

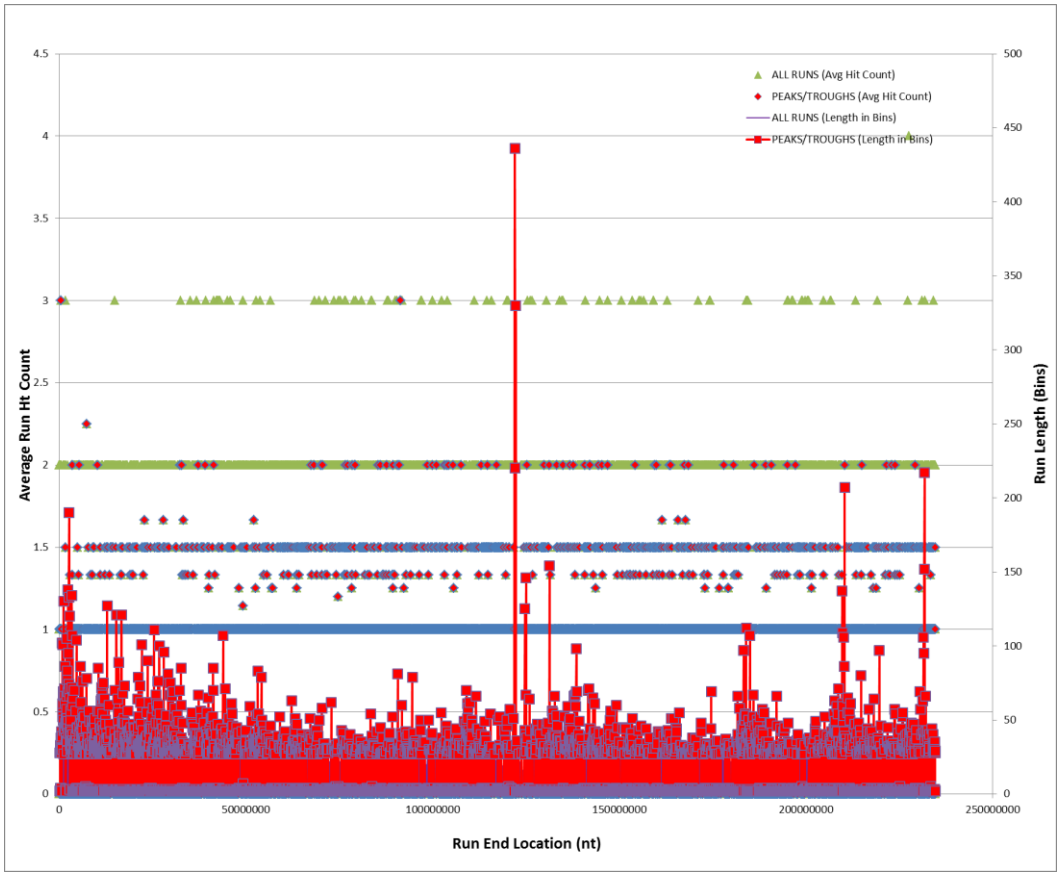


Figure 5.11: Example H. sapiens Run Peak Analysis. Data corresponds to that taken from scanning the positive strand (Normal5to3 scan) of *H. sapiens* chromosome 1. Green triangles mark the locations of run ends with the indicated average hits/bin. Small red diamonds mark those runs which were selected as either peaks or troughs. Purple lines show how the runs varied in length. Red lines with red boxes outline the peaks and troughs that were selected by G-Trip based on run length. The apparent red band near the bottom of the plot represents the space primarily occupied by hit run peaks.

The annotation data is extremely preliminary and has only been partially assessed for chromosome 1 from *H. sapiens* on the positive strand. Chromosome 1 had 6510 peaks, 51% of which were associated with at least one gene, and 3% of which were associated with two genes. Approximately one third of peaks associated with at least one gene were associated with a unique gene. This is compared with 1935 troughs, of which 73% were associated with at least one gene, 21% of which were associated with two or more genes. Roughly two-thirds of the troughs associated with at least one gene were associated with a unique gene. While at first glance, this appears to suggest that troughs are more often associated with genes than are peaks, this is before taking into account the lengths of the runs associated with these. Peaks average only 2 bins in length, or approximately 1000nt. The troughs on the other hand average 38.5 bins in length, or approximately 19,250nt. Over such large spaces, the chance of a trough containing a gene is significantly higher than it is for a peak. In this context, the number of peaks associated with genes is at least preliminarily suggestive that there may be an association between G3 structures and genes. What this association might be remains to be determined. A brief survey of the specific genes associated with the peaks in chromosome 1 did not immediately reveal an association with a known gene function, but as more of the data is analyzed, such an association may become apparent.

FUTURE DIRECTIONS

The current version of G-Trip has demonstrated the ability to detect sequences with the potential for forming G3s. These hits can be scored with penalties for the presence of competing G4 structures, and some preliminary analysis of the distributions of the hits can be conducted. However, there is a great deal of room for improvement,

which can be guided by increased understanding of G3 structures gained through lab experiments.

As discussed above, the scoring system for G-Trip is an arbitrary one and will almost certainly need to be reworked as understanding of the relative stabilities of G3 structures increases. Aside from the specifics of how scores are calculated, there are additional considerations. G4s are not the only stable competing structures that could weaken or invalidate a G3 hit. Regular triplexes and high-stability hairpins could also compete. Currently, these alternative competing structures are not factored into the score. The addition of prediction algorithms for these alternate structures and consideration of their relative contributions into the final scoring would make the G3 predictions far more robust. More robust predictions will also increase the likelihood that predicted structures could have a biological role related to the G3 structure formation itself.

Another important improvement to the scoring scheme for G3 predictions would consider the stability of the predicted G3 structure. While G3 structures may have their own specific characteristics, they are similar enough to G4 structures that a G4-based stability model could be used as a reasonable first approximation. In general, G4 structures benefit from having more tetrads, which correspond to longer G-tracts. Their stability also tends to suffer from longer loop lengths. A careful search of the literature on G4 structures may provide sufficient examples for the creation of a generalized empirical stability prediction algorithm based on analysis of the G-tract and loop lengths. This algorithm could be further refined to better apply it specifically towards G3 structures.

Besides improvements in scoring of G3 hits, G-Trip would also benefit from improvements in its analysis functions. For example, the run peak/trough detection could

be improved by better incorporating the scores of the hits being binned. This would allow not only hit densities to be determined, but allow the assignment of quality values to the regions with high/low hit densities. While this could simply be done by applying a quality value determined from the average of the scores of the hits that compose a given run, it could also be used to improve run selection in the first place, improving the peak/trough detection. Trough detection, in particular, would benefit from these improvements because the assumption that only empty bins constitute troughs is not necessarily an accurate one, as determined in *E. coli* and *S. cerevisiae*, which were both shown to have relatively high hit densities.

In order to further analyze the data, it would be beneficial to include more powerful statistical analysis tools in G-Trip. While such tools need not necessarily be built into the G-Trip program, the raw data is already present in a form that G-Trip can interpret and the further that G-Trip can take the data before relying on other methods, the greater its utility to researchers. A major improvement would be a function that can aid the user in building a statistical prediction model based on the detection pattern used in the prediction of the G3 structures. The model could then be used to probe whether the detected hits represent selective pressure by determining if there are significantly fewer or more hits than would be expected by chance. The existence of selective pressure would be a strong indicator of the existence of a functional role.

Finally, G-Trip could be made faster and more efficient through optimization. While efforts were made to make G-Trip both robust and fast, it has not yet been formally optimized. Optimizations could include revisions to the code, especially at the junctions between functions, in order to increase speed and perhaps handling of memory as well. The algorithms used by G-Trip could also potentially be improved to increase performance.

CHAPTER 5 REFERENCES

1. Bardin,C. and Leroy,J.L. (2008) The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.*, **36**, 477–488.

2. Stadlbauer,P., Trantirek,L., Cheatham,T.E., Koca,J. and Sponer,J. (2014) Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie*, **105C**, 22–35.

3. Rajendran,A., Endo,M., Hidaka,K. and Sugiyama,H. (2014) Direct and Single-Molecule Visualization of the Solution-State Structures of G-Hairpin and G-Triplex Intermediates. *Angew. Chem.-Int. Ed.*, **53**, 4107–4112.

4. Cerofolini,L., Amato,J., Giachetti,A., Limongelli,V., Novellino,E., Parrinello,M., Fragai,M., Randazzo,A. and Luchinat,C. (2014) G-triplex structure and formation propensity. *Nucleic Acids Res.*, **42**, 13393–13404.

5. Kikin,O., D'Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–W682.

6. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.

7. Cock,P.J.A., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B., *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinforma. Oxf. Engl.*, **25**, 1422–1423.

8. Mukundan,V.T. and Phan,A.T. (2013) Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.

Chapter 6: Summary of Conclusions and Future Directions

After conducting the experiments discussed in chapters 2-4, G4-selective photocleavage compounds were synthesized, screened, and then a subset were selected for more in depth study into their photocleavage and binding selectivities. Each member of this subset was a representative of a different chemical family. Chapter 5 discussed the development of a tool to aid in the study of G-triplexes. This chapter summarizes conclusions drawn from the combined results of these studies in order to suggest potential avenues for future investigations.

PHOTOCLEAVAGE ACTIVITY AND SELECTIVITY

The screen results from the first and second generation libraries (Chapter 1) indicated some interesting structure-activity relationships for the photocleavage and binding activity of the compounds. The triazolium group present in the first generation library appeared to greatly enhance binding as determined increases in DNA-compound melt T_m (Chapter 3), but at the cost of lower cleavage relative to the second generation compounds which lacked this group. This enhanced binding only occurred with those first generation compounds that incorporated a longer carbon chain linker, suggesting a minimum amount of spacing required for the favorable association with the G-quadruplex. This spacing was likely required to give the positively charged group access to the negatively charged DNA backbone. The lower cleavage observed with the first generation library relative to the first may also be partly due to the triazolium group if it's association with the DNA orients the photocleavage moiety in a less active position.

Studies into the selectivity of **21b**, NMM, and TEL011 (Chapter 4) suggest that all three are reasonably selective (roughly 100X) for G4 structures over duplex DNA, although additional improvement is still necessary to increase the selectivity, especially

relative to other competing secondary structures such as triplexes. The photocleavage screens (Chapter 3) suggest that photocleavage activity of these compounds is still somewhat low. This can be compensated for by improvements in solubility, which, by increasing photocleavage efficiency, would allow for lower photocleavage activities to be compensated for by increased concentrations of probe. Alternatively, the photocleavage activity needs to be further enhanced. Structures that better orient the photocleavage moiety may accomplish this, perhaps through the addition of a second linker region on the "azide side" of the triazole formed in the "click" reaction. A second linker region could allow more flexibility in the molecules enabling them to better orient themselves on binding. The preparation of additional libraries can also increase the likelihood of identifying an optimal structure to serve as a G4 structure probe.

At the current level of development, the probes are limited to temporal detection resolutions on the scale of 15-30min, which means that only relatively long-lived G4 structures can be detected. While this may not allow for the detailed study of the progress of a biological process, such as unwinding by a helicase, it can be used to in the detection of G4 structures that form readily or often.

In addition to bringing development of a G4 specific small molecular probe closer to completion, these studies have also identified several potential ways for improving G4 ligand design which can be applied not only to design better probes, but to also help inform development of small molecule ligands as therapeutic agents.

FUTURE DIRECTIONS IN SMALL MOLECULAR PROBE DESIGN

The differences observed between the first and second generation library suggest some potential modifications to the structure that could potentially improve both binding and photocleavage. Since the triazolium group appeared to aid in binding with longer

linker lengths, a positive charge in this location is desirable. However, to offset the potential for poor alignment of the photocleavage moiety, a second linker could be introduced to allow the photocleavage moiety to better position itself on binding. This modification could be accomplished by introducing the second linker into the azide precursor and would allow most of the steps of synthesis for the first generation library to remain the same.

The control compound **23** made for the second generation that lacked the linker and photoactive group suggested that binding was inhibited by the linker and/or the photocleavage moiety. While adding a second linker may address this issue, it can also be tested directly through the synthesis of additional control compounds which contain the linker, but which lack a photoactive group. This could help identify whether or not the photocleavage moieties are responsible for disrupting binding.

The strong coordination of copper seen in the first and second generation libraries (Chapter 1) suggests another interesting direction for future investigation. It may be possible that different metal complexes with the scaffolds used in both libraries may have intrinsic DNA cleavage properties, if not through photocleavage, possibly through oxidative cleavage. The preparation of a library of metal-complexes with these scaffolds incorporating different ions may identify additional cleavage agents.

Several additional alternative libraries could also prove fruitful. Of those discussed in Chapter 1, the amine arm approach to functionalizing the 360A-like scaffolds, the asymmetrically functionalized swallow-tail PDIs, and the PDI macrocycles remain as interesting areas of study. The amine arm linker could be used to introduce a pH-dependent charge that could act as an inducible analogue to the charge of the triazolium group from the first library. The asymmetry of the functionalized swallow-tail PDIs could prove a distinct advantage in binding to and photocleaving asymmetric G4

structures while the PDI macrocycles may improve on the low photocleavage efficiency observed in TEL011 (Chapter 3) by allowing for "preaggregation".

To further characterize both the compounds discussed in this dissertation, as well as future compounds, several additional tests could be conducted. The first of these is to obtain the kinetic association constants between these compounds and the labeled DNA probes. This would help in interpreting whether future probes would be most aided by improving binding or improving photocleavage activity. Kinetic parameters could be obtained through several methods including UV-VIS spectrometry, surface plasmon resonance (SPR), or isothermal calorimetry (ITC). In addition, assays testing the cell permeability of these compounds would be useful in determining their utility for *in vivo* use and could help direct the development of future compound libraries.

G-TRIP

The G-Trip program allowed the detection of potential G3-forming sequences and has made some rudimentary attempts at scoring these potential G3 sequences in terms of competition with likely more stable G4 structures (Chapter 5). It has also allowed for preliminary analysis of the density of G3 hits within the source sequence(s). This analysis has shown that there are indeed broad regions within the genomes of *E. coli*, *S. cerevisiae*, and *H. sapiens* that show heightened G3 hit densities.

However, in the future, these preliminary results will require the application of more rigorous statistical models to aid in comparing the occurrence rates observed with what could be expected to occur purely by chance. Such an analysis would be able to determine if there appears to be a selective bias for or against G3 structures in the genomes studied. In addition, an analysis of the hit density in relation to the sequence of

the associated hits may be able to identify particular motifs which most favor G3 structures.

The scoring algorithms of G-Trip would also benefit from expansion to account for other secondary structure competitors such as triplex DNA as well as a score that relates to the relative stability of the predicted G3 structures. These improvements would allow for stronger selection for those G3 structures most likely to be adopted in vivo and could enhance statistical analysis.

Appendices

APPENDIX A: GENERAL METHODS

Common Buffer Preparations

Potassium cacodylate and EDTA (PCE) stock buffers were prepared by titrating a 250mM aqueous solution of cacodylic acid containing 12.5mM EDTA with a 1M potassium hydroxide solution to pH 7.4. If necessary, potassium chloride was added to bring the potassium concentration to 250mM. This stock buffer was then diluted as needed to prepare samples for photocleavage and melts. Lithium cacodylate and EDTA (LCE) buffer was prepared similarly, only 1M lithium hydroxide was used for the titration. Sodium cacodylate and EDTA (SCE) was prepared by dissolving 250mM sodium cacodylate in water with 12.5mM EDTA and adjusting the pH as needed with either 2M HCl or 1M NaOH.

One liter stocks of 10X (0.5M) stock of Tris, Boric Acid, and EDTA (TBE) buffer were prepared by dissolving 108g Tris Base and 55g of Boric Acid and adding 50mL of 0.5M EDTA. The pH was adjusted to 8 with the addition of either 10M HCl or 10M NaOH.

DNA Preparation

HPLC purified F21T (5'-dGGG(TTAGGG)$_3$-3') and cMYC (5'-dTGAGGGTGGGTAGGGTGGGTAG-3') dual labeled with a 5' FAM label and a 3' TAMRA labels were purchased from Integrated DNA Technologies (IDT) and used without further purification. The plasmid ϕX174 RF-1 was purchased from Fisher and used without further purification.

Calf thymus (CT) DNA was purchased from Sigma-Aldrich. 100mg was dissolved in a minimum (<50mL) of BPE (6mM Na_2HPO_4, 2mM NaH_2PO_4, 1mM Na_2EDTA; pH 7.0) buffer and dialyzed with a molecular weight cut off (MWCO) of

7000 against BPE buffer overnight. The DNA solution was then sonicated on ice at 65% amplitude for five minutes before being sparged with argon for five minutes. This process was repeated until the DNA had been sonicated for a total of 30min. The sodium chloride concentration was then increased to 250mM and RNase was added to a concentration of 0.5mg/mL before the sample was heated at 37°C for 30min. The DNA solution was then extracted three times with phenol-chloroform and the combined aqueous layers washed three times with water-saturated ether. The combined aqueous layers were then dialyzed with a MWCO of 7000 against BPES (6mM Na_2HPO_4, 2mM NaH_2PO_4, 1mM Na_2EDTA, 185mM NaCl; pH 7.0) buffer overnight. The DNA solution was concentrated under reduced pressure and was veriifed to have an approximate size of 200bp (1% agarose gel in TBE) and to have a melting point of 65-66°C as detected by absorbance spectrometry at 260nm. The DNA was then aliquoted out and dried down as 500μg stocks for use as either unlabeled compeitor or in the dissociation buffer used for the photocleavage assays. DNA was stored at -20°C until used. CT DNA yields were typically between 80 and 90%.

The pSP189 (1) and pSP189G4 (pSP189 containing a G4 DNA insert: 5'-dAGCTCAATGGGGTTGGGGTGGGGTTGGGGCGC-3' between the EcoRI and XhoI restriction sites) plasmids were provided as kind gifts from the lab of Karen Vasquez. Electrocompetent DH5α cells (kind gifts from the lab of Walter Fast) were transformed with the plasmid using general transformation procedures. pSP189G4 was then harvested and purified using a Qiagen microfuge spin column kit and eluted with 10mM Tris pH 8. Concentration and purity were verified on a Thermo Fisher Nanodrop 2000 spectrometer. Samples were stored at -20°C until used. Deconcatenated pSP189 was obtained through the retriction digest of 3ug of plasmid with 10U XhoI in 50μL 1X NEB buffer 2 (50mM NaCl, 10mM Tris-HCl, 01mM $MgCl_2$, 1mM DTT, pH 7.9) overnight at 37°C. The

solution was diluted to 100μL by the addition of 50mM Tris buffer pH 8, and extracted with phenol-chloroform-isoamyl alcohol. The aqueous layer was then extracted twice with chloroform-isoamyl alcohol and the DNA precipitated by the addition of 240μL ice-cold ethanol and storage at -20°C overnight. The DNA was pelleted by centrifugation at top speed on a conventional benchtop centrifuge at 4°C for 45min. After decanting the solution, the pellet was washed with 200μL of ice-cold 70% aqueous ethanol, centrifuged for a further 45min, the solution decanted, and the pellet dried. The pellet was then resuspended in water. 20ng of the restriction digested DNA was then treated with 10U of T4 ligase in 20μL 1X NEB ligation buffer (50mM Tris-HCl, 10mM $MgCl_2$, 1mM ATP, 10mM DTT, pH 7.5) overnight at room temperature. This DNA was then used to transform DH5α cells. The pSP189 plasmid was purified in the same manner as pSP189G4.

G and GA Ladders of F21T and cMYC were prepared using Maxam-Gilbert sequencing methods. Preparation of the G ladders was accomplished by adding 10μL of 10μM F21T or cMYC in water to 200μL of 50mM SCE buffer. 1μL of DMS was added and the tube incubated at 25°C for 3.5min before adding 50μL quench solution (1.5M sodium acetate pH 7.0, 1M β-mercaptoethanol, 100μg/mL tRNA) and mixing before additional shared steps with the GA ladders were conducted as described below. Preparation of the GA ladders was accomplished by adding 10μL of μM F21T or cMYC in water to 25μL formic acid. The tube was then heated at 25°C for 6min before 200μL stop buffer (0.3M sodium acetate pH 7, 0.1mM EDTA, 25μg/mL tRNA) was added. 750μL ethanol was added to either the G or GA ladder preparations to begin DNA precipitation at -80°C for 1 hour. The precicipate was pelleted, washed with 70% aqueous ethanol, resuspended in 120μL 0.6M sodium acetate, and precipitated once more by addition of 500μL ethanol storage at -80°C for 1 hour. The precipitate was pelleted a

second time, washed twice with 70% aqueous ethanol before being dried completely. Pellets were resuspended in 70μL 10% piperidine (with freshly distilled piperidine) and incubated at 90C for 30min. The samples were evaporated to dryness on a speedvac, washed with 30μL water, dried again,washed a final time with 20μL, and dried once more. The samples were dissolved in 100μL water and used to prepare then 10μL aliquots which were dried down and stored at -20°C until used. On first use, the samples were dissolved in denaturing formamide loading buffer (80% formamide, 10mM NaOH, 1mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue), and heated at 90°C for 3min before loading on a gel (see below)

Internal photocleavage standard stocks of F21T and cMYC were prepared by treatment of 100pmol F21T or cMYC with 10U of S1 nuclease in 10μL for 15min at 37°C before quenching with 240μL of 50mM PCE or SCE buffer.

Photocleavage Assay

Each sample condition was prepared in triplicate. Reactions were done on a fluorescence compatible (black) 384-well plate. Each well held 100nmol of dual-labeled DNA in either PCE or SCE buffer and with the photocleavage compound being screened in a total volume of 20μL. After irradiation in a LuzChem photoreactor with lamps centered at the appropriate wavelength (all irradiations with visible light included a glass plate to filter out UV light), 95μL dissociation buffer containing a vast excess (0.05μg/μL) of CT DNA in water or an appropriate (PCE/SCE) buffer was added. The plate was then sealed with plate film and heated for 30min at 85ºC before being slowly cooled to room temperature overnight. The plate was then spun in a plate centrifuge to remove condensation on the film and the film removed. Internal standards were then loaded into predesignated empty wells and the FAM fluorescence of each well read five times in a Victor 3V plate reader. Results from the wells were averaged and any signal

from buffer was subtracted. The average apparent percent photocleavage was then calculated using the equation below.

Apparent Percent Cleavage = $((F_x-F_{min})/(F_{max}-F_{min}))*100$; where F_x is the background-subtracted well signal, F_{min} is the signal from an untreated control well, and F_{max} is the signal from the appropriate internal DNA standard.

Polyacrylamide Gel Electrophoresis (PAGE)

40% acrylamide (38:2 acrylamide:bisacrylamide) was prepared by mixing 500g acrylamide, 26.3g bisacrylamide, and 132mL 10X TBE buffer followed by dilution to 1315mL with nanopure water. The 20% denaturing acrylamide stock was prepared by mixing 200mL 40% acrylamide buffer with 168g Urea and 40mL 10X TBE buffer followed by dilution to 400mL with nanopure water. Buffers were filtered through a 0.2μm membrane before use.

To prepare for gel electrophoresis, the triplicate wells from the photocleavage assay above were combined. Half of the combined sample was treated with 1M piperidine and heating for 30min at 90ºC while the second half was left untreated. The photocleavage products were dried down with a Savant Speed Vac vacuum centrifuge, resuspended in denaturing formamide loading buffer, and analyzed on 20% denaturing polyacrylamide gels run with 1X TBE buffer. Gels were preheated before loading samples by running at 10W (constant current) for 30min. The gel was then run at 8W (constant current) for approximately 2.25 hours. Gels were imaged through FAM fluorescence on a Typhoon Trio or Typhoon 9500 gel imager. Bands were quantified using the program GelQuant.Net provided by biochemlabsolutions.com and the relative percent cleavage calculated using the equation below.

Relative Percent Cleavage = $((C_x/Tot_x-FC_0)/(1-FC_0))*100$

Where C_x is the sum of the cleavage band intensities, Tot_x is the total band intensities in a lane, and FC_0 is the ratio of cleavage band intensities to total band intensities in a control lane loaded with the untreated control.

DNA Melts

DNA melts used 200nM of either F21T or cMYC in 5mM PCE or SCE and treated with varying concentrations of photocleavage compound and/or unlabeled competitor DNA. The total sample volume loaded into the optical glass cuvettes was 400µL. Before the addition of photocleavage agent and/or unlabeled competitor DNA, the dual labeled DNA sample was melted and reannealed three times both to degass the sample and to increase reproducibility of the secondary structure that was formed. FAM fluorescence was measured on a Cary Eclipse fluorimeter and the melting temperature determined through a nonlinear regression fit to the logistic equation below. The program SciDAVis was used to conduct the nonlinear regression.

$$I = I_0+(I_f-I_0)/(1+\exp(-s*(T-T_m)))$$

Where I is fluorescence intensity, I_0 is initial fluorescence intensity, I_f is final fluorescence intensity after melting, s is the signal increase rate (a scaling factor), T is temperature, and T_m is the melting temperature.

Circular Dichroism (CD) Assays

CD data was collected on a Jasco J-815 CD spectrometer. Samples containing 28-30µM strand of dual or unlabeled DNA in 12.5mM PCE buffer were pre-annealed (heated to 95°C before being slowly cooled to room temperature) and then incubated at 15°C for several minutes before reading was initiated. Samples were loaded into 1mm CD cuvettes and were scanned five times with averaging from 350nm to 200nm (dual labeled DNA was scanned from 600nm to 200nm). The spectrum for a buffer-only control was subtracted from each of the spectra. A 5X concentration of compound

(NMM) was added to the unlabeled DNA sample and incubated for 24 hours at 4°C before being scanned again. The DNA sample containing compound was then reannealed and scanned a third time. After incubating an additional 48 hours at 4°C, a fourth scan was obtained of the same sample.

Dimethyl Sulfate (DMS) Protection Assay

Samples were prepared in PCR tubes so that thermo-cyclers could be used to regulate the reaction temperature. Stock solutions were pre-incubated at 25ºC before being mixed. Each 25μL sample contained 400nM of dual-labeled DNA (F21T or cMYC), 40mM PCE or SCE, 0.06μg/μL of t-RNA as a carrier, and 0.5v% of DMS. After reacting for 10min, 10μL of ice-cold stop buffer (4M β-mercaptoethanol, 1.5M NaOAc pH 5, 2μg/μL t-RNA) was added. The DNA was then concentrated and purified through ethanol precipitation overnight at -20ºC and pelleting at 4ºC for 45min in a bench top microcentrifuge. The pellet was washed with 70% ice-cold aqueous ethanol and repelleted. After being treated with piperidine and heating at 90ºC for 30min, the cleavage products were examined with PAGE as described above. Cleavage was compared with that seen on an unfolded (heated to 95°C but then immediately placed on ice, instead of allowing to slowly reanneal) sample prepared in LCE.

Plasmid Nicking Assay

The "duplex" plasmid (ϕX174 RF-1) was prepared at 50μM bp in 26μL. 4μL of compound in water was then added such as to give varying concentrations of compound relative to the bp concentration of plasmid. The 30μL reaction sample was then loaded into one well of a 96-well-plate and irradiated with lamps appropriate for the photocleavage compound being used for 2 hours. A sample that was not irradiated was also prepared as a control. 15μL of each sample was then mixed with 4μL of 5X DNA loading dye and loaded into a 1% TBE agarose gel run at 80V constant voltage for 60-

90min. The gel was then stained with 0.5µg/mL ethidium bromide in 1X TBE for 15min, destained in water for 15min, and imaged with a BioRad trans UV imaging system.

APPENDIX B: SYNTHESIS OF COMPOUNDS

All reactions were conducted under argon in oven-dried glassware and stirred magnetically. Unless stated otherwise, all materials were obtained from commercial sources and used without further purification. THF was distilled over sodium/benzophenone prior to use. Dichloromethane and 1,2-dichloroethane were distilled over CaH2 prior to use. Unless stated otherwise, organic extracts were dried with Na_2SO_4, filtered through a fritted glass funnel and concentrated under a rotary evaporator (20-30mm Hg). Rf values are reported for analytical thin-layer chromatography (TLC) performed on EM Reagent 0.25mm silica gel 60-F plates with UV light and/or KMnO4 stain visualization. Flash chromatography was performed with EM Reagent silica gel (230-400 mesh), using the indicated mobile phase. Melting points (open capillary) are uncorrected. Unless otherwise noted, 1H and 13C spectra were determined in CDCl3 on a spectrometer operating at 400MHz and 100 MHz, respectively, and are reported in ppm using solvent as an internal standard (7.26ppm for 1H and 77.0ppm for 13C). Unless otherwise noted, all mass spectra were obtained in the positive mode either by chemical ionization using methane as the ionizing gas or by electrospray ionization. Unless otherwise indicated, HPLC traces were obtained for the HCl salts of the final library compounds using a Varian Star HPLC pump and detector system and a Kromasil Eternity C18 column outfitted with a C18 guard column. HPLC samples were prepared in water with a small percentage (4-10%) of DMF to aid in dissolution and to mark the void volume of the column. 20-40µL of sample was loaded into a 50µL sample loop and traces were detected at 440 +/- 10nm. Elution was obtained

by running 10:90% (Acetonitrile:10mM Ammonium Acetate pH 4.5) of mobile phase at 0.7mL/min on a continuous concentration and rate gradient up to 100% Acetonitrile at 1mL/min over 30min and then held at 100% Acetonitrile at 1mL/min for 10min. After each run, the sample loop and column were flushed with a 20-50µL injection of 5% DMF in water to ensure no cross contamination between runs.

***N*,*N*'-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (360A Precursor)** (2)**.** 200mg, (1eq, 1.2mmol) of pyridine-2,6-dicarboxylic acid was added to excess (10mL) of neat thionyl chloride and stirred at reflux for 3 hours while under argon. After 3 hours, the remaining thionyl chloride was distilled off, leaving a crude residue of pyridine-2,6-dicarbonyl dichloride, which was used immediately without purification. 345mg (2eq, 2.4mmol) of 3-aminoquinoline was added to the dichloride residue, followed by 10mL of toluene. The mixture was stirred at reflux for 4 hours while under argon. The resulting precipitated product was filtered, followed by recrystallization from hot methanol, giving 91.1mg (25% yield) of **360A Precursor** as a light tan solid. mp = 303.5-304.0°C. ^{1}H NMR (400MHz, DMF) δ 11.4 (s, 2H), 9.5 (d, 2H), 9.1 (d, 2H), 8.6 (d, 2H), 8.5 (t, 1H), 8.1 (d, 4H), 7.75 (t, 2H), 7.66 (t, 2H), MS (ESI) m/z = (M+1, 100%). IR (KBr) 3236.64, 1682.33, 1544.63, 1491.91, 1468.79, 1448.20, 1370.13, 1209.58, 782.81, 746.48, 683.18, 650.49 cm^{-1}.

3,3'-((pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) (360A) (2)**.** To 53.5mg (1eq, 0.13mmol) of **360A Precursor** in 6mL of 1:1 acetone:DMF was added 0.16mL (20eq, 2.6mmol) of methyl iodide. The solution was stirred at room temperature for 4 days, shielded from light. The bright yellow precipitate was filtered and washed three times with cold methanol, giving 53.7mg of **360A** (60% yield) as the bright yellow diiodide salt. mp = 267.7°C (decomposition). ^{1}H NMR (400MHz, DMF) δ 10.1 (s, 2H), 9.7 (s, 2H), 8.5-8.6 (m, 6H), 8.2 (t,2H), 8.1 (t, 2H), 4.8 (s, 6). IR (KBr)

3320.61, 3019.93, 1698.79, 1544.8, 1521.11, 1447.09, 1420.08, 1376.04, 1237.84, 1131.42, 766.54, 635.48 cm^{-1}.

Dimethyl 4-oxo-1,4-dihydropyridine-2,6-dicarboxylate (2) (3)**.** Distilled methanol that had been stored under argon over 4Å molecular sieves (3.1mL, 31eq, 77.5mmol) was added to a oven-dried flask under argon and the flask was then placed in an ice bath. Slowly, 1.1mL (6.2eq, 15.5mmol) of thionyl chloride was added. The solution was allowed to stir for 5-10min before 500mg (1eq, 2.5mmol) of chelidamic acid was added under increased argon flow. The flask was outfitted with an oven-dried condenser. The mixture was then stirred for 72 hours, under argon, allowing the ice bath to slowly warm to room temperature. The faintly yellow solution was then diluted approximately 2X with methanol, transferred to a larger flask, and the solvent removed under reduced pressure to give a white solid residue. The flask was then placed in an ice bath for 15 min before 3 mL of chilled distilled water was added with swirling, followed by 0.75 mL of chilled 10% sodium carbonate solution and 0.75mL of chilled 50% aqueous methanol. After swirling, the mixture was allowed to stand in the ice bath for 20min before being filtered under reduced pressure and washed with 3mL, 3mL, and 1mL portions of chilled 50% aqueous methanol, giving 500mg of crude white product. The crude product was adsorbed to 1g of silica and purified through chromatography on a silica plug (about 5-6 g SiO_2, EtOAc as eluent), giving 417mg of purified product as a white solid (79% yield). ^{1}H NMR (400MHz, $CDCl_3$) δ 10.0 (s, 1H), 7.5 (s, 2H), 4.0 (s, 6H) (matches lit.[21]). ^{13}C NMR (100MHZ, $CDCl_3$) δ 172.8, 163.3, 144.1, 117.7, 53.3 (matches lit.[21]).

First Generation Mitsunobu Coupling General Procedure: Dimethyl 4-(pent-4-yn-1-yloxy)pyridine-2,6-dicarboxylate (3a). 300mg (1eq, 1.42mmol) of **7** was suspended in 11mL of freshly distilled THF under argon. 745mg (2eq, 2.84mmol) of

triphenylphosphine was then added, followed by 198µL (1.5eq, 2.13mmol) of 4-pentyn-1-ol. The flask was then placed in an ice bath and stirred for 10 minutes. 391µL (1.4eq, 1.99mmol) of diisopropylazodicarboxylate (DIAD) was then added drop-wise. The reaction was stirred, allowing it to return to room temperature, for 72hr at which point the reaction appeared done by TLC (EtOAc, KMnO4 stain). The solvent was removed under reduced pressure, giving a viscous oil that was then subjected to high vacuum for 1 hour before being dissolved in the minimum amount of ethyl acetate. After purification by flash chromatography (50% EtOAc in hexanes), 350mg (89% yield) of the product was obtained as a white powder. mp = 107.6-108.6ºC. ^{1}H NMR (400MHz, CDCl3) δ 7.8 (s, 2H), 4.2 (t, 2H, J=6.1Hz), 4.0 (s, 6H), 2.4 (t of d, 2H, J=6.8Hz), 2.02 (p, 2H, J=6.4Hz), 1.96 (t, 1H, J=2.7Hz). ^{13}C NMR (100MHZ, $CDCl_3$) δ 166.8, 165.1, 149.7, 114.5, 82.5, 69.5, 67.1, 53.2, 27.5, 14.9. MS (ESI) *m/z* = 300.1 ($(M+Na)^+$, 100%), 278.1 ($(M+H)^+$, 25%). HRMS calc for $C_{14}H_{15}NO_5Na^+$ 300.08424, found 300.08404. IR (KBr) 3270.68, 2964.05, 1727.91, 1604.48, 1444.42, 1371.14, 1270.86, 1112.73, 1045.23, 1008.59, 883.24, 788.74, 705.82, 592.04 cm^{-1}.

Dimethyl 4-(hex-5-yn-1-yloxy)pyridine-2,6-dicarboxylate (3b). Following the general procedure above but employing 5-hexyn-1-ol afforded 180mg (87% yield) of **3b** after flash chromatography as a white solid. ^{1}H NMR (400MHz, $CDCl_3$) δ 7.8 (s, 2H), 4.1 (t, J=6.3Hz, 2H), 4.0 (s, 6H), 2.3 (t of d, J=7.0Hz, 2H), 1.9-2.0 (m, 3H), 1.7 (p, J=7.2Hz, 2H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 166.9, 165.1, 149.7, 114.5, 83.5, 69.0, 68.4, 53.2, 27.6, 24.6, 18.0. MS (ESI) m/z = 605.2 ($(2M+Na)^+$, 100%), 314.1 ($(M+Na)^+$, 25%), 292.1 ($(M+H)^+$,10%). HRMS calc for $C_{15}H_{18}NO_5^+$ 292.11790, found 292.11800.

Dimethyl 4-(hept-6-yn-1-yloxy)pyridine-2,6-dicarboxylate (3c). Following the general procedure above but employing 6-heptyn-1-ol afforded 164mg (87% yield) of **3c** after flash chromatography as a white solid. ^{1}H NMR (400MHz, $CDCl_3$) δ 7.8 (s, 2H),

4.11 (t, J=6.4Hz,2H), 3.96 (s, 6H), 2.20 (p, J=2.7Hz, 2H), 1.9 (t, J=2.7Hz, 1H), 1.8 (p, J=6.8Hz, 2H), 1.6 (m, 4H). ^{13}C NMR (100MHZ, CDCl3) δ 167.0,165.1, 149.6, 114.4, 84.0, 68.8, 68.6, 53.2, 28.2, 27.9, 24.9, 18.2. MS (ESI) *m/z* = 633.2 ($(2M+Na)^+$, 100%), 328.1 ($(M+Na)^+$, 33%), 306.1 ($(M+H)^+$, 15%). HRMS calc for $C_{16}H_{19}NO_5Na^+$ = 328.11550, found 328.11580.

Dimethyl 4-(non-8-yn-1-yloxy)pyridine-2,6-dicarboxylate (3d). Following the general procedure above but employing 7-octyn-1-ol afforded 221mg (quantitative yield) of 3d after flash chromatography as a white powder. ^{1}H NMR (400MHZ, $CDCl_3$) δ 7.7 (s, 2H), 4.1 (t, J=6.3Hz, 2H), 4.0 (s, 6H), 2.1 (t, J=6.9Hz, 2H), 1.9 (t, J=2.6Hz, 1H), 1.8 (p, J=7.0Hz, 2H), 1.3-1.5 (m, 8H). ^{13}C NMR (100MHZ, CDCl3) δ 167.0, 165.1, 149.6, 114.4, 84.4, 68.9, 68.2, 53.1, 28.60, 28.55, 28.4, 28.2, 25.6, 18.3. MS (CI) *m/z* 334 ($(M+H)^+$,100%). HRMS for $C_{18}H_{25}NO_5^+$ 334.1654, found 134.1655.

Trimethyl Aluminum Amidation General Procedure: 4-(pent-4-yn-1-yloxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (4a). 320mg (1eq, 1.04mmol) of **3a** and 357mg (2.4eq, 2.5mmol) of 3-aminoquinoline were dissolved in 15mL distilled 1,2-dichloroethane under argon. 4.5mL (4.2eq, 4.5mmol) of 1.0M trimethylaluminum in heptane was then added under increased argon flow resulting in a yellow solution. The flask was then fitted with an oven-dried condenser and placed in an oil bath at 94°C and stirred at reflux for 90min at which point the reaction appeared complete by TLC (EtOAc). The dark red solution was cooled to room temperature and quenched with 3mL of methanol, resulting in the formation of a yellow gel. The gel was diluted and partially dissolved with additional methanol and chloroform, silica gel added, and the solvent removed under reduced pressure. The silica-adsorbed residue was partially purified through a silica column plug (EtOAc). Combining the fractions from the column gave 564mg of the crude product after removal of the solvent under reduced pressure. The

crude product was then suspended in 20mL of methanol and stirred for 2 hours before being filtered under reduced pressure and washed three times with methanol giving 487mg of the product as a yellow solid. The filtrate was concentrated under reduced pressure, suspended in 3-5mL of methanol and filtered again after 30min, giving an additional 8mg of the product as a yellow solid. The combined crops gave 495mg (89% yield) of the product. mp = 235°C. ^{1}H NMR (400MHz, $CDCl_3$, 4% MeOD) δ 11.1 (s, 0.5H), 9.02-9.04 (m, 4H), 7.971 (s, 2H), 7.969 (d, J=8.0Hz, 2H), 7.8 (d, J=8.1Hz, 2H), 7.6 (t, J=7.6Hz, 2H), 7.5 (t, J=7.2Hz, 2H), 4.3 (t, J=6.1Hz, 2H), 2.4 (t of d, J=6.9Hz, 2H), 2.1 (p, J=6.5Hz, 2H), 2.0 (t, J=2.6Hz, 1H). ^{13}C (100MHZ, $CDCl_3$, 4% MeOD) δ 167.8, 162.8, 150.7, 144.6, 144.5, 131.6, 128.7, 128.2, 127.9, 127.3, 125.6, 125.5, 112.1, 82.6, 69.5, 67.2, 27.5, 14.9. MS (ESI) *m/z* = 502.3 (M+H, 100%); HRMS calc for $C_{30}H_{24}N_5O_3^+$ 502.18737, found 502.1873. IR (KBr) 3126.18, 1675.34, 1606.65, 1543.35, 1492.28, 1376.56, 1340.15, 1225.99, 1045.45, 902.46, 750.69 cm^{-1}.

4-(hex-5-yn-1-yloxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (4b). Following the general procedure above but employing **3b** afforded 146mg (64% yield) of **4b** as a yellow solid. mp = 204°C. ^{1}H NMR (400MHz, $CDCl_3$ (2% MeOD)) δ 10.9 (s, 1H), 9.0 (s, 2H), 8.9 (s, 2H), 7.92 (d, J=8.4Hz, 2H), 7.88 (s, 2H), 7.7 (d, J=8.1Hz, 2H), 7.6 (t, J= 7.5Hz, 2H), 7.5 (t, J=7.4Hz, 2H), 4.2 (t, J=6.3Hz, 2H), 2.3 (t, J=6.9Hz, 2H), 2.00 (t, J=2.6Hz, 1H), 1.99 (p, J=6.3Hz, 2H), 1.7 (p, J=7.1Hz, 2H). ^{13}C NMR (100MHz, $CDCl_3$ (2% MeOD)) δ 167.8, 162.7, 162.6, 150.6, 144.64, 144.59, 131.5, 131.4, 128.6, 128.3, 128.1, 127.8, 127.3, 125.4, 125.3, 112.0, 83.6, 69.0, 68.5, 27.7, 24.7, 18.0. MS (ESI) *m/z* = 516.2 ($M+H^+$, 100%), 252.2 (M^{2+}, 70%). HRMS calc for $C_{31}H_{26}N_5O_3^+$ 516.20302, found 516.20327. IR (KBr) 3293.32, 2941.37, 1667.96, 1604.95, 1537.16, 1491.22, 1468.93, 1423.99, 1374.2, 1337.97, 1278.61, 1226.4, 1176.96, 1144.2, 1098.44, 1034.38, 899.15, 780.04, 748.42, 661.48cm^{-1}.

4-(hept-6-yn-1-yloxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (4c). Following the general procedure above but employing **3c** afforded 104.6mg (37% yield) of **4c** as a pale yellow solid. mp = 199.8-202.1°C (decomp). ^{1}H NMR (400MHz, DMF) δ 11.4 (s, 2H), 9.5 (d, J=2.5Hz, 2H), 9.1 (d, J=2.4Hz, 2H), 8.1 (t, J=9.3Hz, 4H), 7.8 (t, J=7.6Hz, 2H), 7.7 (t, J=7.5Hz, 2H), 4.4 (t, J=6.5Hz, 2H), 2.8 (t, J=2.6Hz, 1H), 2.3 (p, J=4.0Hz, 2H), 1.9 (p, J=7.0Hz, 2H), 1.7 (m, 4H). ^{13}C NMR (100MHZ, DMF) δ 168.8, 163.0, 151.6, 146.4, 145.9, 133.0, 129.6, 128.9, 128.8, 128.6, 127.8, 124.4, 112.1, 84.9, 70.9, 69.7, 28.8, 25.6, 18.5. MS (ESI) *m/z* = 530.22(M+H^{+}, 100%). HRMS calc for $C_{32}H_{28}N_5O_3^{+}$ 530.21867, found 530.21987. IR (KBr) = 3354.65, 3199.37, 1686.04, 1637.74, 1609.93, 1539.93, 1492.12, 1399.91, 1375.81, 1337.48, 1207.02, 1175.14, 1145.52, 1028.99, 901.20, 782.60, 752.33, 736.65, 656.88 cm^{-1}.

4-(non-8-yn-1-yloxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (4d). Following the general procedure above but employing **3d** afforded 134mg (36% yield) of 4d as pale yellow crystals. mp = 160.7-162.4°C (decomp). ^{1}H NMR (400MHZ, CDCl3) δ 10.2 (s, 2H), 8.9 (d, J=2.5Hz, 2H), 8.5 (d, J=2.3Hz, 2H), 7.9 (d, J=8.3Hz, 2H), 7.7 (s, 2H), 7.53 (t, J=8.1Hz, 2H), 7.50 (t, J=7.7Hz, 2H), 7.4 (t, J=7.5Hz, 2H), 4.0 (t, J=6.6Hz, 2H), 2.2 (t of d, J=7.0Hz, 2H), 2.0 (t, J=2.6Hz, 1H), 1.8 (p, J=6.7Hz, 2H), 1.3-1.6 (m, 8H).^{13}C NMR (100MHZ, CDCl3) δ 167.9, 162.3, 150.2, 145.2, 144.9, 130.8, 128.71, 128.67, 127.8, 127.6, 127.2, 125.2, 111.8, 84.6, 69.2, 68.3, 28.8, 28.7, 28.6, 28.3, 25.7, 18.44, 18.38. MS (ESI) *m/z* = 558.25 (M+H^{+}, 100%). HRMS calc for $C_{34}H_{32}N_5O_3^{+}$ 558.24997, found 558.25046. IR (KBr) 3299.15, 2933.39, 2373.65, 2344.64, 1870.59, 1846.21, 1793.81, 1773.91, 1751.61, 1735.65, 1718.92, 1700.86, 1685.47, 1664.83, 1637.67, 1607.89, 1560.25, 1542.40, 1490.85, 1458.79, 1421.72, 1375.18, 1399.93, 1208.00, 1143.95, 1097.50, 1015.73, 897.82, 778.95, 746.47, 669.71, 473.38 cm^{-1}.

Azide Coupling General Procedure: 2-(2-azidoethyl)-1H-benzo[de]isoquinoline-1,3(2H)-dione (5). 102mg (1eq, 0.34mmol) of 2-(2-bromoethyl)-1H-benzo[de]isoquinoline-1,3(2H)-dione was dissolved in 5.5mL of DMF and placed under argon. 64mg (3eq, 0.99mmol) of sodium azide was then added and the flask heated with stirring in a 60°C oil bath for 24 hours, at which point the reaction appeared complete by TLC (1:1 Hexanes: Chloroform). After cooling to room temperature, the solvent was removed under reduced pressure and heating in a 60°C water bath. The resulting crude solid was dissolved/partitioned with ethyl acetate (40mL) and water (25mL). The organic phase was then washed twice more with water (15mL ea.) and once with brine (30mL) before being dried over sodium sulfate. Removal of the solvent under reduced pressure gave 84mg (93% yield) **5** as a light tan/yellow solid. ^{1}H NMR (400MHz, $CHCl_3$) δ 8.6 (d, J=7.3Hz, 2H), 8.2 (d, J=8.4Hz, 2H), 7.7 (t, J=7.8Hz, 2H), 4.4 (t, J=6.3Hz, 2H), 3.6 (t, J=6.3Hz, 2H). ^{13}C NMR (100MHz, $CHCl_3$) δ 164.2, 134.2, 131.5, 131.4, 128.1, 126.9, 122.2, 48.8, 38.7

(4-(azidomethyl)phenyl)(phenyl)methanone (6). Following the general procedure above, but instead using (4-(bromomethyl)phenyl)(phenyl)methanone afforded 59mg (quantitative) of **6** as a yellow oil. ^{1}H NMR (400MHz, $CDCl_3$) δ 7.77-7.81 (m, 4H), 7.6 (t of t, J=7.4Hz, 1H), 7.4-7.5 (m, 4H), 4.4 (s, 2H). ^{13}C NMR (100MHz, $CDCl_3$) δ 196.1, 139.9, 137.35, 137.32, 132.5, 130.5, 129.9, 128.3, 127.8, 54.2.

3-azidopropan-1-ol (misc_3). Following the general procedure above, but instead using 3-chloro-1-hydroxy propane, afforded 1.38g of clear liquid (determined to be 62wt% of **misc_3** in DMF by NMR). ^{1}H NMR (400MHz, CDCl3) δ 3.7 (t, J=6Hz, 2H), 3.4 (t, J=6.6Hz, 2H), 2.3 (s, 1H), 1.8 (p, J=6.4Hz, 2H). ^{13}C NMR (100MHZ, CDCl3) δ 59.6, 48.4, (3rd peak believed to overlap with DMF peak at 36.6)

3-azidopropyl 9,10-dioxo-9,10-dihydroanthracene-2-carboxylate (7). 100mg (1eq, 0.37mmol) of 9,10-dioxo-9,10-dihydroanthracene-2-carbonyl chloride was dissolved in 3mL freshly distilled dichloromethane in a flask under argon. This flask was then cooled to 0°C. In a separate flask, 150mg (2.55eq, 0.92mmol) of 62wt% **misc_3** in DMF was dissolved in 1mL of distilled dichloromethane under argon. 52μL (1eq, 0.37mmol) of freshly distilled triethylamine was then added. The solution containing the alcohol and the amine was then added slowly under argon to the flask containing the acyl chloride at 0°C. The reaction was then allowed to return to room temperature and was stirred for 19 hours when it appeared complete by TLC (75% ethyl acetate in hexanes). After flash chromatography (25% ethyl acetate gradually up to 50% ethyl acetate in hexanes), 74mg of **7** as a light yellow solid was obtained (approximately 80% purity, 45% yield). mp = <125°C. ^{1}H NMR (400MHz, CDCl3) δ 8.9 (s, 1H), 8.3-8.4 (m, 4H), 7.80-7.83 (m, 2H), 4.5 (t, J=6.0Hz, 2H), 3.5 (t, J= 6.6Hz, 2H), 2.1 (p, J=6.4Hz, 2H), MS (ESI) m/z = 358.1 $(M+Na^+)^+$, 100%); HRMS calc for $C_{18}H_{13}N_3NaO_4^+$ 358.07983, found 358.07973. IR (KBr) 3133.59, 2105.89, 1727.91, 1679.70, 1590.99, 1402.00, 1332.57, 1276.65, 1249.65, 1164.79, 1037.52, 939.16, 802.24, 703.89 cm^{-1}.

Copper-catylized coupling General Procedure: 4-(3-(1-(2-(1,3-dioxo-1H-benzo[de]isoquinolin-2(3H)-yl)ethyl)-1H-1,2,3-triazol-4-yl)propoxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (8a). The alkyne **3a** 20 mg (1eq, 0.04mmol) and 12.7 mg (1.2eq, 0.048mmol) of azide **5** (4) were dissolved in 3mL DMF and placed under argon. A solution of 0.1M copper triflate in water (440 μL, 0.044mmol) was added, resulting in a dark green solution. A freshly prepared solution of 0.1M sodium ascorbate in water (160 μL, 0.016mmol) was added, and the flask was placed in an oil bath at 25°C and stirred for 68 hours until the reaction appeared complete by TLC (EtOAc). The DMF was removed under high vacuum and the residue partitioned

between about 40 mL of chloroform and 40 mL of water. This mixture was stirred vigorously for 24 hours before being transferred to a separation funnel. The aqueous layer was extracted twice more with smaller volumes of chloroform (about 10-20mL) and the organic layers were combined, and the solvent removed. The residue was dissolved in a mixture of methanol and chloroform and absorbed to silica by removal of the solvent under reduced pressure. A mini column was then run (EtOAc up to 10% MeOH in EtOAc), giving 23mg (76% yield) of **8a** as a pale yellow solid. ^{1}H NMR (400MHz, $CDCl_3$ (4% MeOD)) δ 11.1 (s, exchanged), 9.1 (s, 4H), 8.5 (d, J=7.6Hz, 2H), 8.2 (d, J=7.8Hz, 2H), 8.0 (s, 4H), 7.9 (d, J=7.7Hz, 2H), 7.6-7.7 (m, 4H), 7.5-7.6 (m, 3H), 4.72 (s, 2H), 4.62 (s, 2H), 4.2 (s, 2H), 2.9 (t, J=7.3Hz, 2H), 2.2 (t, J=6.1Hz, 2H). MS (ESI) *m/z* = 790.26 ($(M+Na)^+$,100%), 768.28 ($(M+H)^+$, 25%) HRMS calc for $(C_{44}H_{33}N_9O_5Na)^+$ 790.24970, found 790.25100.

4-(3-(1-(4-benzoylbenzyl)-1H-1,2,3-triazol-4-yl)propoxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (9a). Following the general procedure above but employing azide **6** afforded 5.5mg (52% yield) of the **9a** as a pale yellow solid. ^{1}H NMR (400MHz, $CDCl_3$ (1.5% MeOD)) δ 10.8 (s, exchanged amide), 9.1 (s, 2H), 9.0 (s, 2H), 7.99 (d, J=7.2Hz, 2H), 7.91 (s, 2H), 7.8 (d, J=8.6Hz, 2H), 7.68-7.72 (m, 4H), 7.6 (t, J= 7.3Hz, 2H), 7.52 (t, J=7.3Hz, 2H), 7.49 (t, J= 1.3Hz, 1H), 7.4 (t, J=7.6Hz, 2H), 7.3 (t, J=8.3Hz, 2H), 5.6 (s, 2H), 4.2 (t, J=6.2Hz, 2H), 2.9 (t, J=7.3Hz, 2H), 2.3 (p, J=6.6Hz, 2H). MS (ESI) *m/z* = 739.3 ($(M+H)^+$, 50%), 761.3 ($(M+Na)^+$, 40%). HRMS calc for $C_{44}H_{35}N_8O_4^+$ = 739.27760, found 739.27870.

3-(4-(3-((2,6-bis(quinolin-3-ylcarbamoyl)pyridin-4-yl)oxy)propyl)-1H-1,2,3-triazol-1-yl)propyl 9,10-dioxo-9,10-dihydroanthracene-2-carboxylate (10a). Following the general procedure above but employing azide **7** afforded 7 mg of **10a** (21% yield). ^{1}H NMR (400MHz, $CDCl_3$) δ 10.3 (s, 2H), 9.1 (s, 2H), 8.8 (s, 1H), 8.7 (s, 2H),

8.2-8.3 (m, 2H), 8.16-8.18 (m, 1H), 7.93 (d, J=8.0Hz, 2H), 7.8 (s, 2H), 7.71-7.74 (m, 2H), 7.69 (d, J=8.9Hz, 2H), 7.6 (t, J=7.6Hz, 2H), 7.45-7.49 (m, 3H), 4.5 (t, J=6.7Hz, 2H), 4.3 (t, J=6.7Hz, 2H), 4.2 (t, J=5.8Hz, 2H), 3.0 (t, J=7.2Hz, 2H), 2.4 (p, J=6.4Hz, 2H), 2.3 (p, J=6.7Hz, 2H). MS (ESI) m/*z* = 859.3 ($(M+Na)^+$, 50%). HRMS calc for $C_{48}H_{36}N_8O_7Na^+$ 859.25990, found 859.26200.

4-(4-(1-(2-(1,3-dioxo-1H-benzo[de]isoquinolin-2(3H)-yl)ethyl)-1H-1,2,3-triazol-4-yl)butoxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (8b). Following the general procedure above but employing alkyne **4b** afforded 13.4mg (44% yield) of **8b**. ^1H NMR (400MHz, $CDCl_3$ (20% MeOD)) δ 9.0 (s, 2H), 8.9 (s, 2H), 8.4 (d, J=7.0Hz, 2H), 8.1 (d, J=8.0Hz, 2H), 7.91 (d, J=8.5Hz, 2H), 7.89 (s, 2H), 7.8 (d, J=8.3Hz, 2H), 7.64 (t, J=7.8Hz, 2H), 7.58 (t, J=7.6Hz, 2H), 7.48 (t, J=7.6Hz, 2H), 7.45 (s, 1H), 4.6 (t, J=6.3Hz, 2H), 4.5 (t, J=6.0Hz, 2H), 4.1 (t, J=6.3Hz, 2H), 2.7 (t, J=7.5Hz, 2H), 1.8 (m, 4H). MS (ESI) *m/z* =804.3 ($(M+Na)^+$,100%). HRMS calc for $C_{45}H_{35}N_9O_5Na^+$ 804.26530, found 804.26690.

4-(4-(1-(4-benzoylbenzyl)-1H-1,2,3-triazol-4-yl)butoxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (9b). Following the general procedure above but employing alkyne **4b** and azide **6** afforded 27.9mg (95% yield) of **9b**. ^1H NMR (400MHz, $CDCl_3$) δ 10.8 (s, exchanged), 9.1 (s, 2H), 9.0 (s, 2H), 8.01 (d, J=8.8Hz, 2H), 7.93 (s, 2H), 7.8 (d, J=8.4Hz, 2H), 7.71-7.73 (m, 4H), 7.6 (t, J=7.5Hz, 2H), 7.5-7.6 (m, 3H), 7.42 (t, J=7.7Hz, 2H), 7.35 (s, 1H), 7.3 (d, J=8.0Hz, 2H), 5.6 (s, 2H), 4.2 (t, J=5.4Hz, 2H), 2.8 (t, J=6.8Hz, 2H), 1.90-1.91 (m, 4H). MS (ESI) *m/z* = 775.28 ($(M+Na)^+$,100%), 753.29 ($(M+H)^+$, 33%). HRMS calc for $(C_{45}H_{36}N_8O_4Na)^+$ 775.27520, found 775.27540.

4-((5-(1-(4-benzoylbenzyl)-1H-1,2,3-triazol-4-yl)pentyl)oxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (9c). Following the general procedure

above but employing alkyne **4c** and azide **6** afforded 34.6mg (88% yield) of **9c** as an off-white solid mp = 125°C. ^{1}H NMR (400MHz, $CDCl_3$) δ 10.2 (s, 2H), 9.1 (d, J =2.5Hz, 2H), 8.8 (d, J=2.3Hz, 2H), 8.0 (d, J= 9.3Hz, 2H), 7.9 (s, 2H), 7.8 (d, J= 8.2Hz, 2H), 7.69-7.72 (m, 4H), 7.6 (t, J= 7.7Hz, 2H), 7.5-7.6 (m, 3H), 7.4 (t, J=7.0Hz, 2H), 7.26-7.28 (m, 3H), 5.6 (s, 2H), 4.1 (t, J=6.4Hz, 2H), 2.8 (t, J=7.5Hz, 2H), 1.9 (p, J=6.9Hz, 2H), 1.8 (p, J=7.6Hz, 2H), 1.5 (p, J=7.6Hz, 2H). IR (KBr) 3128.25, 2935.02, 2385.38, 1794.12, 1774.05, 1735.74, 1719.09, 1701.17, 1686.19, 1655.31, 1637.99, 1607.92, 1578.00, 1560.39, 1543.06, 1509.50, 1490.94, 1459.13, 1400.02, 1099.73, 750.87, 612.45 cm^{-1}. MS (ESI) *m/z* = 789.29 ($(M+Na)^+$, 100%), 767.31 ($(M+H)^+$, 70%). HRMS calc for $(C_{46}H_{38}N_8O_4Na)^+$ 789.29080, found 789.29120.

4-((7-(1-(4-benzoylbenzyl)-1H-1,2,3-triazol-4-yl)heptyl)oxy)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (9d). Following the general procedure above but employing alkyne **4d** and azide **6** afforded 20.9mg (49% yield) of **9d** as a white powder. ^{1}H NMR (400MHZ, $CDCl_3$) δ 10.1 (s, 2H), 9.1 (d, J=2.9Hz, 2H), 8.8 (d, J=2.6Hz, 2H), 8.0 (d, J=8.6Hz, 2H), 7.9 (s, 2H), 7.7-7.8 (m, 6H), 7.62 (t, J=7.7Hz, 2H), 7.5-7.56 (m, 3H), 7.4 (t, J=7.2Hz, 2H), 7.2-7.3 (m, 3H), 5.5 (s, 2H), 4.2 (t, J=6.4Hz, 2H), 2.7 (t, J=7.7Hz, 2H), 1.8 (p, J=7.0Hz, 2H), 1.7 (p, J=7.5Hz, 2H), 1.5 (p, J=6.4Hz, 2H), 1.4-1.47 (m, 4H). MS (ESI) *m/z* = 795.3 ($(M+H)^+$, 50%), 817.3 ($(M+Na)^+$, 100%), 1612.7 ($(2M+Na)^+$, 30%). HRMS calc for $C_{48}H_{42}N_8O_4Na^+$ = 817.32210, found 817.32210.

3-(4-(4-((2,6-bis(quinolin-3-ylcarbamoyl)pyridin-4-yl)oxy)butyl)-1H-1,2,3-triazol-1-yl)propyl 9,10-dioxo-9,10-dihydroanthracene-2-carboxylate (10b). Following the general procedure above but employing alkyne **4b** and azide **7** afforded 9.7mg of **10b** as a white solid (29% yield). ^{1}H NMR (400MHz, $CDCl_3$ (5% MeOD)) δ 9.0 (s, 4H), 8.8 (s, 1H), 8.2-8.4 (m, 4H), 7.99 (d, J= 8.2Hz, 2H), 7.95 (s, 2H), 7.8 (d, J=8.4Hz, 2H), 7.75-7.77 (m, 2H), 7.6 (t, J= 7.3Hz, 2H), 7.54 (t, J=7.5Hz, 2H), 7.45 (s,

1H), 4.5 (t, J= 7.0Hz, 2H), 4.4 (t, J=6.1Hz, 2H), 4.2 (t, J=6.4Hz, 2H), 2.8 (t, J=6.7Hz, 2H), 2.4 (p, J=6.4Hz, 2H), 1.9 (m, 4H). MS (ES) m/z = 873.3 ($(M+Na)^+$, 50%). HRMS calc for $C_{49}H_{38}N_8O_7Na^+$ 873.27560, found 873.27450.

First Generation Methylation General Procedure: 3,3'-((4-(4-(1-(2-(1,3-dioxo-1H-benzo[de]isoquinolin-2(3H)-yl)ethyl)-3-methyl-1H-1,2,3-triazol-3-ium-4-yl)butoxy)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) trifluoromethanesulfonate (11b). 13.4mg (1eq, 0.016mmol) of **8b** was dissolved in 3mL of 3.8% methanol in chloroform and placed under argon. The flask was placed in an ice bath and 10μL (5.6eq, 0.09mmol) of methyl triflate was added slowly. The solution was then stirred for 4 hours, allowing the bath to return to room temperature. A significant amount of precipitate began to form after 2 hours. This precipitate was isolated by filtration under reduced pressure and washed twice with small (1mL) portions of chloroform, giving 13.4mg (66% yield). ^{1}H NMR (400MHz, d_6-DMSO) δ 11.9 (s, 2H), 10.0 (s, 2H), 9.6 (s, 2H), 8.9-9.0 (m, 2H), 8.4-8.6 (m, 5H), 8.3 (t, J=8.0Hz, 2H), 8.05-8.11 (m, 2H), 7.92 (s, 2H), 7.86 (t, J=7.5Hz, 2H), 7.7-7.8 (m, 2H), 5.0 (t, J=5.5Hz, 2H), 4.8 (s, 6H), 4.6 (t, J= 5.7Hz, 2H), 4.3 (s, 2H), 4.2 (s, 3H), 3.0 (t, J=6.6Hz, 2H), 1.8 (m, 4H). MS (MALDI) *m/z* = (1124 (M-OTf),100%); HRMS calc for $C_{50}H_{44}N_9O_{11}F_6S_2^+$ 1124.25004, found 1124.24914.

3,3'-((4-(3-(1-(2-(1,3-dioxo-1H-benzo[de]isoquinolin-2(3H)-yl)ethyl)-3-methyl-1H-1,2,3-triazol-3-ium-4-yl)propoxy)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) trifluoromethanesulfonate (11a). Following the general procedure above but starting with the triazole **8a** afforded 3.5mg (21% yield) of **11a** ^{1}H NMR (400MHZ, d_6-DMSO) δ 10.1 (s, 2H), 9.6 (s, 2H), 9.0 (s, 2H), 8.6 (t, J=8.1Hz, 2H), 8.4 (m, 3H), 8.3 (t, J=8.0Hz, 2H), 8.07-8.11 (m, 3H), 7.93-7.97 (m, 2H), 7.7-7.8 (m, 4H), 5.0 (s, 2H), 4.8 (s, 6H), 4.6 (s, 2H), 4.4 (s, 2H), 4.2 (s,

3H), 3.1 (t, J=7.2Hz, 2H), 2.2 (s, 2H). MS (MALDI) *m/z* = 1110.1 $(M\text{-}OTf)^+$, 8%). HRMS calc for $C_{49}H_{42}N_9O_{11}F_6S_2^+$ 1110.23439, found 1110.2335.

3,3'-((4-(3-(1-(4-benzoylbenzyl)-3-methyl-1H-1,2,3-triazol-3-ium-4-yl)propoxy)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) trifluoromethanesulfonate (12a). Following the general procedure above but starting with the triazole **9a** afforded 8.7mg (50% yield) of **12a**. ^{1}H NMR (600MHz, d_6-DMSO) δ 11.8 (s, 2H), 10.0 (s, 2H), 9.6 (s, 2H), 9.0 (s, 1H), 8.6 (d, J=9.0Hz, 2H), 8.5 (d, J=7.8Hz, 2H), 8.3 (t, J=7.9Hz, 2H), 8.1 (t, J=7.8Hz, 2H), 8.0 (s, 2H), 7.8 (d, J=6.6Hz, 2H), 7.6-7.7 (m, 5H), 7.5 (t, J=6.8Hz, 2H), 6.0 (s, 2H), 4.8 (s, 6H), 4.5 (t, J=6.0Hz, 2H), 4.3 (s, 3H), 3.1 (t, J=7.8Hz, 2H), 2.3 (p, J=6.9Hz, 2H). MS (ESI) *m/z* = 261 (M^{3+}, 80%); HRMS calc for $C_{47}H_{43}N_8O_4^{3+}$ 261.11300, found 261.11310.

3,3'-((4-(4-(1-(4-benzoylbenzyl)-3-methyl-1H-1,2,3-triazol-3-ium-4-yl)butoxy)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) trifluoromethanesulfonate (12b). Following the general procedure above but starting with the triazole **9b** afforded 5.7mg (35% yield) of **12b** ^{1}H NMR (400MHz, d_6-DMSO) δ 10.0 (s, 2H), 9.6 (s, 2H), 8.9 (s, 2H), 8.6 (t, J=7.2Hz, 4H), 8.3 (t, J=8.0Hz, 2H), 8.1 (t, J=7.8Hz, 2H), 8.0 (s, 2H), 7.8 (d, J=8.2Hz, 2H), 7.62-7.7 (m, 4H), 7.56 (t, J=8.0Hz, 2H), 6.0 (s, 2H), 4.8 (s, 6H), 4.4 (m, 2H), 4.2 (s, 3H), 3.0 (t, J=7.0Hz, 2H), 1.9 (m, 4H). MS (MALDI) *m/z* = (1095 (M-OTf),100%); HRMS calc for $C_{50}H_{45}N_8O_{10}F_6S_2^+$ 1095.25988, found 1095.258.

3,3'-((4-((5-(1-(4-benzoylbenzyl)-3-methyl-1H-1,2,3-triazol-3-ium-4-yl)pentyl)oxy)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) trifluoromethanesulfonate (12c). Following the general procedure above but starting with the triazole **9c** afforded 10.1mg (53% yield) of **12c** as a white powder. mp =211°C. ^{1}H NMR (400MHZ, d_7-DMF) δ 11.9 (s, 2H), 10.3 (d, J=2.6Hz, 2H), 9.8 (s, 2H), 9.1 (s,

1H), 8.7 (d, J=9.2Hz, 2H), 8.6 (d, J=8.4Hz, 2H), 8.4 (s, 2H), 8.3 (t, J=8.0Hz, 2H), 8.2 (t, J=7.5Hz, 2H), 8.02-8.05 (m, 2H, +DMF), 7.9 (d, J=8.1Hz, 2H), 7.7-7.8 (m, 5H), 7.6 (t, J=7.6Hz, 2H), 6.2 (s, 2H), 5.0 (s, 6H), 4.5 (s, 3H), 3.1 (t, J=7.6Hz, 2H), 1.9-2.0 (m, 4H), 1.7 (p, J=7.7Hz, 2H). IR (KBr) = 3122.83, 2372.10, 2344.75, 1774.14, 1735.85, 1719.20, 1701.46, 1686.39, 1655.16, 1637.81, 1560.41, 1543.41, 1475.80, 1458.82, 1399.46, 1275.61, 1156.44, 1030.19, 637.63 cm^{-1}. MS (ESI) *m/z* = 270.5 (M^{3+}, 1.5%), 398.2 ((M-Me)$^{2+}$, 100%). HRMS calc for $C_{49}H_{47}N_8O_4^{3+}$ = 270.45679, found 270.45621. HRMS calc for $C_{48}H_{44}N_8O_4^{2+}$ 398.17373, found 398.17423. I

3,3'-((4-((7-(1-(4-benzoylbenzyl)-3-methyl-1H-1,2,3-triazol-3-ium-4-yl)heptyl)oxy)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) trifluoromethanesulfonate (12d). Following the general procedure above but starting with the triazole **9d** afforded 2.5mg (24% yield) of **12d** as a white powder. ^{1}H NMR (400MHZ, d_7-DMF) δ 12.0 (s, 2H), 10.3 (s, 2H), 9.8 (s, 2H), 8.7 (d, J=8.6Hz, 2H), 8.6 (d, J=8.6Hz, 2H), 8.3 (t, J=8.6Hz, 2H), 8.2 (t, J=8.1Hz, 2H), 8.1 (s, 2H), 7.9 (d, J=6.7Hz, 2H), 7.7-7.8 (m, 4H), 7.6 (t, J=7.7Hz, 2H), 6.2 (s, 2H), 5.0 (s, 6H), 4.4 (s, 3H), 3.1 (t, J=8.4Hz, 2H), 1.8-1.9 (m, 4H), 1.5-1.6 (m, 6H). MS (MALDI) *m/z* = 1137.0 ((M-OTf)$^+$,15%). HRMS calc for $C_{53}H_{51}N_8O_{10}F_6S_2$

N-(2-(dimethylamino)ethyl)-9,10-dioxo-9,10-dihydroanthracene-2-carboxamide (14). 100mg (1eq, 0.37mmol) of 9,10-dioxo-9,10-dihydroanthracene-2-carbonyl chloride was suspended in 4mL of dichloromethane. A solution of 56µL N,N-dimethylethylenediamine in 6mL dichloromethane was slowly added by addition funnel over 35 minutes. The mixture was then stirred at 35°C for 2 hours, when an additional 50µL of diamine were added drop wise and the mixture stirred for an additional hour when the reaction appeared to be complete by TLC (75% EtOAc in Hexanes). The precipitate was isolated by filtration under reduced pressure and was washed 3X with

dichloromethane, giving 46.7mg of starting material (by TLC). The filtrate was concentrated under reduced pressure and the residue partitioned between dichloromethane and water. The aqueous layer was extracted an additional two times with dichloromethane, and the combined organic layers washed once with 2.3M NaOH and once with brine before being dried over sodium sulfate. Removal of the solvent by rotary evaporation gave 54mg (46% yield) of **14** as a yellow solid. ^{1}H NMR (400MHZ, CDCl3) δ 8.5 (s, 1H), 8.2-8.3 (m, 4H), 7.6-7.7 (m, 2H), 7.2 (s, 1H), 3.5 (q, J=5.6Hz, 2H), 2.5 (t, J=6.0Hz, 2H), 2.3 (s, 6H). ^{13}C NMR (100MHZ, CDCl3) δ 182.43, 182.36, 165.6, 139.6, 134.9, 134.31, 134.28, 133.28, 133.25, 133.0, 127.7, 127.3 125.2, 57.6, 45.1, 37.5.

2-aminoquinolin-4-ol (16) (5). 3.56g (1.1eq, 50.86mmol) of freshly sublimed malononitrile was dissolved in 15.4mL of DMF. 7.1mL (1.1eq, 50.86mmol) of TEA was added with stirring and the flask placed in an oil bath at 55°C. A solution of 7.54g (1eq, 46.24mmol) of **1** in 31mL DMF was slowly added via an addition funnel over 90min. After addition, the oil bath temperature was increased to 60°C, and the solution was stirred for an additional 2 hours at which point the dark-brown/black solution was cooled to room temperature. After cooling, the contents of the reaction flask were poured into 375mL of ice-cold 1M HCl, resulting in a precipitate suspended in a bright yellow solution. The precipitate was isolated by filtration under reduced pressure and washed with a small amount (about 5-10mL) of ice-cold 1M HCl, giving 4.88g of crude intermediate as a brown solid. The crude intermediate was placed into a new flask and suspended in 120mL of 48% HBr. The flask was then heated at reflux in an oil bath at 140°C for 24 hours at which point a clear dark-brown solution had formed. After cooling to room temperature, the flask was placed in an ice bath for 25min. The resulting light-tan precipitate was isolated by filtration under reduced pressure and washed with a small amount (about 5mL) of ice-cold water. The precipitate was placed in another flask and

dissolved with heating in about 175mL water. The solution was made basic by addition of ammonium hydroxide, resulting in an off-white precipitate that was isolated by filtration under reduced pressure. The precipitate was washed, first with water and then IPA, giving 3.45g (47% yield) of the final product as an off-white solid. The ^{1}H NMR (400MHZ, MeOD) δ 8.1 (d, J=8.0Hz, 1H), 7.5 (t, J=8.0Hz, 1H), 7.3 (d, J=8.2Hz, 1H), 7.2 (t, J=7.6Hz, 1H), 5.7 (s, 1H). ^{13}C NMR (100MHZ, MeOD) δ 178.3, 157.0, 139.6, 132.3, 125.6, 123.73, 123.71, 117.4, 91.9.

4-(2-(dimethylamino)ethoxy)quinolin-2-amine (17) (6). 300mg (1eq, 1.87mmol) of **2** was placed under high vacuum (100um) for 2 hours at 100°C before being flushed with argon. 689mg (2eq, 3.74mmol) of triphenylphosphine was added, and the flask returned to high vacuum at room temperature for 1 hour before being refilled with argon. 60mL of freshly distilled THF was added, and the flask placed in an ice bath. 170µL (1.3eq, 2.431mmol) of N,N-dimethylethanolamine was added. After stirring for 10 minutes, 0.3mL (1.2eq, 2.24mmol) of diisopropylazodicarboxylate (DIAD) was added drop-wise over 15min. The flask was allowed to return to room temperature and was stirred under argon for 6 days until the reaction appeared complete by TLC (75% MeOH in EtOAc; 10% MeOH, 5% TEA in EtOAc). The solution was filtered to remove a small amount of precipitate (unrecoverable from filter) and then the filtrate was concentrated under reduced pressure and the residue purified on a silica column (40g, 10-20% MeOH, 5% TEA, in EtOAc), giving 330mg (76% yield) as a white solid. ^{1}H NMR (400MHZ, MeOD) δ 8.0 (d, J=8.2Hz, 1H), 7.4-7.5 (m, 2H), 7.2 (t, J=7.5Hz, 1H), 6.2 (s, 1H), 4.3 (t, J=5.4Hz, 2H), 2.9 (t, J=5.4Hz, 2H), 2.4 (s, 6H). ^{13}C NMR (100MHZ, MeOD) δ 163.8, 161.0, 148.9, 131.3, 124.9, 123.0, 122.7, 118.5, 91.4, 67.5, 58.6, 46.0.

dimethyl 4-methoxypyridine-2,6-dicarboxylate (22) (7). 100mg (1eq, 0.48mmol) of **5** was placed in an oven-dried flask and placed under high vacuum

(100µm) for 30min before backfilling with argon. This was dissolved in 3mL of acetone with heating. 82.7mg (1.2eq, 0.58mmol) of potassium carbonate was added, followed by 60µL (2eq, 0.96mmol) of methyl iodide. The flask was placed in an oil bath at 65°C and refluxed overnight. However, by morning the acetone had all evaporated. The residue was resuspended in acetone; TLC (25% hexanes in EtOAc) verified that the reaction was complete. The mixture was filtered over celite and washed with acetone. The filtrate was concentrated under reduced pressure before being dissolved in diethyl ether and washed with 1M HCl, saturated sodium bicarbonate, and brine in succession before being dried over anhydrous sodium sulfate. This gave 40.8mg (37.7% yield) of product as a white solid. ^{1}H NMR (400MHZ, CDCl3) δ 7.8 (s, 2H), 3.99 (s, 6H), 3.96 (s, 3H).

Trimethyl Aluminum Amidation General Proceduere: N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-(pent-4-yn-1-yloxy)pyridine-2,6-dicarboxamide (18a). 37mg (1eq, 0.136mmol) of dimethyl 4-(pent-4-yn-1-yloxy)pyridine-2,6-dicarboxylate (**3a**) and 70mg (2.2eq, 0.298mmol) of **17** were combined in an oven-dried flask and placed under high vacuum for 30min before being placed under argon. 3mL of distilled 1,2-dichloroethane (stored under argon over 3Å sieves) was then added with stirring, resulting in a faintly yellow solution. 0.57mL (4.2eq, 0.570mmol) of 1M trimethylaluminum in heptane was then added drop wise, giving a yellow solution. The flask was outfitted with an oven-dried condenser and the flask placed in an oil bath at 94°C. After refluxing for two hours, the reaction appeared complete by TLC (10% MeOH in $CHCl_3$) and the dark-red solution was cooled to room temperature before being quenched with 2mL methanol, giving an orange gel. Chloroform was added and the gel suspension was gravity filtered through a cotton plug and washed alternatively with methanol and chloroform four times. The filtrate was concentrated under reduced pressure and then redissolved in a minimal amount of

chloroform before being purified on a silica column (25g, 5% MeOH in CHCl3), giving 47mg (51% yield) of the product as a pale yellow solid. ^{1}H NMR (400MHZ, CDCl3) δ 11.2 (s, 2H), 8.13 (s, 2H), 8.09 (d, J=8.4Hz, 2H), 7.9 (s, 2H), 7.8 (d, J=8.3Hz, 2H), 7.6 (t, J=7.7Hz, 2H), 7.4 (t, J=7.5Hz, 2H), 4.4 (t, J=5.5Hz, 4H), 4.2 (t, J=6.3Hz, 2H), 2.9 (t, J=5.5Hz, 4H), 2.42 (t, J=5.6Hz, 2H), 2.41 (s, 12H), 2.04 (p, J=6.6Hz, 2H), 2.01 (t, J=2.6Hz, 1H). ^{13}C NMR (100MHZ, CDCl3) δ MS (ESI) = 676.3 ($(M+H)^{+}$, 100%), 338.7 ($(M+2H)^{2+}$, 25%), 698.3 ($(M+Na)^{+}$, 6%). HRMS (ESI) calc for $C38H41N7O5^{+}$ = 676.32420, found 676.32470.

N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-(hex-5-yn-1-yloxy)pyridine-2,6-dicarboxamide (18b). Following the general procedure above, except using **3b**, gave 54.8mg (45% yield) of **18b** as a pale yellow-white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.3 (s, 2H), 8.15 (s, 2H), 8.11 (d, J=7.8Hz, 2H), 7.9 (s, 2H), 7.8 (d, J=8.3Hz, 2H), 7.7 (t, J=7.6Hz, 2H), 7.4 (t, J=7.2Hz, 2H), 5.2 (s, 2H), 4.4 (t, J=5.5Hz, 4H), 4.1 (t, J=6.3Hz, 2H), 2.9 (t, J=5.5Hz, 4H), 2.4 (s, 12H), 2.3 (t, J=6.9Hz, 2H), 2.00 (t, J=2.6Hz, 1H), 1.97 (p, J=6.1Hz, 2H), 1.7 (p, J=7.3Hz, 2H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 168.1, 163.2, 162.3, 152.1, 150.4, 146.9, 130.3, 127.0, 124.3, 122.1, 119.7, 111.7, 93.9, 83.6, 69.0, 68.5, 67.2, 57.9, 46.1, 27.7, 24.7, 18.1. MS (ESI) = 690.3 ($(M+H)^{+}$,100%), 345.7 ($(M+2H)^{2+}$, 30%), 712.3 ($(M+Na)^{+}$,10%). HRMS (ESI) calc for $C39H43N7O5^{+}$ = 690.33980, found 690.34010.

N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-(hept-6-yn-1-yloxy)pyridine-2,6-dicarboxamide (18c). Following the general procedure above, except using **3c**, gave 43.2mg (45% yield) of **18c** as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.0 (s, 2H), 8.11 (s, 2H), 8.09 (d, J=9.1Hz, 2H), 7.9 (s, 2H), 7.8 (d, J=8.5Hz, 2H), 7.6 (t, J=7.6Hz, 2H), 7.4 (t, J=7.6Hz, 2H), 4.4 (t, J=5.4Hz, 4H), 4.1 (t, J=6.5Hz, 2H), 2.9 (t, J=5.5Hz, 4H), 2.4 (s, 12H), 2.2 (t, J=5.2Hz, 2H), 2.0 (t, J=2.7Hz, 1H), 1.8 (p,

J=6.7Hz, 2H), 1.56-1.60 (m, 4H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 168.0, 163.1, 162.2, 152.1, 150.2, 146.8, 130.2, 127.0, 124.2, 122.0, 119.5, 111.6, 93.8, 84.0, 68.8, 68.6, 67.1, 57.8, 46.0, 28.2, 28.0, 24.9, 18.3. MS (ESI) m/z = 704.35 ($(M+H)^+$, 100%). HRMS (ESI) calc for $(C40H46N7O5)^+$ =704.35550, found 704.35490.

N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-(non-8-yn-1-yloxy)pyridine-2,6-dicarboxamide (18d). Following the general procedure above, except using **3d**, gave 41.5mg (63.0% yield) of **18d** as a pale yellow oil. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.3 (s, 2H), 8.13 (s, 2H), 8.09 (d, J=8.3Hz, 2H), 7.9 (s, 2H), 7.8 (d, J=8.2Hz, 2H), 7.6 (t, J=7.6Hz, 2H), 7.4 (t, J=7.6Hz, 2H), 5.5 (s, 2H), 4.4 (t, J=5.6Hz, 4H), 4.1 (t, J=6.5Hz, 2H), 2.9 (t, J=5.5Hz, 4H), 2.4 (s, 12H), 2.2 (t, J=7.0Hz, 2H), 2.0 (t, J=2.6Hz, 1H), 1.8 (p, J=6.8Hz, 2H), 1.5 (p, J=6.8Hz, 2H), 1.4-1.5 (m, 6H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 168.1, 163.1, 162.3, 152.0, 150.3, 149.8, 130.3, 127.0, 124.3, 122.0, 119.6, 111.7, 93.8, 84.5, 69.0, 68.3, 67.1, 57.8, 46.0, 28.7, 28.6, 28.5, 28.3, 25.7, 18.3. MS (ESI) = 366.4 ($(M+2H)^{2+}$, 100%), 731.9 ($(M+H)^+$, 30%). HRMS (CI) calc for $C_{42}H_{50}N_7O_5^+$ = 732.3873, found 732.3865.

N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-methoxypyridine-2,6-dicarboxamide (23). Following the general procedure above, except using **22**, gave 26.6mg (23.7% yield) of **23** as pale yellow oil. The HCl salt was created by dissolution in concentrated hydrogen chloride and precipitation from diethyl ether to give 11.6mg of the salt as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.4 (s, 2H), 8.2 (s, 2H), 8.1 (d, J=8.2Hz, 2H), 7.9 (s, 2H), 7.8 (d, J=7.9Hz, 2H), 7.6 (t, J=7.6Hz, 2H), 7.4 (t, J=7.8Hz, 2H), 5.0 (s, 2H), 4.4 (t, J=5.3Hz, 4H), 4.0 (s, 3H), 2.9 (t, J=5.5Hz, 4H), 2.4 (s, 12H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 168.7, 163.2, 162.3, 152.1, 150.4, 146.8, 130.4, 127.0, 124.4, 122.1, 119.6, 111.4, 93.8, 67.1, 57.8, 56.1, 46.0. MS (ESI) = 624.3 ($(M+H)^+$,100%).

HRMS (ESI) calc for $C34H37N7O5^+$ = 624.29290, found 624.29300. HPLC 97.9% pure by integration.

Copper-catalyzed coupling general procedure: N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-(3-(1-(2-(1,3-dioxo-1H-benzo[de]isoquinolin-2(3H)-yl)ethyl)-1H-1,2,3-triazol-4-yl)propoxy)pyridine-2,6-dicarboxamide (19a). 24mg (1eq, 0.037mmol) of **18a** and 12mg (1.2eq, 0.044mmol) of azide **5** were combined in a flask and dissolved in 4mL DMF and placed under argon. 407µL (1.1eq, 0.041mmol) of 100mM copper triflate solution was then added, resulting in a green solution. After stirring for 10min under argon, 148µL (0.4eq, 0.015mmol) of freshly prepared 100mM sodium ascorbate solution was added. After stirring at room temperature for 24 hours, the solvent was on a rotary evaporator (50µm Hg) and the residue partitioned between chloroform (approx. 40mL) and a 250mM aqueous sodium hydroxide solution containing 45mg thiourea (approx. 40mL). A small amount of methanol was also added so that the residue completely dissolved. The partitioned mixture was stirred for 24 hours at room temperature before the layers were separated in a separatory funnel. The aqueous layer was washed three times with chloroform (12.5mL) and the combined organic layers were washed with brine before being dried over sodium sulfate. The solvent was then removed under reduced pressure and the residue purified on a silica column (20g, 10% MeOH, 1% TEA in $CHCl_3$), giving 7mg (20% yield) of **19a**. The hydrochloride salt was prepared by dissolving in a minimum of 1M HCl and adding to a 100X volumetric excess of diethyl ether. Recovery of the precipitate gave 2.7mg of the hydrochloride salt for use in screening. ^{1}H NMR (400MHZ, CDC_{l3}) δ 11.4 (s, 2H), 8.5 (d, J=7.4Hz, 2H), 8.2 (s, 2H), 8.1-8.2 (m, 4H), 8.0 (s, 2H), 7.9 (d, J=7.6Hz, 2H), 7.6-7.7 (m, 4H), 7.5 (s, 1H), 7.4 (t, J=7.7Hz, 2H), 4.8 (t, J=5.9Hz, 2H), 4.6 (t, J=5.9Hz, 2H), 4.4 (t, J=5.5Hz, 4H), 4.2 (t, J=6.3Hz, 2H), 2.9-3.0 (m, 6H), 2.4 (s,

12H), 2.2 (p, J=6.7Hz, 2H). MS (ESI) = 942.4 ($(M+H)^+$,100%), 471.7 ($(M+2H)^{2+}$,50%), 964.4 ($(M+Na)^+$, 25%). HRMS (ESI) calc for $C_{52}H_{51}N_{11}O_7^+$ = 942.40460, found 942.40390. EA = 62.15% C, 4.65% H, 13.62% N. HPLC 75.0% pure by integration.

N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-(4-(1-(2-(1,3-dioxo-1H-benzo[de]isoquinolin-2(3H)-yl)ethyl)-1H-1,2,3-triazol-4-yl)butoxy)pyridine-2,6-dicarboxamide (19b). Following the general procedure above except using **18b** with azide **5** gave 17.7mg (64% yield) of **19b**. From this, 14.8mg of the HCl salt was obtained as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.4 (s, 2H), 8.5 (d, J=7.3Hz, 2H), 8.1-8.2 (m, 6H), 7.90 (s, 2H), 7.86 (d, J=8.8Hz, 2H), 7.66-7.70 (m, 4H), 7.5 (s, 1H), 7.4 (t, J=7.5Hz, 2H), 4.7 (t, J=6.0Hz, 2H), 4.6 (t, J=6.4Hz, 2H), 4.4 (t, J=5.5Hz, 4H), 4.1 (t, J=5.8Hz, 2H), 3.0 (t, J=5.5Hz, 4H), 2.8 (t, J=6.9Hz, 2H), 2.4 (s, 12H), 1.82-1.85 (m, 4H). MS (ESI) = 956.4 ($(M+H)^+$, 100%), 478.7 ($(M+2H)^{2+}$, 85%), 978.4 ($(M+Na)^+$, 15%), 994.4 ($(M+K)^+$, 3%). HRMS (ESI) calc for $C_{53}H_{53}N_{11}O_7^+$ = 956.42020, found 956.42080. HPLC 97.9% pure by integration.

N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-((5-(1-(2-(1,3-dioxo-1H-benzo[de]isoquinolin-2(3H)-yl)ethyl)-1H-1,2,3-triazol-4-yl)pentyl)oxy)pyridine-2,6-dicarboxamide (19c). Following the general procedure above except using **18c** with azide **5** gave 11mg (40% yield) of **19c** as a pale yellow residue. From this, 6.3mg of the HCl salt was obtained as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.3 (s, 2H), 8.5 (d, J=7.3Hz, 2H), 8.1-8.2 (m, 6H), 7.93 (s, 2H), 7.88 (d, J=8.2Hz, 2H), 7.66-7.70 (m, 4H), 7.44 (s, 1H), 7.42 (t, J=7.4Hz, 2H), 4.7 (t, J=5.9Hz, 2H), 4.6 (t, J=6.3Hz, 2H), 4.4 (t, J=5.3Hz, 4H), 4.2 (t, J=6.6Hz, 2H), 3.0 (t, J=5.3Hz, 4H), 2.7 (t, J=7.5Hz, 2H), 2.4 (s, 12H), 1.9 (p, J=7.1Hz, 2H), 1.7 (p, J=7.6Hz, 2H), 1.5 (p, J=7.6Hz, 2H). MS (ESI) = 486.7 ($(M+2H)^{2+}$, 100%), 970.4 ($(M+H)^+$, 25%), 992.4 ($(M+Na)^+$, 12%). HRMS (ESI) calc for $C_{54}H_{55}N_{11}O_7$ is 970.43590, found 970.43540. HPLC 99.4% pure by integration.

N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)-4-((7-(1-(2-(1,3-dioxo-1H-benzo[de]isoquinolin-2(3H)-yl)ethyl)-1H-1,2,3-triazol-4-yl)heptyl)oxy)pyridine-2,6-dicarboxamide (19d). Following the general procedure above except using **18d** with azide **5** gave 5.3mg (19.7% yield) of **19d** as a pale yellow oil. From this, 4.4mg of the HCl salt was obtained as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.5 (s, 2H), 8.5 (d, J=6.9Hz, 2H), 8.1-8.2 (m, 6H), 8.0 (d, J=3.2Hz, 2H), 7.9 (d, J=8.0Hz, 2H), 7.7 (t, J=8.2Hz, 4H), 7.41-7.45 (m, 3H), 4.7 (t, J=6.4Hz, 2H), 4.6 (t, J=6.2Hz, 2H), 4.4 (t, J=5.2Hz, 4H), 4.2 (t, J=6.4Hz, 2H), 3.0 (t, J=5.4Hz, 4H), 2.7 (t, J=7.8Hz, 2H), 2.4 (s, 12H), 2.2 (t, J=8.3Hz, 2H), 1.8 (p, J=7.9Hz, 2H), 1.6 (p, J=5.4Hz, 2H), 1.5-1.6 (m, 4H). MS (ESI) = 366.4 ($(M+2H)^{2+}$, 100), 998 ($(M+H)^{+}$, 30%). HRMS (CI) calc for $C_{56}H_{60}N_{11}O_7^{+}$ = 998.4677, found 998.4676. HPLC 95.3% pure by integration.

4-(3-(1-(4-benzoylbenzyl)-1H-1,2,3-triazol-4-yl)propoxy)-N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)pyridine-2,6-dicarboxamide (20a). Following the general procedure above except using **18a** with azide **6** gave 11.7mg (34% yield) of **20a** as a clear residue. From this, 9.3mg of the HCl salt was obtained as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.4 (s, 2H), 8.2 (s, 2H), 8.1 (d, J=78Hz, 2H), 7.91 (s, 2H), 7.86 (d, J=8.4Hz, 2H), 7.8 (d, J=8.1Hz, 2H), 7.73 (d, J=8.5Hz, 2H), 7.67 (t, J=7.7Hz, 2H), 7.5 (t, J=7.4Hz, 1H), 7.38-7.44 (m, 4H), 7.33-7.35 (m, 3H), 5.6 (s, 2H), 4.4 (t, J=5.5Hz, 4H), 4.2 (t, J=5.8Hz, 2H), 2.9-3.0 (m, 6H), 2.4 (s, 12H), 2.3 (p, J=6.8Hz, 2H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 163.3, 162.3, 152.2, 150.4, 147.3, 146.9, 139.1, 137.1, 132.7, 130.7, 130.4, 130.0, 128.3, 127.7, 127.0, 124.4, 122.1, 121.1, 119.7, 111.8, 93.9, 68.0, 67.2, 57.8, 53.6, 46.0, 28.3, 21.9. MS (ESI) = 913.4 ($(M+H)^{+}$,100%), 457.2 ($(M+2H)^{2+}$, 50%), 913.4 ($(M+Na)^{+}$, 2%). HRMS (ESI) calc for $C_{52}H_{52}N_{10}O_6^{+}$ = 913.41440, found 913.41550. HPLC 99.2% pure by integration.

4-(4-(1-(4-benzoylbenzyl)-1H-1,2,3-triazol-4-yl)butoxy)-N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)pyridine-2,6-dicarboxamide (20b). Following the general procedure above except using **18b** with azide **6** gave 16.5mg (61% yield) of **20b** as a clear residue. From this, 5.2mg of the HCl salt was obtained as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.4 (s, 2H), 8.2 (s, 2H), 8.1 (d, J=8.4Hz, 2H), 7.91 (s, 2H), 7.87 (d, J=8.2Hz, 2H), 7.73-7.8 (m, 4H), 7.67 (t, J=7.7Hz, 2H), 7.6 (t, J=7.4Hz, 1H), 7.4-7.5 (m, 4H), 7.34 (d, J=8.4Hz, 2H), 7.31 (s, 1H), 5.6 (s, 2H), 4.4 (t, J=5.5Hz, 4H), 4.2 (t, J=6.1Hz, 2H), 2.9 (t, J=5.4Hz, 4H), 2.8 (t, J=7.7Hz, 2H), 2.4 (s, 12H), 1.87-1.94 (m, 4H). MS (ESI) = 464.2 ($(M+2H)^{2+}$, 50%), 927.4 ($(M+H)^+$, 25%), 949.4 ($(M+Na)^+$, 12%), 965.4 ($(M+K)^+$, 0.5%). HRMS (ESI) calc for $C_{53}H_{54}N_{10}O_6^+$ = 927.43010, found 927.42930. HPLC 98.3% pure by integration.

4-((5-(1-(4-benzoylbenzyl)-1H-1,2,3-triazol-4-yl)pentyl)oxy)-N2,N6-bis(4-(2-(dimethylamino)ethoxy)quinolin-2-yl)pyridine-2,6-dicarboxamide (20c). Following the general procedure above except using **18c** with azide **6** gave 5.3mg (20% yield) of **20c** as a pale yellow residue. From this, 5.8mg of the HCl salt was obtained as a yellow-white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.4 (s, 2H), 8.22 (s, 2H), 8.16 (d, J=8.3Hz, 2H), 8.0 (s, 2H), 7.9 (d, J=8.4Hz, 2H), 7.73-7.8 (m, 4H), 7.70 (t, J=7.6Hz, 2H), 7.6 (t, J=8.1Hz, 1H), 7.4-7.5 (m, 4H), 7.34 (d, J=8.1Hz, 2H), 7.28 (s, 1H), 5.6 (s, 2H), 4.4 (t, J=5.5Hz, 4.2 (t, J=6.3Hz, 2H), 3.0 (t, J=5.7Hz, 4H), 2.8 (t, J=7.6Hz, 2H), 2.4 (s, 12H), 1.9 (p, J=7.6Hz, 2H), 1.8 (p, J=7.7Hz, 2H), 1.6 (p, J=7.5Hz, 2H). MS (ESI) = 471.2 ($(M+2H)^{2+}$, 100%), 941.4 ($(M+H)^+$, 30%), 963.4 ($(M+Na)^+$, 15%). HRMS (ESI) calc for $C_{54}H_{56}N_{10}O_6^+$ = 941.44570, found 941.44560. HPLC 94.9% pure by integration.

3-(4-(3-((2,6-bis((4-(2-(dimethylamino)ethoxy)quinolin-2-yl)carbamoyl)pyridin-4-yl)oxy)propyl)-1H-1,2,3-triazol-1-yl)propyl 9,10-dioxo-9,10-dihydroanthracene-2-carboxylate (21a). Following the general procedure above

except using **18a** with azide **7** gave 14.3mg (64% yield after adjusting for TEA contamination) of **21a** as a pale yellow residue. From this, 8mg of the HCl salt was obtained as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$, contained 4X excess TEA) δ 11.3 (s, 2H), 8.7 (s, 2H), 8.2 (s, 2H), 8.18-8.20 (m, 1H), 8.13-8.15 (m, 1H), 8.06-8.07 (m, 3H), 7.84 (s, 2H), 7.75 (d, J= 8.8Hz, 2H), 7.7 (p, J=2.8Hz, 2H), 7.6 (t, J=7.8Hz, 2H), 7.5 (s, 1H), 7.4 (t, J=7.8Hz, 2H), 4.5 (t, J=6.8Hz, 2H), 4.40 (t, J=5.5Hz, 4H), 4.36 (t, J=6.2Hz, 2H), 4.2 (t, J=6.2Hz, 2H), 3.0 (t, J=5.3Hz, 4H), 2.9 (t, J=7.3Hz, 2H), 2.5 (s, 12H), 2.4 (p, J=6.5Hz, 2H), 2.3 (p, J=6.5Hz, 2H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 163.0, 162.1, 152.0, 150.3, 146.7, 134.4, 134.32, 134.24, 133.1, 127.5, 127.24, 127.20, 126.9, 124.4, 122.0, 119.5, 111.7, 66.7, 62.2, 57.6, 51.71, 51.68, 49.0, 29.4, 28.1. MS (ESI) = 1011.4 ((M+H)+, 100%), 506.2 ($(M+2H)^{2+}$, 90%), 1033.4 ($(M+Na)^{+}$, 6%). HRMS (ESI) calc for $C_{56}H_{54}N_{10}O_9$ = 1011.41480, found 1011.41480. HPLC 80.2% pure by integration

3-(4-(4-((2,6-bis((4-(2-(dimethylamino)ethoxy)quinolin-2-yl)carbamoyl)pyridin-4-yl)oxy)butyl)-1H-1,2,3-triazol-1-yl)propyl 9,10-dioxo-9,10-dihydroanthracene-2-carboxylate (21b). Following the general procedure above except using **18b** with azide **7** gave 12.4mg (58% yield) of **21b** as a yellowish solid. From this, 9.3mg of the HCl salt was obtained as a tan solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.3 (s, 2H), 8.8 (s, 2H), 8.3 (s, 2H), 8.2-8.3 (m, 2H), 8.1-8.2 (m, 3H), 7.9 (s, 2H), 7.84 (d, J= 8.2Hz, 2H), 7.75 (t, J=4.7Hz, 2H), 7.7 (t, J=7.5Hz, 2H), 7.43 (s, 1H), 7.40 (t, J=7.7Hz, 2H), 4.5 (t, J=6.9Hz, 2H), 4.41-4.43 (m, 6H), 4.2 (t, J=5.7Hz, 2H), 3.0 (t, J=5.3Hz, 4H), 2.8 (t, J=6.8Hz, 2H), 2.4-2.5 (m, 2H), 2.4 (s, 12H), 1.92-1.94 (m, 4H). MS (ESI) = 513.2 ($(M+2H)^{2+}$, 100%), 1025.4 ($(M+H)^{+}$, 40%), 1047.4 ($(M+Na)^{+}$, 10%). HRMS (ESI) calc for $C_{57}H_{56}N_{10}O_9^{+}$ = 1025.43040, found 1025.43420. HPLC 72.4% pure by integration

3-(4-(5-((2,6-bis((4-(2-(dimethylamino)ethoxy)quinolin-2-yl)carbamoyl)pyridin-4-yl)oxy)pentyl)-1H-1,2,3-triazol-1-yl)propyl 9,10-dioxo-9,10-dihydroanthracene-2-carboxylate (21c). Following the general procedure above except using **18c** with azide **7** gave 25.5mg (79% yield) of **21c** as a clear residue. From this, 13.5mg of the HCl salt was obtained as a white solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 11.3 (s, 2H), 8.8 (s, 1H), 8.3 (s, 2H), 8.17-8.23 (m, 2H), 8.07-8.12 (m, 3H), 7.84 (s, 2H), 7.79 (d, J=8.3Hz, 2H), 7.6-7.7 (m, 4H), 7.49-7.51 (m, 1H), 7.35-7.40 (m, 3H), 4.5 (t, J=6.8Hz, 2H), 4.1-4.2 (m, 6H), 3.0 (t, J=5.0Hz, 4H), 2.8 (t, J=7.5Hz, 2H), 2.43-2.45 (m, 2H), 2.43 (s, 12H), 1.9 (p, J=7.7Hz, 2H), 1.8 (p, J=7.7Hz, 2H), 1.7 (p, J=6.1Hz, 2H), 1.5 (p, J=7.7Hz, 2H). ^{1}H NMR (400MHZ, DMSO) δ 12.1 (s, 2H), 10.7 (s, 2H), 8.6 (s, 2H), 8.4 (d, J=8.5Hz, 2H), 8.2-8.3 (m, 2H), 8.1-8.2 (m, 3H), 8.0 (d, J=9.6Hz, 2H), 7.87-7.88 (m, 4H), 7.57-7.61 (m, 3H), 4.7 (t, J=4.4Hz, 4H), 4.5 (t, J=6.7Hz, 2H), 4.3 (t, J=6.4Hz, 2H), 4.2 (t, J=6.3Hz, 2H), 3.8 (s, 4H), 3.0 (s, 12H), 2.6 (t, J=7.3Hz, 2H), 2.3 (p, J=6.5Hz, 2H), 1.8 (p, J=7.1Hz, 2H), 1.6 (p, J=7.3Hz, 2H), 1.4 (p, J=7.2Hz, 2H). (ESI) = 520.2 ($(M+2H)^{2+}$, 100%), 1039.4 ($(M+H)^+$, 25%), 1061.4 ($(M+Na)^+$, 8%). HRMS (ESI) calc for $C_{58}H_{58}N_{10}O_9^+$ = 1039.44610, found 1039.44800. HPLC 90.0% pure by integration.

N,N'-[10,10]-Perylene-3,4:9,10-bis(dicarboximide)-cyclophane (24). A 250mL three neck flask was stoppered on its central neck. The left and right necks were outfitted with U-shaped Claisen adapters with the straight portion of each also stoppered. To the left "arm" was added 18mg (1eq, 0.104mmol) of 1,10-diaminodecane, which was then suspended in 24mL DMF. To the right arm was added 41mg (1eq, 0.104mmol) of perelene-3,4,9,10-tetracorboxylic acid dianhydride, which was then suspended in 18mL DMF. 230mL of DMF was added to the three necked flask. Each "arm" was outfitted with a reflux condenser at the top of the bend (See below for diagram of setup). The

DMF in the flask was refluxed with stirring by heating on a manifold, allowing for slow addition of the reagents. After refluxing for 2 weeks, the reaction was still incomplete (there was still amine and dianhydride present in each arm), however the reaction was worked up regardless. The reaction vessel was cooled such that it was safe to handle and its contents poured while still warm into a 500mL flask. The solvent was then removed under high vacuum (50μm) rotary evaporator. The residue was partitioned between chloroform and 1M HCl (color changed from deep red/black to green on contact with acid) and the aqueous layer extracted twice with chloroform. The combined organic layers were then washed once with 250mM NaOH, once with water, and once with brine before being dried over sodium sulfate (the washings were red, but the organic layer remained green). This gave 20.7mg of crude product and TLC (10% MeOH in $CHCl_3$) showed several potential orange or red product spots. The crude product was dissolved in a mixture of methanol and chloroform, adsorbed to silica (500mg), and then purified on a silica column (30g, 2% MeOH in $CHCl_3$). Three products were isolated from the column. MALDI and CI HRMS showed that the first product to elute (2.7mg) contained the desired product, but it was impure by NMR. A second silica column was then conducted (2.5g, 1% MeOH, $CHCl_3$) to give 0.7mg (0.64% yield) of the final pure (by CI HRMS) product as a red solid. HRMS calculated for $(C68H57N4O8)^+$ = 1057.4132, found 1057.4099.

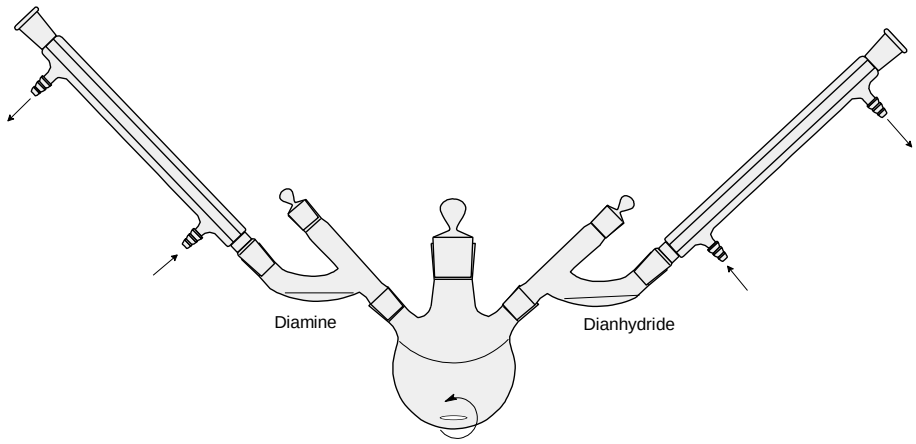


Figure B1: Experimental setup used for the synthesis of the PDI macrocycle **24**.

non-2-yn-1-ol (misc_1) (8)**.** 1mL (1eq, 6.8mmol) of 1-octyne was dissolved in 70mL freshly distilled THF in an oven-dried flask under argon. 10.2mL (10eq, 68mmol) of TMEDA was then added and the resulting clear solution was placed in an acetone/dry ice bath and stirred for 2min. 3.4mL (1.25eq, 8.5mmol) of 2.5M n-BuLi in hexanes was then added drop wise from a gas-tight syringe. After stirring 5min in the acetone/dry ice bath, 1.89g (7.5eq, 51.14mmol) of paraformaldehyde suspended in 20mL of distilled THF was added with swirling via cannula. The mixture was stirred in the bath an additional 15min before being removed from the acetone/dry ice bath and being allowed to warm for 2min before being added to an oil bath at 80°C and refluxing for 40min when the mixture appeared close to homogeneity and TLC (EtOAc, KMnO4 stain) suggested the reaction was complete. The mixture was quenched with the addition of 100mL of saturated aqueous ammonium chloride. The aqueous layer was washed 3X with 70-100mL portions of hexanes and the combined organic layers were then washed with 100mL 1M HCl and 150mL of brine before being dried over sodium sulfate. Concentration on a water-aspirator-equipped rotary evaporator gave 1.48g of the crude product, which was then purified on a silica flash column (50g, 20% EtOAc in Hexanes)

to give 1.19g of 62wt% **misc_1** in 1:1 EtOAc:hexanes (77.3% yield) as a clear liquid. ^{1}H NMR (400MHz, CDCl3) δ 4.2 (t, J=2.2Hz, 2H), 2.2 (t, J=7.1Hz, 2H), 1.8 (s, 1H), 1.5 (p, J=7.2Hz, 2H), 1.2-1.4 (m, 6H), 0.9 (t, J=7.0Hz, 3H). ^{13}C NMR (100MHZ, CDCl3) δ 86.5, 78.3, 51.3, 31.3, 28.54, 28.52, 22.5, 18.7, 14.0.

non-8-yn-1-ol (misc_2) (9)**.** 324.5mg (6eq, 45.4mmol) of lithium wire was added to an oven-dried flask under argon. 32.5mL of 1,3-diaminopropane was then added and the flask was equipped with an oven-dried condenser and drying tube (filled with KOH). The mixture was stirred in an oil bath at 40°C for 30min at which point the lithium dissolved, generating a dark inky-blue solution. After stirring at 70°C for 3 hours, the blue color had discharged and yielded a white suspension. After cooling to room temperature, 3.42g (4eq, 30.5mmol) of potassium tert-butoxide was added gradually, giving a white suspension in a yellow solution. The mixture was stirred for 30min at room temperature. 1.06g (1eq, 7.6mmol) of **misc_1** was then added slowly. The needle was rinsed with 1mL 1,3-diaminopropane and the mixture stirred for 40min. The dark orange mixture was then quenched by addition to 50mL ice water and the aqueous layer was extracted 4X with 50mL portions of hexanes. The combined organic layers were washed with 100mL water, 100mL 1M HCl, and 65mL brine before being dried over sodium sulfate and concentrated on a rotary evaporator, giving 681mg of 37wt% **misc_2** in hexanes (23.6% yield) as a pale yellow solution. ^{1}H NMR (400MHZ, CDCl3) 3.6 (t, J=6.7Hz, 2H), 2.1 (t, J=7.0Hz, 2H), 2.0 (s, 1H), 1.9 (t, J=2.7Hz, 1H), 1.4-1.5 (m, 6H), 1.28-1.31 (m, 4H). ^{13}C NMR (100MHZ, CDCl3) δ 84.5, 68.1, 62.7, 34.6, 32.6, 28.9, 28.6, 28.3, 25.6, 18.3 (one extra peak that likely belongs to a hexanes isomer).

Amidation of In-Situ-Generated Acyl Chloride General Procedure: 4-chloro-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (misc_5). 500mg (1 eq, 2.73mmol) of chelidamic acid monohydrate was dissolve in 2.5mL of 5:2

thionylchloride:dimethylformamide. The solution was stirred at reflux under argon for 2 hours. The thionyl chloride was then distilled off at reduced pressure and 10mL of toluene was added to the acid chloride product. The toluene solution was then added drop wise to another flask containing 3-aminoquinoline. The resulting solution was stirred at reflux under argon for 4 hours. The precipitated product was filtered under reduced pressure and washed 3 times with toluene, followed by extensive washing in methanol, giving 0.19g (35% yield) of **misc_5** as a tan solid. mp = 318°C (decomposition). ^{1}H NMR (400MHz, DMSO) δ 9.6 (d, 2H), 9.2 (s, H), 8.4 (s, 2H), 8.1 (dd, 4H), 7.8 (t, 2H), 7.7 (t, 2H). MS (ESI) m/z = (M+1, 100%). IR (KBr) 1685.30, 1573.14, 1540.35, 1489.53, 1370.38, 1217.36, 768.43, 747.63.

N2,N6-bis(6-bromoquinolin-3-yl)pyridine-2,6-dicarboxamide (misc_21). The general procedure above was followed except that 2,6-dicarboxypyridine and 6-bromo-3-aminoquinoline were reacted to give 37mg (56%) of **misc_21** as a light brown solid. mp = 305.1-310.8°C (decomposition). ^{1}H NMR (400MHz, DMF) δ 11.7 (s, 2H), 9.6 (d, 2H, J=2.4Hz), 9.1 (d, 2H, J=2.4Hz), 8.6 (d, 2H, J=7.8Hz), 8.5 (t, 2H, J=7.6Hz), 8.4 (d, 2H, J=2.0Hz), 8.0 (s, 1H), 7.8 (d, 2H, J=9.0Hz). MS (ESI) m/z =576 (M, 100%); HRMS calc for $C_{25}H_{15}Br_2N_5O_2^+$ was 573.9519, found 573.9522. IR (KBr) 3123.06, 2372.48, 2344.24, 1773.69, 1735.45, 1718.80, 1700.99, 1685.99, 1654.79, 1637.50, 1560.27, 1542.82, 1525.02, 1508.72, 1475.94, 1458.57, 1399.54 cm^{-1}.

N2,N6-bis(7-bromoquinolin-3-yl)pyridine-2,6-dicarboxamide (misc_22). The general procedure above was followed except that 2,6-dicarboxypyridine and 7-bromo-3-aminoquinoline were reacted to give 37mg (56% yield) of **misc_22** as a light tan solid. mp = 319.9-330.2 (decomposition). NMR contained buried peaks and impurities, which makes assignment less certain (16H shown). ^{1}H NMR (400MHz, DMSO) δ 12.2 (s, 2H), 9.9 (d, 2H, J=2.4Hz), 9.4 (d, 2H, J=2.0Hz), 8.7 (d, 2H, J=7.8Hz), 8.6 (s, 1H), 8.4 (d, 2H,

J=14.2Hz), 8.3 (d, 2H, J=8.8Hz), 8.1 (s, 1H), 8.0 (t, 2H, J=9.5Hz). MS (CI) m/z =578 (M+2, 100%); HRMS calc for $C_{25}H_{15}Br_2N_5O_2^+$ 577.9650, found 575.9668. IR (KBr) 3128.06, 1686.53, 1599.30, 1541.92, 1475.32, 1400.54, 1222.36, 1189.25, 901.54, 810.53, 748.88, 648.81 cm^{-1}.

N2,N6-bis(8-bromoquinolin-3-yl)pyridine-2,6-dicarboxamide (misc_23). The general procedure above was followed except that 2,6-dicarboxypyridine and 8-bromo-3-aminoquinoline were reacted to give 28mg (42% yield) of **misc_23** as a white solid. mp = 248°C (decomposition). NMR contained buried peaks and impurities, which makes assignment less certain (17H shown). ^{1}H NMR (400MHz, DMF) δ 12.0 (s, 2H), 9.8 (d, 2H, J=2.4Hz), 9.2 (d, 2H, J=2.0Hz), 8.5 (m, 2H), 8.4 (m, 2H), 8.1 (m, 4H), 7.6 (t, 3H, J=7.8Hz). MS (ESI) m/z =577 (M+1H), 100%); HRMS calc for $C_{25}H_{15}Br_2N_5O_2^+$ was 577.9645, found 577.9644, IR (KBr) 3123.67, 1687.24, 1561.08, 1534.50, 1475.58, 1401.43, 1372.60, 1207.77, 1145.64, 990.02, 763.34, 657.02 cm^{-1}.

2-(2-(ethylamino)ethyl)isoindoline-1,3-dione (misc_6). 0.3mL (1eq, 2.75mmol) of N1-ethylethane-1,2-diamine was added to 2.7mL of distilled water followed by 407mg (1eq, 2.75mmol) of phthalic anhydride. After refluxing for 3 hours, the solvent was removed under reduced pressure and the resulting oily residue suspended in 2mL acetone, resulting in precipitate of the amide, 2-((2-(ethylamino)ethyl)carbamoyl)benzoic acid, which was isolated by filtration under reduced pressure. Heating a sample of the amide with a heat gun resulted in a viscous orange oil, however, it did not appear to contain the product. Heating the solid amide at reflux in glacial acetic acid gave **misc_6** as a clear oil (contained about 1.6eq of acetic acid/acetate). The product was found to be unstable in water and appeared to decompose to the diacid. For reactions, the monoamide would be formed and stored as a pure solid and only converted to **misc_6** on the day of the next step reaction. ^{1}H NMR (400MHz, $CHCl_3$) δ 9.5 (s, 3H), 7.75-7.78 (m,

2H), 7.63-7.66 (m, 2H), 3.9 (t, J= 6.0Hz, 2H), 3.2 (t, J=6.0Hz, 2H), 3.0 (q, J=7.2Hz, 2H), 1.2 (t, J=7.3Hz, 3H). 13C NMR (100MHz, $CHCl_3$) δ 168.0, 133.9, 132.0, 123.2, 44.6, 42.0, 34.3, 11.3.

2-chloro-N-(2-(1,3-dioxoisoindolin-2-yl)ethyl)-N-ethylacetamide (misc_7). First, the phthalimide **misc_6** was prepared following the procedure described above, except that the residual acetic acid was removed under high vacuum (100µm) over 4 hours. Next, the oil from the first step was dissolved in 8mL of methylene chloride under argon. The flask was then placed in an ice bath. 1mL (3.5eq, 7.42mmol) of distilled triethylamine was gradually added drop wise alternatingly with the drop wise addition of 0.25mL (1.5eq, 3.17mmol) of chloroacetylchloride under increased argon flow, forming a dark solution. The reaction solution was stirred for 80 minutes, when it appeared to be complete by TLC (EtOAc and 75% MeOH in EtOAc). A large amount of white precipitate was observed (triethylammonium chloride salt), which was removed by filtration under reduced pressure. The filtrate was diluted to 20mL with methylene chloride and washed consecutively with 10mL 0.1M NaOH, 10mL water, 10mL 0.1M HCl, 10mL brine, and dried over sodium sulfate. After purification on a silica column (50% EtOAc), 246mg (40% yield) of **misc_7** was obtained as a clear oil. ^{1}H NMR (400MHz, $CHCl_3$) δ 7.9-8.0 (m, 2H), 7.67-7.69 (m, 2H), 4.1 (s, 0.5H), 4.0 (s, 1.5H), 3.8-3.9 (m, 2H), 3.63 (t, J=5.4Hz, 1.5H), 3.57 (t, J=6.8Hz, 0.5H), 3.5 (q, J=7.3Hz, 0.5H), 3.4 (q, J=7.1Hz, 1.5H), 1.18-1.24 (m, 3H). ^{13}C NMR (100MHz, $CHCl_3$) δ 168.3, 168.0, 166.8, 134.4, 134.0, 132.0, 131.7, 123.6, 123.3, 45.2, 43.6, 43.0, 41.2, 40.89, 40.83, 36.1, 35.4, 14.2, 14.0, 12.4

N-(2-(1,3-dioxoisoindolin-2-yl)ethyl)-N-ethylhept-6-ynamide (misc_9). First, 143µL (1eq, 1.12mmol) of 6-heptanoic acid was added to an oven-dried conical flask under argon. The flask was then placed in a water bath at 10°C (causing the acid to

freeze). 121µL (1.8eq, 1.4mmol) of oxalyl chloride was then added slowly with stirring under increased argon flow. The flask was then stirred for one hour, allowing the bath to return to room temperature. Next, 522µL (4.2eq, 3.7mmol) of freshly distilled triethylamine was added to 234mg (1eq, 0.75mmol) of 70wt% (in acetic acid) **misc_6** in 2mL methylene chloride and under argon. The flask was placed in a water bath. The solution from step one was added drop wise by syringe under increased argon flow. 1mL of methylene chloride was added to dissolve a small amount of precipitate that formed. The reaction solution was then stirred for 1 hour, allowing the bath to return to room temperature. At this point the reaction appeared complete by TLC (50% EtOAc in hexanes). The reaction solution was diluted to 10mL with methylene chloride and the organic layer was washed consecutively with 6mL 0.2M NaOH, two times with 10mL water, 10mL 0.1M HCl, 10mL brine, and was dried over sodium sulfate. The organic layer was concentrated under reduced pressure and purified with a silica column (25%-100% EtOAc) to give 125mg (51%) of **misc_9** as a clear oil. A significant side product (the acetyl instead of the chloro) resulted from the presence of acetic acid, partially explaining the poor yield. NMR reveals formation of rotamers. ^{1}H NMR (400MHz, $CHCl_3$) δ 7.8-7.9 (m, 2H), 7.66-7.68 (m, 2H), 3.9 (t, J= 5.7Hz, 1.5H), 3.8 (t, J= 7.8Hz, 0.5H), 3.6 (t, J=5.7Hz, 1.5H), 3.5 (t, J= 7.4Hz, 0.5H), 3.4 (q, J= 7.0Hz, 0.5H), 3.3 (q, J= 7.2Hz, 1.5H), 2.4 (t, J= 7.7Hz, 0.5H), 2.21 (t, J= 7.4Hz, 1.5H), 2.16 (t of d, J=7.0Hz, 0.5H), 2.1 (t of d, J= 7.1Hz, 1.5H), 1.9 (t, J=2.6Hz, 1H), 1.4-1.6 (m, 3.5H), 1.1 (t, J=6.6Hz, 3H). ^{13}C NMR (100MHz, $CHCl_3$) δ 172.8, 168.3, 134.3, 133.9, 132.1, 131.7, 123.5, 123.2, 84.2, 68.41, 68.36, 44.8, 43.3, 42.5, 40.7, 35.8, 32.2, 28.1, 28.0, 24.2, 18.2, 14.0, 12.9.

General Phthalimide Deprotection Procedure N-(2-aminoethyl)-N-ethylhept-6-ynamide (misc_10). 168mg (1eq, 0.51mmol) of **misc_9** was dissolved in 4mL ethanol.

The solution was sparged with argon for 5min and then the flask was placed in an ice bath. 25µL (1.5eq, 0.77mmol) of anhydrous hydrazine was then added drop wise under argon. The solution was refluxed for 2 hours when it appeared complete by TLC (EtOAc) and had a large amount of white precipitate. 12mL of 1M NaOH was added to the flask and was extracted twice with 20mL of chloroform, and once with 50mL chloroform. The combined organic layers were washed with 50mL brine and dried over sodium sulfate. Removal of the solvent under reduced pressure gave 86mg of the crude product, which was purified on a silica column (10% MeOH, 0.5% TEA, in EtOAc) to give 72.5mg (72% yield) of **misc_10** as a colorless oil. ^{1}H NMR (400MHz, $CHCl_3$) δ 6.5 (s, 1H), 3.3 (q, J=5.3Hz, 2H), 2.9 (s, 2H), 2.7 (t, J=5.6Hz, 2H), 2.6 (q, J=7.1Hz, 2H), 2.11-2.15 (m, 4H), 1.9 (s, 1H), 1.7 (p, J=7.3Hz, 2H), 1.5 (p, J=7.3Hz, 2H), 1.0 (t, J= 7.1Hz, 3H). ^{13}C NMR (100MHz, $CHCl_3$) δ 172.8, 84.0, 68.4, 48.2, 43.5, 38.7, 35.9, 27.8, 24.7, 18.0, 14.7. MS (ESI) m/z = 197.2 ($(M+H)^+$, 100%); HRMS calc for $C_{11}H_{21}N_2O^+$ = 197.16484, found 197.16460.

pent-4-yn-1-amine (misc_31). First, 1920mg (1.2eq, 12.92mmol) of phthalimide and 5.72g (1.2eq, 21.52mmol) of triphenylphosphine were added to an oven-dried flask and placed under high vacuum (100µm Hg) for 30min before being backfilled with argon. The mixture was then dissolved in 80mL of freshly distilled THF. 1mL (1eq, 10.76mmol) of 4-pentyn-1-ol was added and the flask cooled in an ice bath. 2.4mL (1.2eq, 12.92mmol) of DIAD was then added drop wise over 25min. The pale yellow solution was stirred for 3 days under argon when the reaction appeared complete by TLC (20% EtOAc in Hexanes). The solvent was removed under reduced pressure and the residue purified on a silica column (130g, prepack 5% EtOAc in hexanes, elute 10% EtOAc in hexanes) to give 2.25g (98% step 1 yield) of phthalimide protected amine as

fluffy white crystals. Next, all (1eq, 10.55mmol) of the intermediate was deprotected using the general procedure described above except that the organic extract was distilled an azeotrope (believed to be with methanol at 70°C) containing the amine began to collect in the receiving flask. The source flask was shown to contain amine, ether, methanol, and residual water. The amine was distilled once more over sodium hydroxide under high vacuum (100µm Hg) at 40°C using ice water in the condenser and with the receiving flask sitting in dry ice. This gave approximately 300mg (17% yield) of 49wt% **misc_31** in water and methanol. ^{1}H NMR (400MHZ, CDCl3) δ 4.9 (s, 2H), 2.8 (t, J=7.2Hz, 2H), 2.2 (t, J=7.1Hz, 2H), 1.9 (t, J=2.7Hz, 1H), 1.6 (p, J=7.1Hz, 2H).

benzyl (2-(((benzyloxy)carbonyl)amino)ethyl)(ethyl)carbamate (misc_12) (10)**.** 0.6mL (1eq, 5.5mmol) of N1-ethylethane-1,2-diamine was added to 28mL of 2M sodium carbonate at 5°C in a water bath. 7mL of freshly distilled dioxane was then added and the solution stirred as the temperature was reduced to 2°C. 1.9mL (2.4eq, 13.2mmol) of benzyloxycarbonyl chloride dissolved in 7mL of freshly distilled dioxane was then added drop wise with vigorous stirring under argon while 14mL 1M sodium hydroxide was simultaneously added, with the temperature maintained below 8°C. After addition was complete, 18mL of chilled distilled water was added to facilitate stirring overnight, with the temperature allowed to increase to 20°C. Most of the dioxane was removed under reduced pressure (rotary evaporator) before the solution was extracted once with 75mL of diethyl ether and twice with 35mL diethyl ether (TLC suggested the second two extractions were unnecessary). The organic layer was then washed consecutively with 75mL of 1M $KHSO_4$, 75mL 1M sodium bicarbonate, and 75mL brine before being dried over sodium sulfate. Removal of the solvent under reduced pressure (rotary evaporator) followed by rigorous drying under high vacuum (100µm Hg) at 50°C

yielded 1.87g of approximately 94wt% **misc_12** (90% yield) as a pale yellow oil containing residual dioxane and benzyl chloride (or possibly benzyl alcohol). ^{1}H NMR (400MHz, $CDCl_3$) δ 7.3-7.4 (m, 10H), 5.10 (s, 2H), 5.06 (s, 2H), 3.3-3.4 (m, 6H), 1.09 (s, 3H). ^{13}C NMR (100MHz, $CDCl_3$) δ 165.9, 156.9, 136.6, 128.7, 128.0, 127.7, 46.5, 42.6, 40.2, 13.8.

benzyl-(2-(((benzyloxy)carbonyl)(ethyl)amino)ethyl)(tert-butoxycarbonyl)carbamate (misc_13) (10)**.** 1.87g (1eq, 5.25mmol) of **misc_12** was dissolved in 19mL of dry acetonitrile. 132mg (0.2eq, 1.08mmol) of dimethylaminopyridine (DMAP) was then added. The flask was chilled in a water bath to 10°C and placed in argon. 1.41g (1.2eq, 6.3mmol) of di-tert-butyl dicarbonate was then added with stirring. The flask was then removed from the water bath and stirred at room temperature for 2.5 hours when the reaction appeared complete by TLC (50% EtOAc in Hexanes). The solution was then stirred at room temperature under air until morning to remove excess Boc_2O. The acetonitrile was removed under reduced pressure giving a brown, oily residue. This was partitioned between 130mL diethyl ether and 64mL of 1M $KHSO_4$. The organic phase was washed consecutively with 65mL 1M $KHSO_4$, 65mL 1M sodium bicarbonate, and 65mL brine before being dried over sodium sulfate. After treating with activated carbon, the solvent was removed under reduced pressure and the residue dried under high vacuum at 50°C to give 1.96g of approximately 94wt% **misc_13** (77% yield) as a reddish oil. The NMR spectrum appeared to contain rotamers. ^{1}H NMR (400MHz, $CDCl_3$) δ 7.31-7.37 (m, 10H), 5.0-5.9 (m, 4H), 3.76-3.83 (m, 2H), 3.39-3.41 (m, 2H), 3.2-3.3 (m, 2H), 1.42-1.46 (m, 2H), 1.0-1.1 (m, 3H). ^{13}C NMR (100MHz, $CDCl_3$) δ 155.9, 155.7, 153.7, 153.5, 151.8, 136.8, 128.5, 128.4, 128.3, 127.8, 127.7, 83.1, 68.4, 67.0, 44.9, 42.6, 27.9, 13.1.

tert-butyl (2-(ethylamino)ethyl)carbamate (misc_14) (10)**.** 1.79g (1eq, 3.93mmol) of **misc_13** was placed under argon. 112.5mg of 10% Pd/C was added, followed by 54mL of 80% aqueous acetic acid. The argon was removed and replaced with hydrogen gas in three pump/purge cycles. Each time the hydrogen pressure was depleted or an aliquot was taken for TLC, these cycles were repeated. After 26 hours, the reaction appeared complete by TLC (25% EtOAc in Hexanes) as suggested by the disappearance of UV-absorbent spots. The reaction mixture was filtered through celite and the solvent removed under reduced pressure. The residue was partitioned between 160mL ether and 160mL 30% sodium carbonate. The aqueous layer was washed twice with 60mL portions of ether and the combined organic layers were washed twice with 60mL portions of brine and dried over sodium sulfate. Removal of the solvent under reduced pressure (rotary evaporator) gave 710mg (96% yield) of **misc_14** (approximately 80-85mol% pure) as a reddish oil. ^{1}H NMR (400MHz, $CDCl_3$) δ 5.0 (s, 1H), 3.2 (q, J=5.8Hz, 2H), 2.7 (t, J=5.9Hz, 2H), 2.6 (q, J=7.2Hz, 2H), 1.7 (s, 1H), 1.4 (s, 9H), 1.1 (t, J=7.1Hz, 3H). ^{13}C NMR (100MHz, $CDCl_3$) δ 156.1, 79.1, 48.9, 43.7, 40.3, 28.4, 15.2

tert-butyl (2-(2-chloro-N-ethylacetamido)ethyl)carbamate (misc_15). 300mg (1eq, 1.6mmol) of **misc_14** in 5mL of dichloromethane was placed under argon. 0.36mL (1.6eq, 2.56mmol) of freshly distilled triethylamine was added and the flask was placed in an ice bath. Slowly, 190µL (1.5eq, 2.39mmol) of chloroacetylchloride was added under argon. The solution was stirred for 3.5 hours (a precipitate began to form after 2 hours), allowing the bath to return to room temperature. Some additional triethylamine was added, and 20 minutes later, the reaction appeared complete by TLC (75% MeOH in EtOAc, KMnO4 stain). The reaction mixture was diluted in dichloromethane (dissolving the formed precipitate) and partitioned against water. The aqueous layer was washed

twice with dichloromethane and the combined organic layers were washed with 1M HCl, 1M NaHCO3, brine, and finally dried over sodium sulfate. Removal of the solvent under reduced pressure gave 407mg (96% yield) of **misc_15** as a reddish oil. The NMR spectrum appeared to contain two rotamers. ^{13}C reported peak shifts are averages of peaks believed to be due to rotamers. ^{1}H NMR (400MHz, $CDCl_3$) δ 5.0 (s, 1H), 4.09 (s, 0.5H), 4.05 (s, 1.5H), 3.2-3.4 (m, 6H), 1.40 (s, 3H), 1.38 (6H), 1.2 (t, J=7.2Hz, 2H), 1.1 (t, J=7.1Hz, 1H). ^{13}C NMR (100MHz, $CDCl_3$) δ 168.2, 156.2, 79.3, 46.1, 43.5, 41.1, 38.8, 28.3, 13.3.

4-bromo-*N*,*N*'-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (misc_16). 450mg of Chelidamic acid monohydrate (1eq, 2.5mmol) was added to 2.7g (7.37eq, 7.37mmol) of phosphorous pentabromide. The solid mixture was melted with stirring at 90°C for 4 hours under argon, while fitted with a reflux condenser. After being allowed to cool to room temperature, 5mL of chloroform was added and the mixture stirred for 30 minutes. An additional 3mL of chloroform was used to rinse the reaction flask once. The mixture was filtered, yielding the putative intermediate (4-bromopyridine-2,6-dicarbonyl dibromide) in the filtrate. The purple solution was used without further purification and was added to 783mg (2eq, 5.46mmol) of 3-aminoquinoline in 2mL of chloroform. 10mL of distilled water was then added to the reaction mixture, which was stirred for 16 hours with the pH maintained between 7 and 8.5 with a 10M sodium hydroxide solution. Small amounts of chloroform were added to maintain the solvent ratio close to 1:1. The mixture was filtered, and the white solid which was washed twice with small amounts (3-5mL) of chloroform, dried completely, and then washed twice with small amounts (3-5mL) of methanol over a filter at reduced pressure to give 827mg (67% yield) of **misc_16** as a white solid. $AgNO_3$ test for halides was positive for

bromide. mp = 320-321°C (decomposition). ^{1}H NMR (400MHz, DMF) δ 9.4 (d, J=2.8Hz, 2H), 9.0 (d, J=2.4Hz, 2H), 8.6 (s, 2H), 8.1 (d, J=8.4Hz, 4H), 7.74 (t, J=7.6Hz, 2H), 7.65 (t, J=7.2Hz, 2H), MS (ESI) m/z = 498 (M+1, 100%); HRMS calc for C25H16BrN5O2 ($M+H^{+}$) 498.0565, found 498.0560. IR (KBr) 3265.87, 1675.00, 1544.48, 1492.16, 1467.65, 1370.59, 1215.23, 783.00, 748.38, 712.45 cm^{-1}

4-ethynyl-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (misc_17). First, 11.6mg of purified tetrakis triphenylphosphine palladium(0) was obtained by repeated washing in distilled and degassed ethanol under argon atmosphere followed by drying under high vacuum and flushing with argon. To an oven-dried 100mL round bottom flask containing argon was added 233mg (1eq, 0.47mmol) **misc_16**, an oven dried stir bar, and 2.2 mg (0.02eq, 0.011mmol) of purified CuI. The flask was flushed with argon after each addition. To this dry mixture was added 50mL freshly distilled THF and 1mL freshly distilled triethylamine forming a solution containing a suspension of starting material. The mixture was covered in foil and stirred while being vigorously degassed with argon for 20 minutes. 0.1mL (1.5eq, 0.70mmol) of trimethylsilylacetylene was then added and degassing continued for 5 more minutes. Working under an argon atmosphere, the 11.6mg (0.021eq, 0.010mmol) of palladium catalyst was added by rinsing with a small volume of the reaction mixture and subsequent transfer back to the reaction flask. The reaction suspension was heated to 78°C for 20 minutes at which point all starting material had entered solution. The temperature was reduced to 74°C and the reaction was allowed to continue for 16 hours. The solvent was then removed under reduced pressure and the flask washed once with methanol, which was again removed under reduced pressure while the crude product was adsorbed to silica. A short column yielded 111mg of the TIPS-protected intermediate as a white solid. The intermediate was

negative for bromine in a silver nitrate test and appeared clean by ^{1}H NMR and TLC. Next, all (1eq, 0.215mmol) of the intermediate was dissolved in 10mL of THF. The flask was then placed in a -78°C bath (dry ice in acetone) and allowed to cool while stirring for 10 minutes. 0.43mL of TBAF (2eq, 0.4305mmol) was added drop wise while the solution continued to stir for 10 more minutes. 200mg of silica was added to the flask and the solvent removed under reduced pressure. The product was obtained after running a short column and combining fractions gave about 70mg (34% yield starting from **misc_16**) of **misc_17** as a white powder. mp = 272°C (decomposition). ^{1}H NMR (400MHz, DMF) δ 11.4 (s, 2H), 9.4 (d, 2H), 9.1 (d, 2H), 8.5 (s, 2H), 8.08-8.05 (m, 4H), 7.74 (t, 2H), 7.66 (t,2H), 5.0 (s, 1H), MS (ESI) m/z = (M+1, 100%); HRMS calc for C27H18N5O2 (M+1) 444.1416, found 444.1458. IR (KBr) 3197.07, 2111.68, 1674.15, 1606.00, 1576.70, 1551.85, 1492.05, 1426.06, 1405.35, 1374.75 cm^{-1}.

3,3'-((4-bromopyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) (misc_18). To 104mg (1eq, 0.21mmol) of **misc_16** in 10mL of DMF, was added 0.25mL (20eq, 4.01mmol) of methyl iodide. The mixture was allowed to stir for 3.25 days at room temperature, protected from light. The bright orange-yellow precipitate was obtained after filtering over a fine filter and washing three times with small portions (3-5mL) of cold methanol and twice with 1:1 methanol: chloroform. Additional product was obtained by removing DMF and methanol under high vacuum from the first filtrate and filtering again over a fine filter, washing four times with 1:1 methanol: chloroform. Combined, this gave 115mg (70% yield) of **misc_18** as an orange-yellow solid. mp = 286.6-292.2°C (decomposition). ^{1}H NMR (400MHz, DMSO) δ 10.1 (d, J=1.6Hz, 2H), 9.7 (s, 2H), 8.7 (s, 2H), 8.5-9.0 (m, 4H), 8.2 (t, J=8.0Hz, 2H), 8.1 (t, J=7.4Hz, 2H), 4.8 (s, 6H), MS (ESI) m/z = 263.5 (M^{2+}, 100%); HRMS calc for $C_{27}H_{22}BrN_5O_2^{2+}$ 263.5510,

found 263.5471. IR (KBr) 3171.49, 1687.33, 1607.47, 1543.65, 1519.16, 1400.34, 1376.80, 1237.85, 1217.4, 771.57 cm^{-1}.

6-nitro-4-(pent-4-yn-1-yloxy)quinoline (misc_19). First, 76µL (1.2eq, 1.1mmol) of 4-pentyn-1-ol was added drop-wise to 33mg (1eq, 1.3mmol) of 60% NaH in 5mL of dry DMF under argon at 0°C. The resulting suspension was stirred for two hours while returning to room temperature to generate a 0.14M alkoxide stock solution. Next, 4mL (1.2eq, 0.57mmol) of the alkoxide stock solution was added with stirring to 103mg (1eq, 0.48mmol) of 4-chloro-6-nitroquinoline under argon, resulting in a very dark brown solution. The reaction solution was heated at 65°C for 18 hours when the reaction appeared complete by TLC (EtOAc, $KMnO_4$ stain). The reaction was quenched by the addition of two drops of glacial acetic acid. The solvent was removed under reduced pressure using a high-vacuum (50µm Hg) rotary evaporator. The resulting crude solid was then redissolved in methanol/chloroform, adsorbed to silica and separated *via* flash chromatography (20% EtOAc in hexanes up to 75% EtOAc in hexanes). Recrystallization from EtOAc and hexanes gave 60.6mg (41.5% yield) of the product as reddish crystals. ^{1}H NMR (400MHZ, $CDCl_3$) δ 9.0 (s, 1H), 8.8 (d, J=5.2Hz,1H), 8.4 (d, J=9.2Hz, 1H), 8.0 (d, J=9.2Hz, 1H), 6.8 (d, J=5.2Hz, 1H), 4.3 (t, J=6.2Hz, 2H), 2.5 (t, J=6.7Hz, 2H), 2.2 (p, J=6.4Hz, 2H), 2.0 (t, J=2.6Hz, 1H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 162.7, 154.9, 151.3, 144.8, 130.6, 123.2, 120.5, 119.5, 102.1, 82.7, 69.7, 67.6, 27.5, 15.4. MS (Negative Ion CI) m/z = 256 (M^-).

2,9-di(pentan-3-yl)anthra[2,1,9-def:6,5,10-d'e'f']diisoquinoline-1,3,8,10(2H,9H)-tetraone (misc_34) (11)**.** 150mg (1eq, 0.38mmol) of perelene-3,4,9,10-tetracorboxylic acid dianhydride and 110µL (2.4eq, 0.92mmol) of pentan-3-amine were mixed with 2g of imidazole under argon. The mixture was then heated to 150-155°C for 30min, resulting in a red solution that gradually thickened into a slurry. The temperature

was lowered to 140°C and the mixture stirred for 3.5 hours. The reaction mixture was then cooled to room temperature with the addition of a small volume of ethanol to dissolve recrystallizing imidazole. About 40mL of 2M HCl was then added with vigorous stirring and the mixture allowed to stand for 1 hour before being filtered under reduced pressure and washed with water. The solid was dried on the filter under reduced pressure for an hour before being collected and further dried under high vacuum (100μm) for 4 hours. This gave about 212mg (quantitative yield) of **misc_34** as a dark red solid. ^{1}H NMR (400MHZ, $CDCl_3$) δ 8.6 (d, J=8.0Hz, 4H), 8.5 (d, J=8.1Hz, 4H), 5.00-5.08 (m, 2H), 2.2-2.3 (m, 4H), 1.9-2.0 (m, 4H), 0.9 (t, J=7.5Hz, 12H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 134.3, 131.3, 129.4, 126.2, 123.5, 122.9, 57.7, 25.0, 11.4.

5-bromo-2,9-di(pentan-3-yl)anthra[2,1,9-def:6,5,10-d'e'f']diisoquinoline-1,3,8,10(2H,9H)-tetraone (misc_35) (12)**.** 250mg (1eq, 0.47mmol) of **misc_34** was dissolved in 15mL methylene chloride. 1.65mL (68eq, 32mmol) of elemental bromine was added drop wise with stirring. The reaction mixture was placed in a water bath shielded from light and stirred for 6 days at ambient temperature. The solution was then sparged with air to drive off excess bromine. The residue was dissolved in a small amount of chloroform and purified on a flash column (70g, 50% $CHCl_3$/CH_2Cl_2). The first band yielded 16.7mg (5% yield) of the undesired dibromo product. The second band to elute gave 64.9mg (23% yield) of the desired monobrominated product **misc_35**. 176.9mg (71% recovery) of pure starting material **misc_34** was also recovered. ^{1}H NMR (400MHZ, $CDCl_3$) δ 9.7 (d, J=8.3Hz, 1H), 8.8 (s, 1H), 8.59-8.63 (m, 3H), 8.49-8.52 (m, 2H), 4.96-5.07 (m, 2H), 2.2-2.3 (m, 4H), 1.9-2.0 (m, 4H), 0.9 (t, J=7.5Hz, 12H). ^{13}C NMR (100MHZ, $CDCl_3$) δ 163.7, 1391, 133.7, 133.4, 133.34, 133.31, 131.0, 130.5, 128.9, 128.6, 128.0, 127.9, 126.9, 123.6, 122.9, 120.8, 57.9, 57.7, 25.0, 24.9, 11.3.

APPENDIX C: SUPPLEMENTAL FIGURES AND TABLES

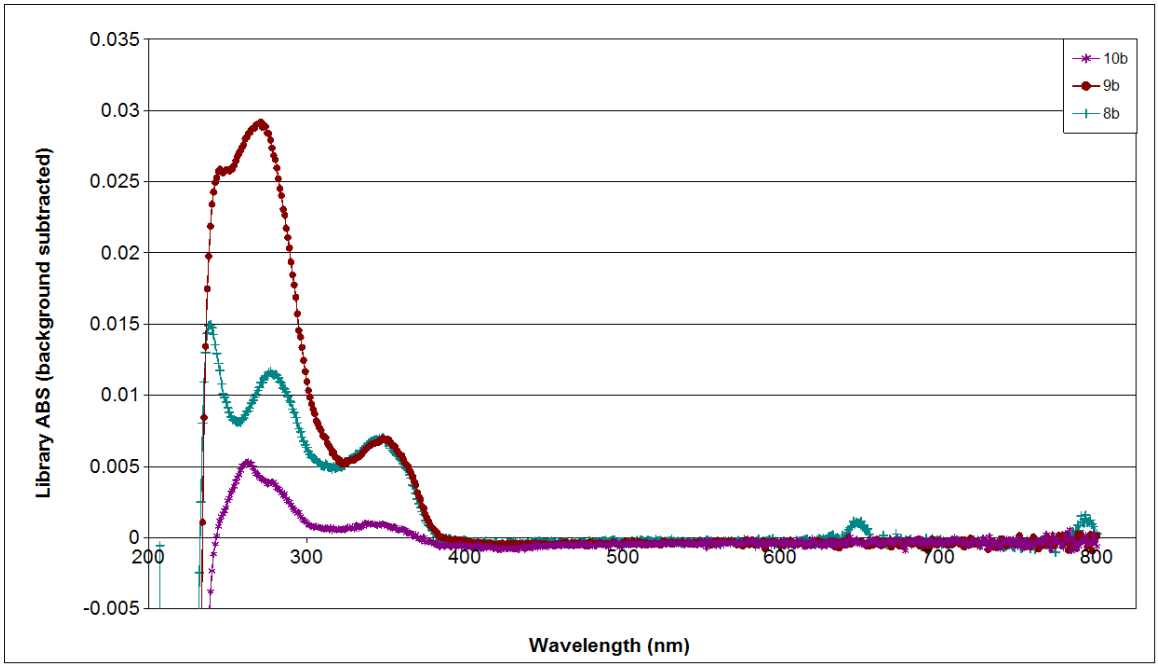


Figure C1: Representative first generation library compound absorbance spectra. Spectra were obtained for 1μM compound in water containing a small percentage of DMSO and have been corrected for any background absorbance. See Chapter 3.

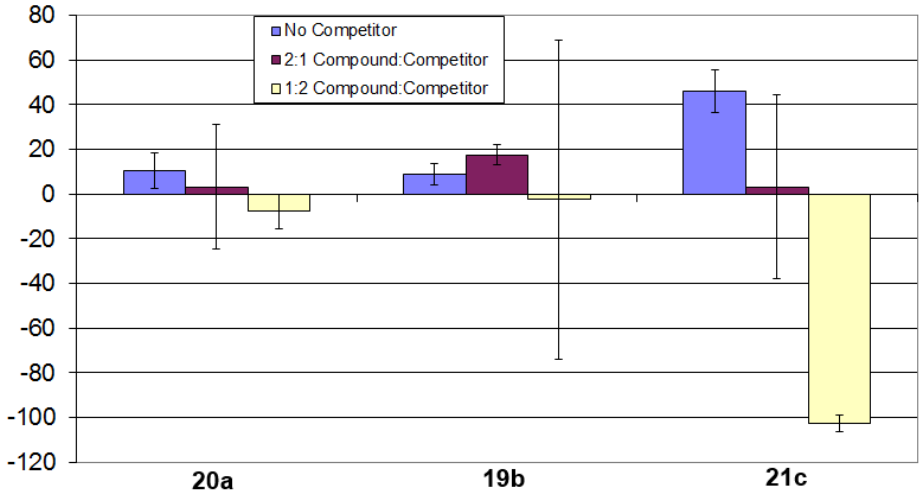
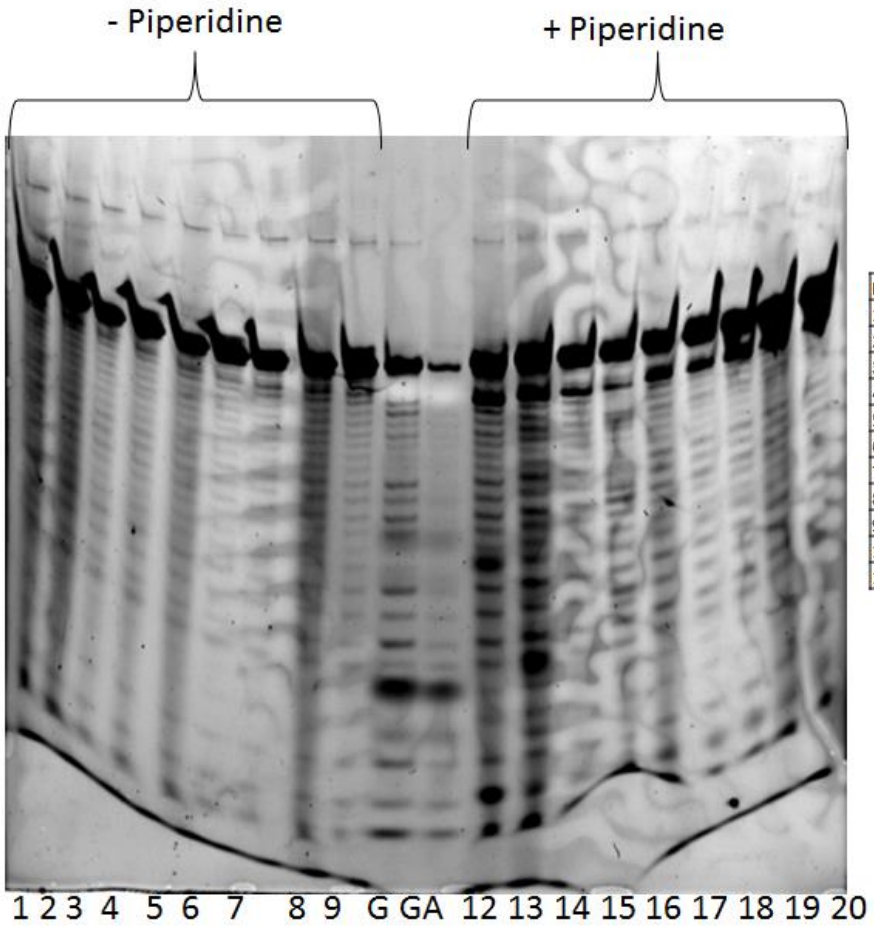


Figure C2: Effect of photocleavage-inactive competitor 23 on photocleavage. F21T was treated with a subset of compounds from the second generation library in the presence of varying concentrations of **23** and irradiated for 30min with UVA-centered lamps. **23** appeared to inhibit photocleavage in a concentration dependent manner. See Chapter 3.



Lane #	Sample Loaded
1/12	19b Untreated
2/13	19b 1000nM
3/14	20a Untreated
4/15	20a 1000nM
5/16	Untreated
6/17	21a Untreated
7/18	NA
8/19	21c Untreated
9/20	21c 1000nM
10	G
11	GA

Figure C3: Example Second Generation Gel (Competition with 23). F21T was treated with a subset of compounds from the second generation library in the presence or absence of 1000nM **23** and irradiated for 30min with UVA-centered lamps. **23** appeared to inhibit photocleavage in a concentration dependent manner. See Chapter 3.

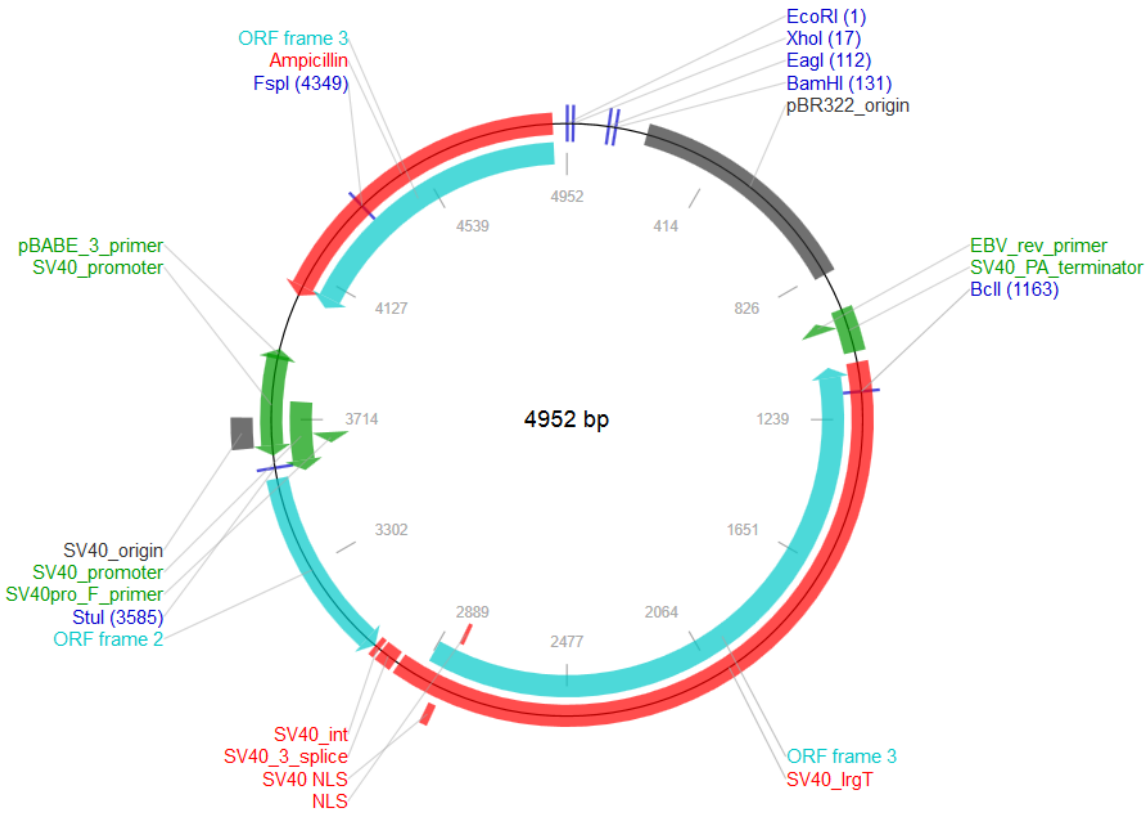


Figure C4: Vector Map of Plasmid pSP189. The G4-forming sequence was inserted between the EcoRI and XhoI restriction sites to produce pSP189G4. The vector map was generated at www.addgene.org. See chapter 4.

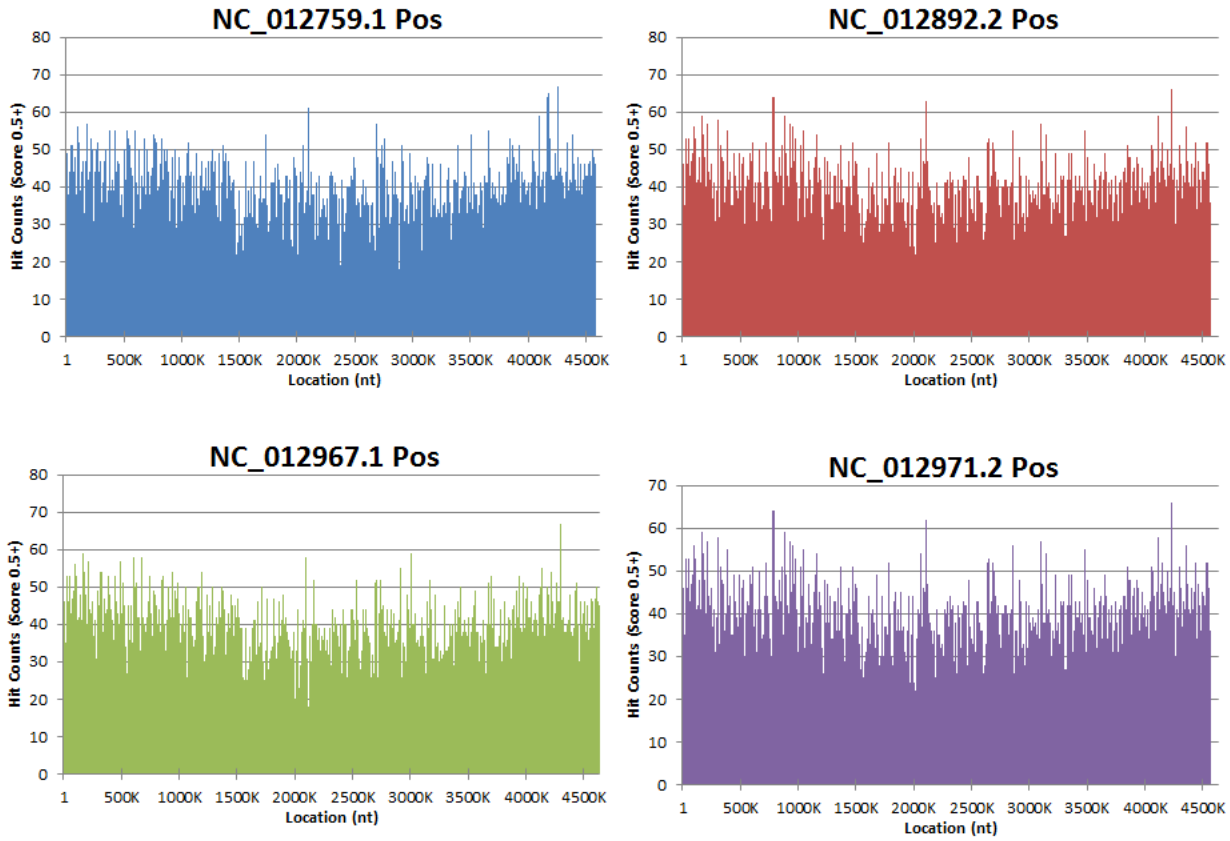


Figure C5: E. coli Strain B Hit Distributions in Positive Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

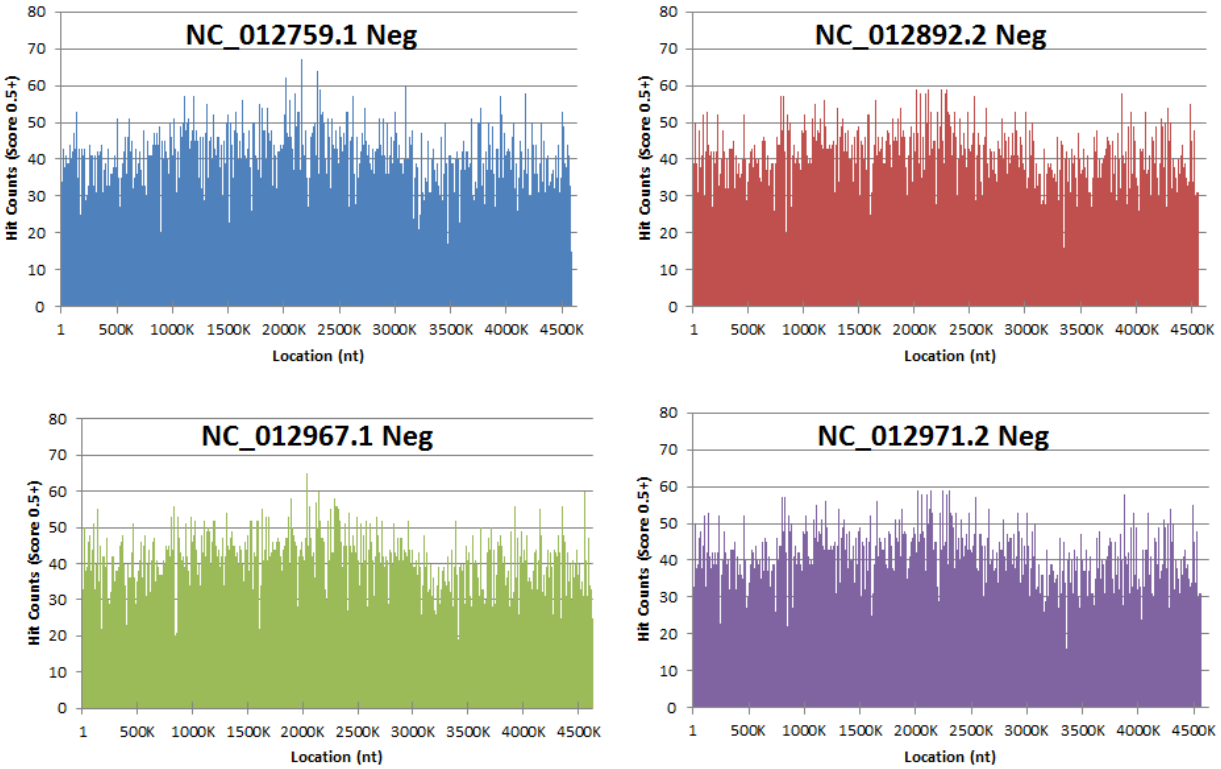


Figure C6: E. coli Strain B Hit Distributions for Negative Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

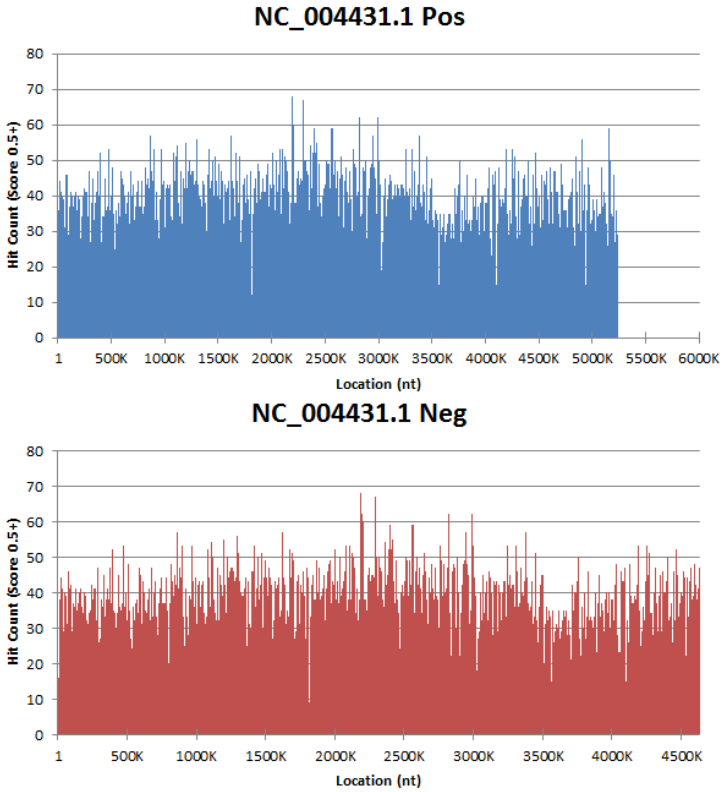


Figure C7: E. coli Strain C Hit Distributions for Both Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

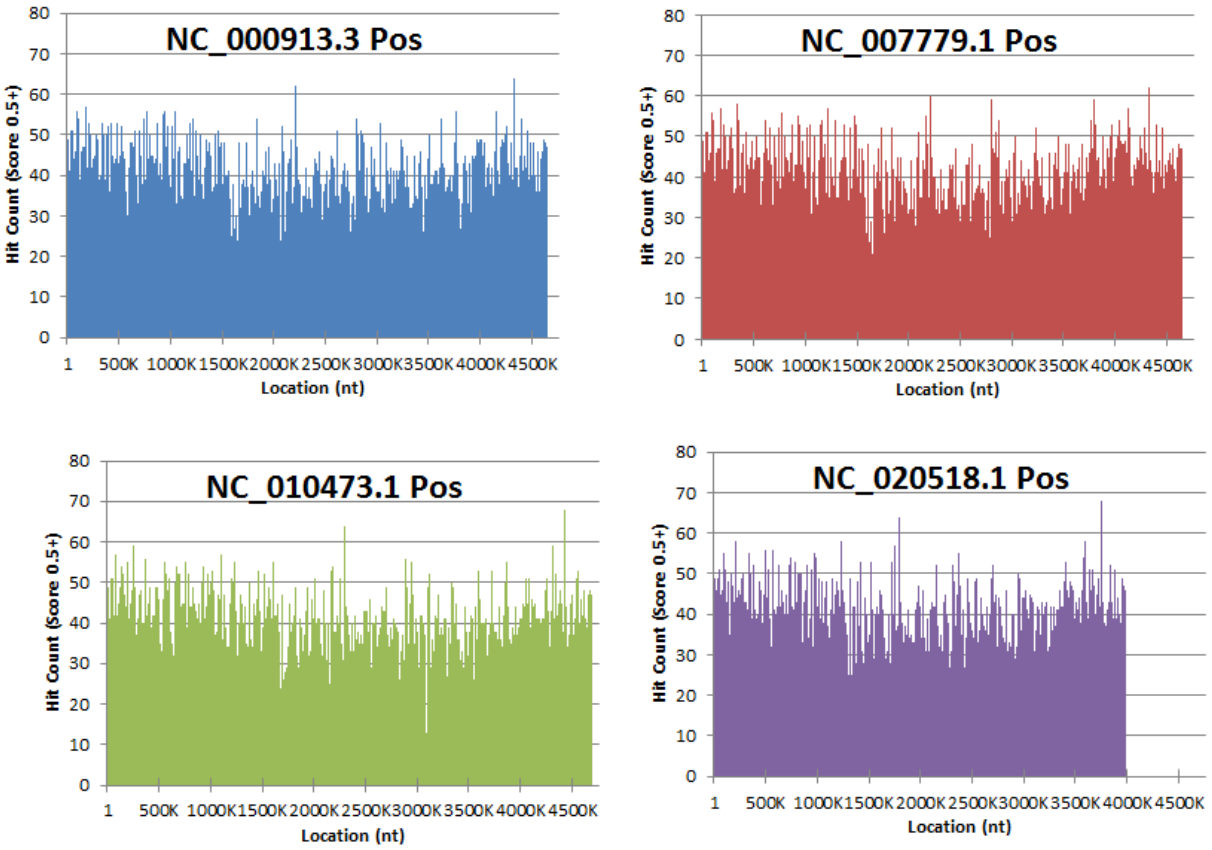


Figure C8: E. coli Strain K12 Hit Distributions for Positive Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

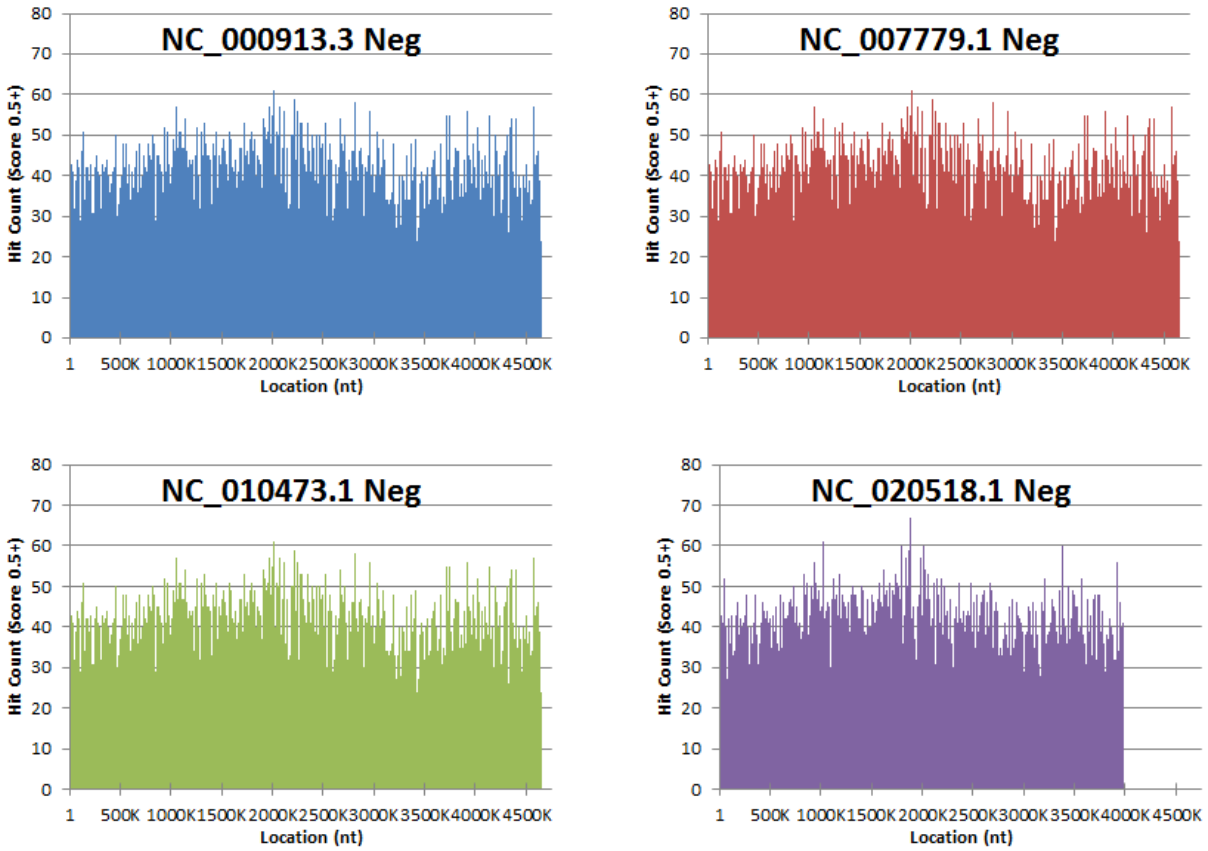


Figure C9: E. coli Strain K12 Hit Distributions for Negative Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

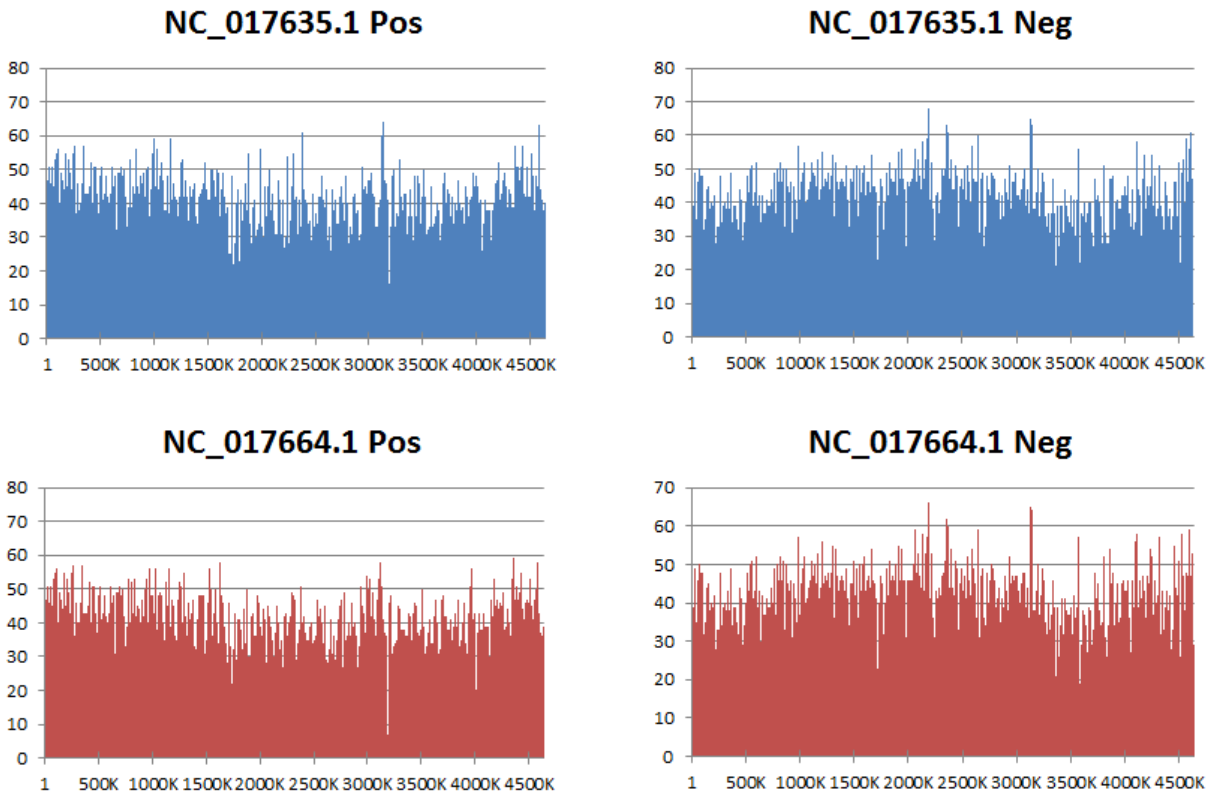


Figure C10: E. coli Strain W Hit Distributions for Both Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

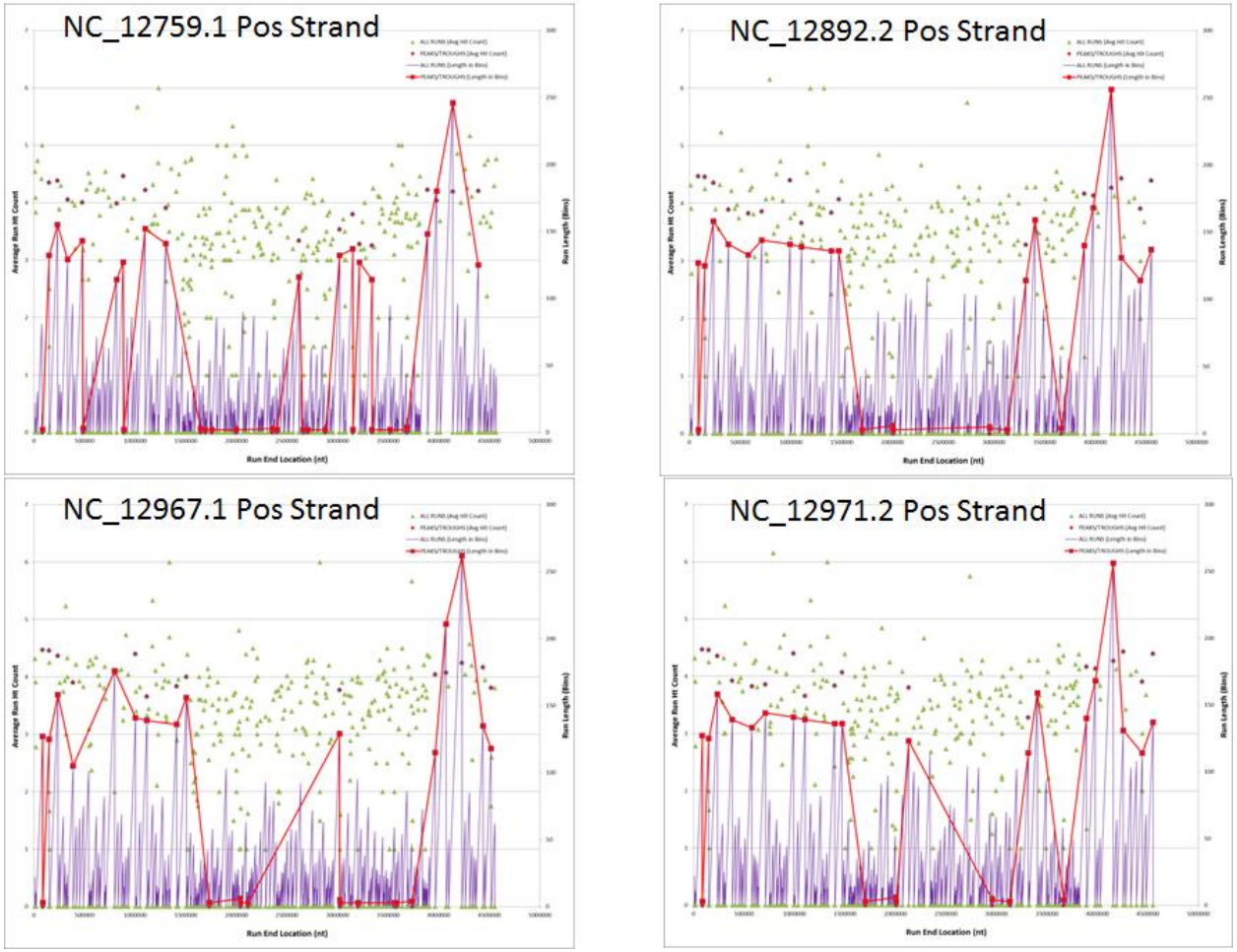


Figure C11: E. coli Strain B Run Distributions for Positive Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

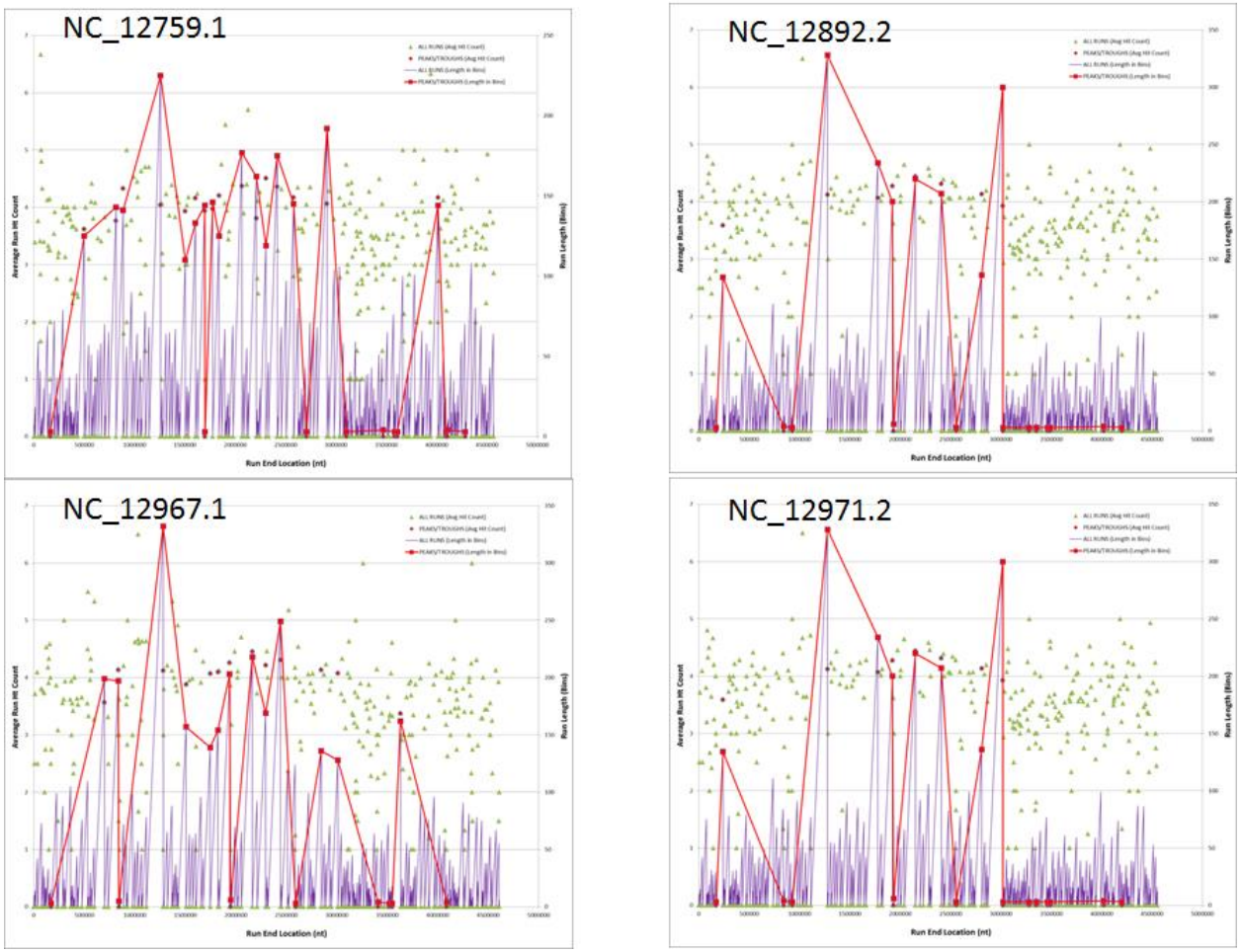


Figure C12: E. coli Strain B Run Distributions for Negative Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

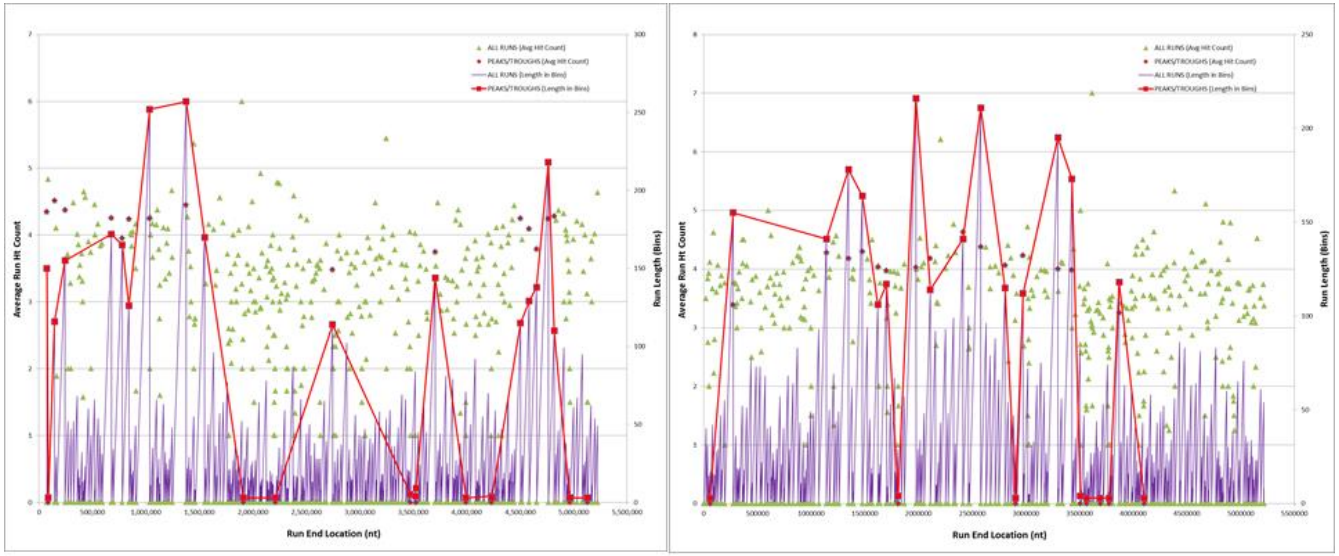


Figure C13: E. coli Strain C Run Distributions for Both Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

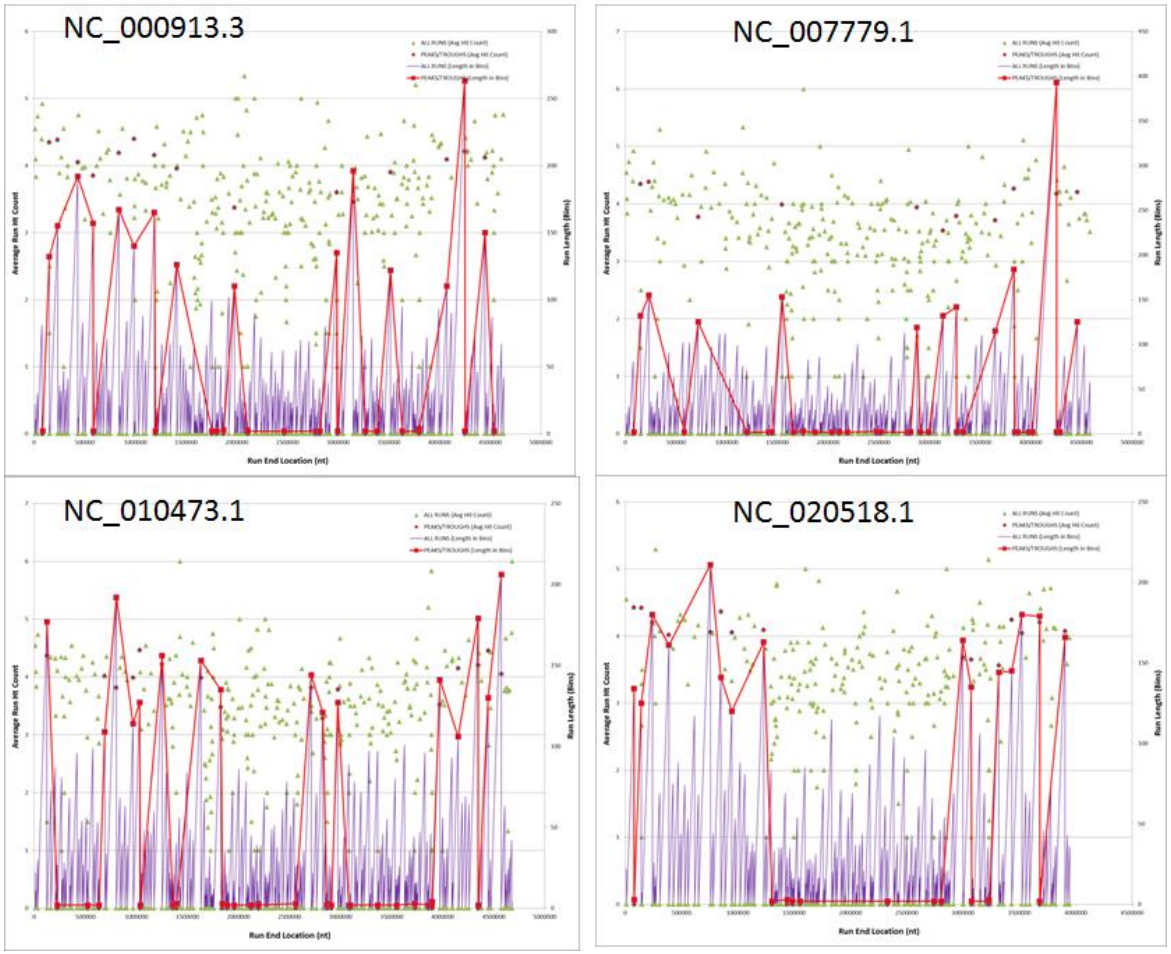


Figure C14: E. coli Strain K12 Run Distributions for Positive Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

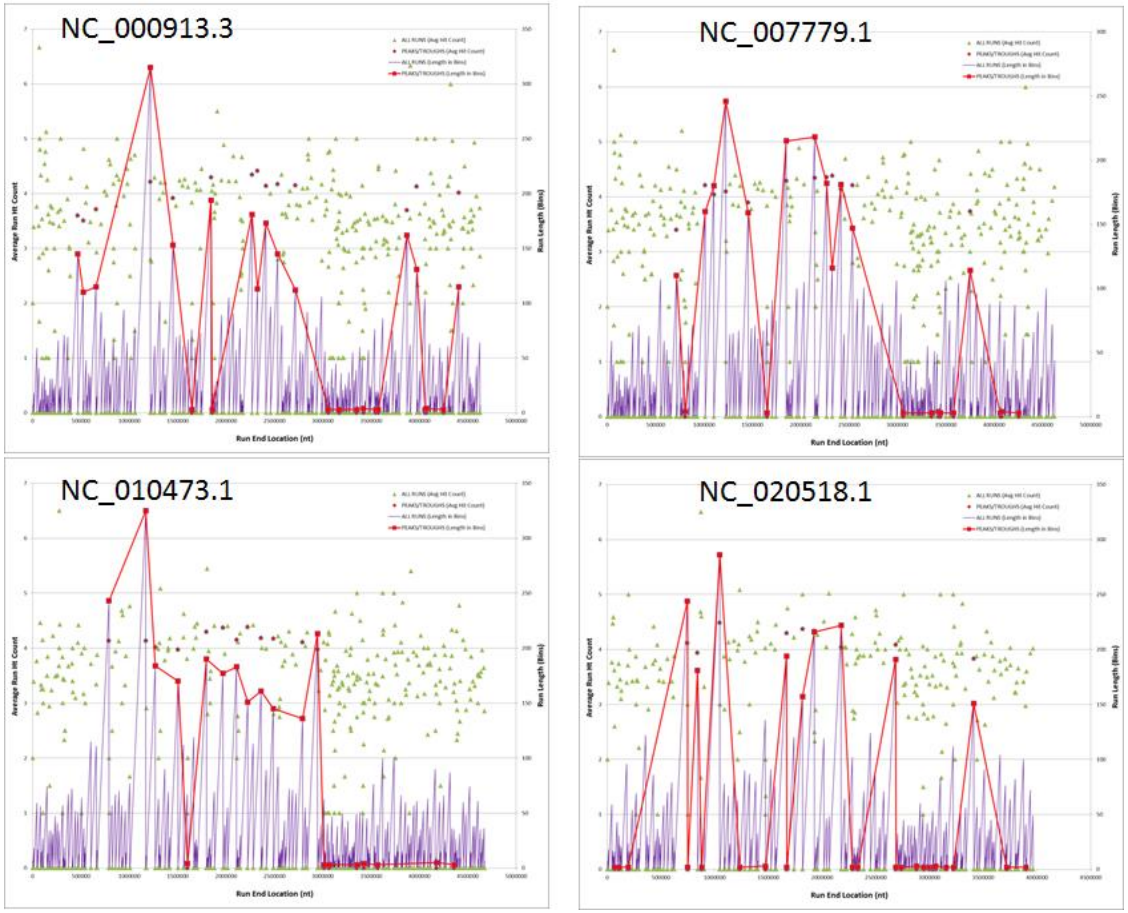


Figure C15: E. coli Strain K12 Run Distributions for Negative Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

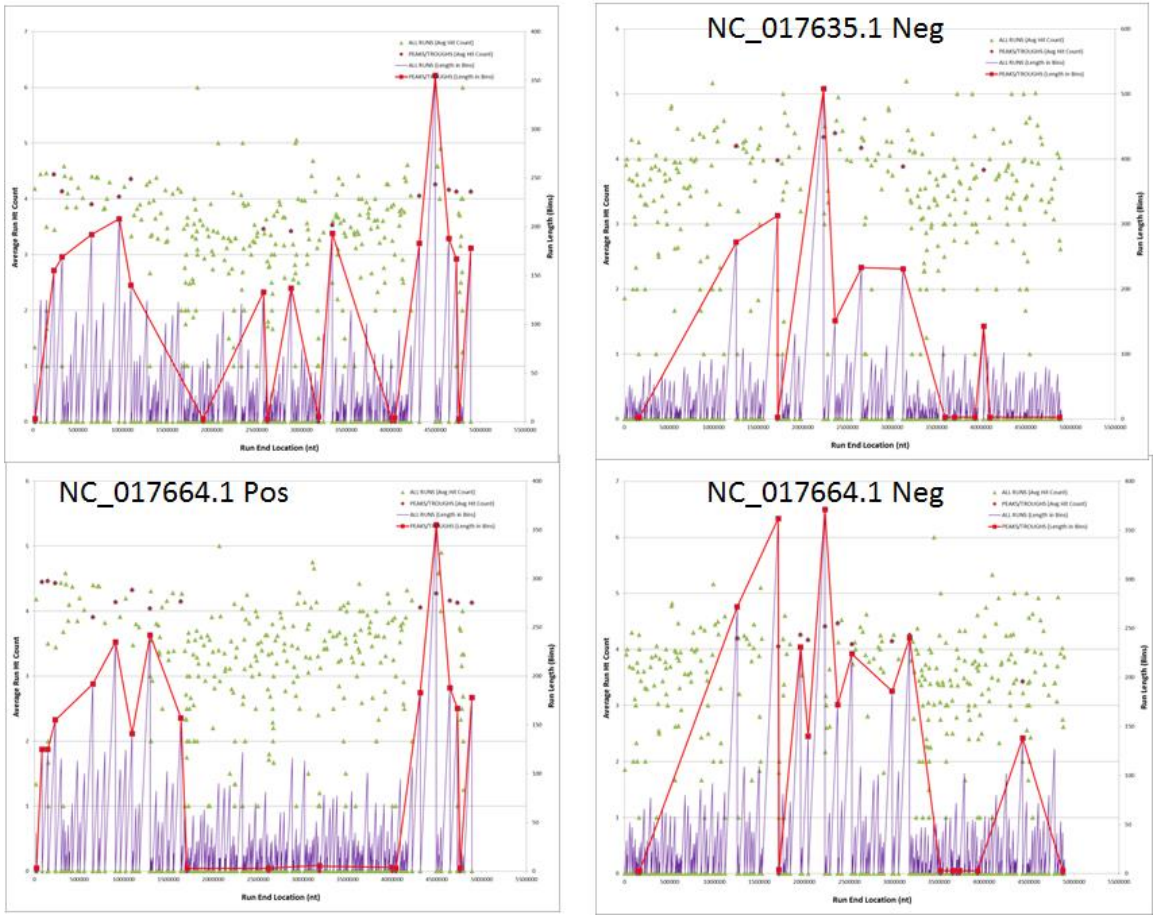
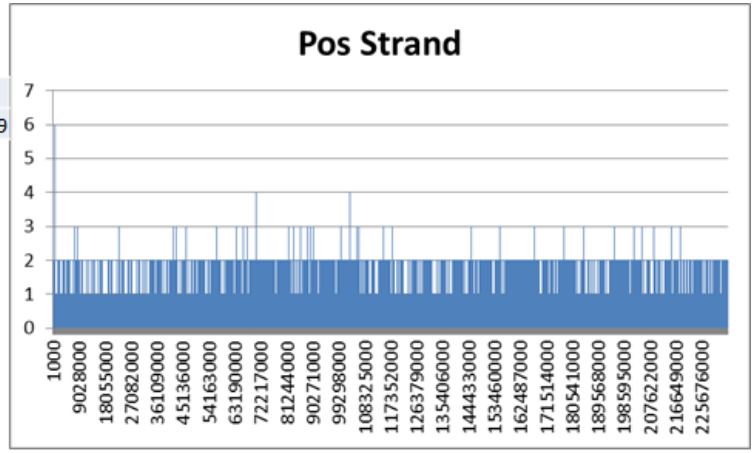


Figure C16: E. coli Strain W Run Distributions for Both Strands. The bin size used was 500nt. Results are for all hits with total scores >= 0.5. The substrain ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
234708820	1000	0.9999

Subsequence ID	Subsequence Length
NT_077402.3	197666
NT_187170.1	40302
NT_077912.2	188020
NT_032977.10	121390471
NT_187171.1	198076
NT_187172.1	278512
NT_187173.1	2282185
NT_187174.1	63597
NT_187175.1	83495
NT_187176.1	251763
NT_004487.20	80374348
NT_167186.2	25337487
NT_187361.1	175055
NT_187362.1	32032
NT_187363.1	127682
NT_187364.1	66860
NT_187365.1	40176
NT_187366.1	42210
NT_187367.1	176043
NT_187368.1	40745
NT_187369.1	41717
NT_187515.1	354444
NT_187517.1	256271
NT_187514.1	109528
NT_187520.1	185285
NW_003315905.1	182439
NW_003315906.1	110268
NW_003315907.2	366580
NT_187521.1	50258
NT_187519.1	911658
NT_187516.1	425601
NT_187518.1	165834
NT_187646.1	162212



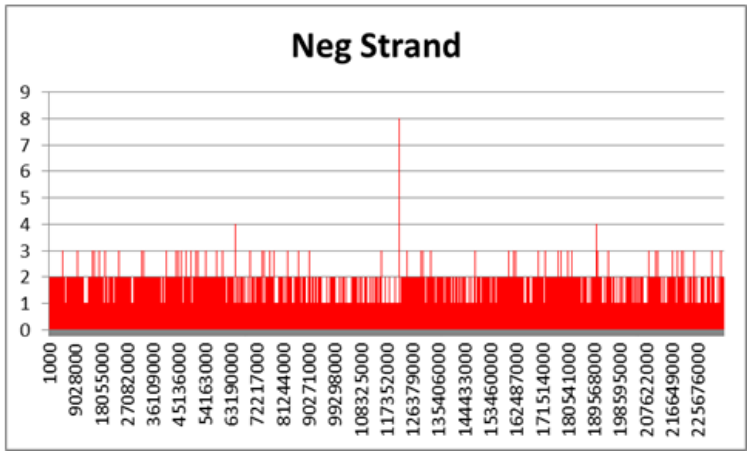


Figure C17: H. Sapiens Chromosome 1 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
2432654	1000	0.9999

Subsequence ID	Subsequence Length
NT_005334.17	16135119
NT_022184.16	73184560
NT_187177.1	871832
NT_034508.3	735634
NT_187178.1	1902412
NT_187179.1	152458
NT_005403.18	147687514
NT_187370.1	161471
NT_187371.1	153799
NT_187525.1	136240
NT_187526.1	70887
NT_187529.1	223625
NT_187522.1	120616
NW_003315908.1	143390
NT_187524.1	133041
NT_187531.1	138019
NT_187530.1	110395
NT_187528.1	110099
NW_003571033.2	96131
NW_003315909.1	123821
NT_187527.1	174166
NT_187523.1	161578
NT_187648.1	214158
NT_187647.1	161218

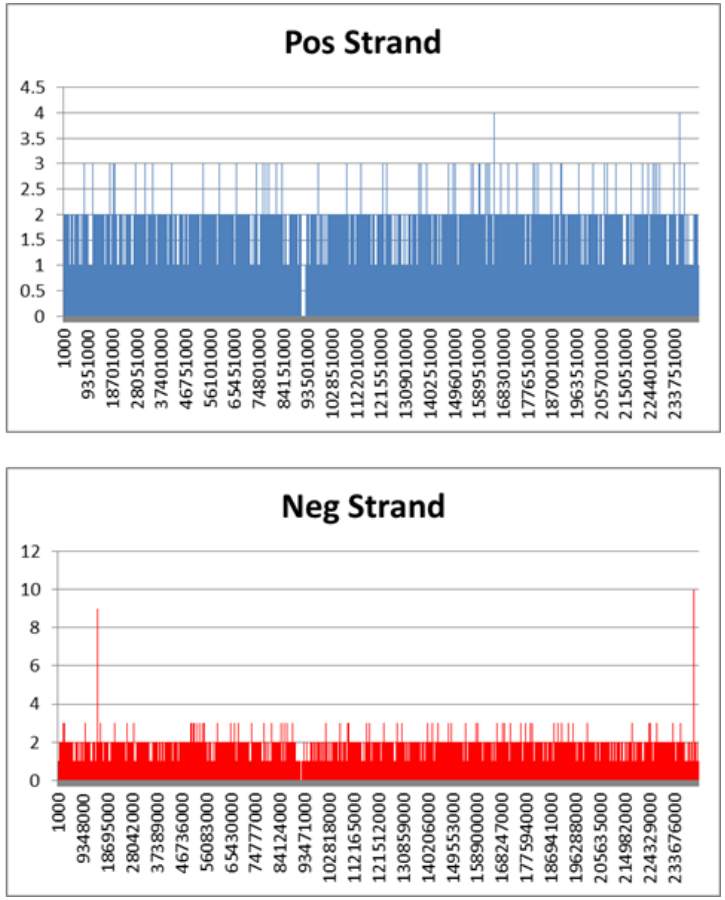


Figure C18: H. Sapiens Chromosome 2 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
201074875	1000	0.9999

Subsequence ID	Subsequence Length
NT_022517.19	90712458
NT_187180.1	461128
NT_187181.1	13936
NT_187182.1	305597
NT_187183.1	2102155
NT_005612.17	104529985
NT_167215.1	155397
NW_003871060.2	173151
NT_187535.1	109187
NT_187537.1	224108
NW_003315913.1	180671
NT_187533.1	173649
NT_187536.1	248252
NT_187538.1	113034
NT_187532.1	205312
NT_187534.1	162429
NT_187539.1	184404
NT_187649.1	162896
NT_187678.1	166540
NT_187688.1	163458
NT_187689.1	197351
NT_187690.1	164170
NT_187691.1	165607

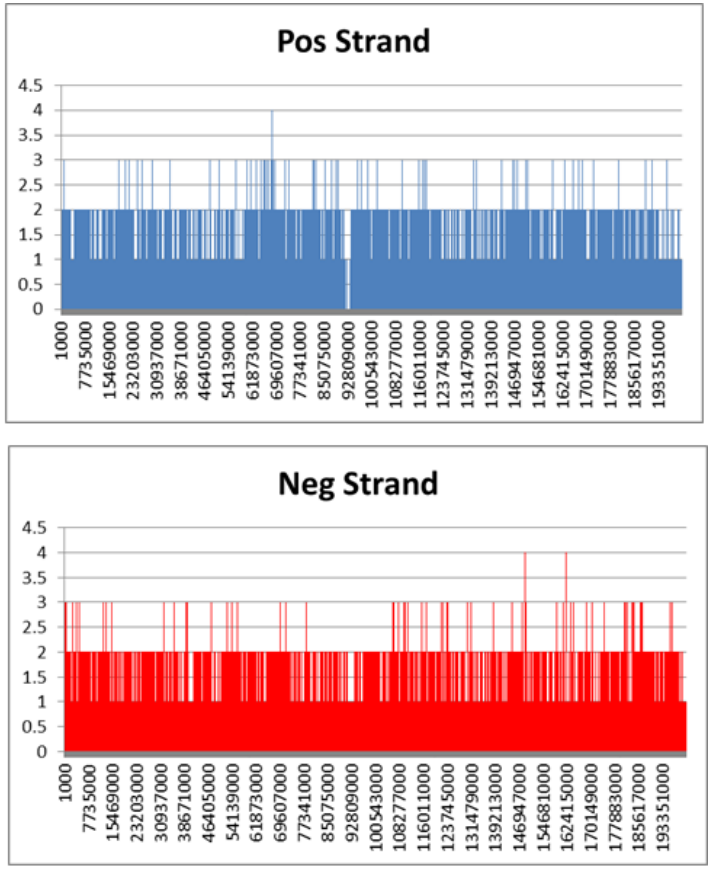


Figure C19: H. Sapiens Chromosome 3 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
193195953	1000	0.9999

Subsequence ID	Subsequence Length
NT_006051.19	8787477
NT_006316.17	23002818
NT_022794.11	1000447
NT_006238.12	16497908
NT_037645.2	171176
NT_187185.1	2031890
NT_022853.16	7084842
NT_016354.20	131283174
NT_113793.3	209709
NT_187540.1	220246
NW_003315915.1	376187
NT_187541.1	111943
NT_167250.2	586476
NT_187544.1	158965
NW_003315914.1	164536
NT_187542.1	119912
NT_187545.1	205944
NT_187543.1	244096
NT_187650.1	378547
NT_187679.1	555799

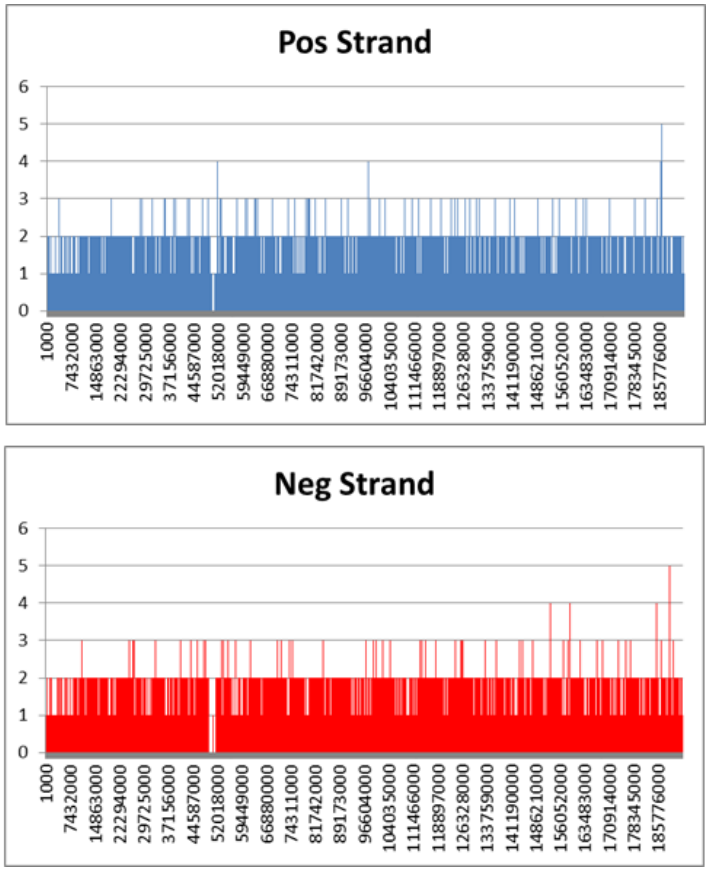
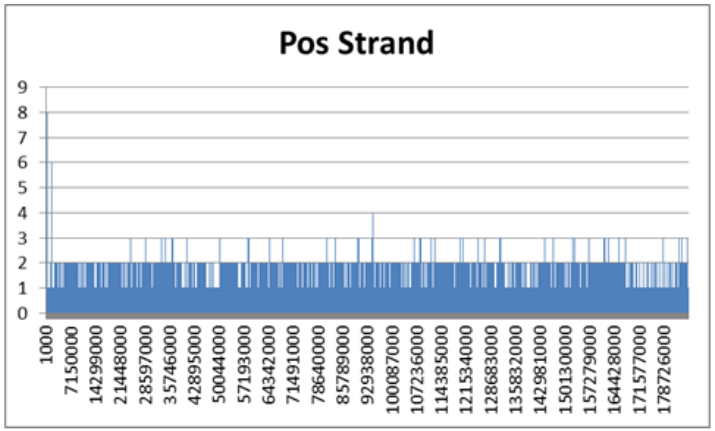


Figure C20: H. Sapiens Chromosome 4 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
185900378	1000	0.9999

Subsequence ID	Subsequence Length
NT_006576.17	46425900
NT_187186.1	83162
NT_187187.1	227563
NT_187188.1	264463
NT_187189.1	45506
NT_187190.1	46345
NT_187191.1	142630
NT_187192.1	12915
NT_187193.1	2282185
NT_187194.1	9517
NT_187195.1	45737
NT_187196.1	20508
NT_187197.1	53672
NT_187198.1	338504
NT_034772.7	89342852
NT_029289.12	16306665
NT_023133.14	25716935
NT_113948.1	92689
NT_187550.1	126136
NT_187548.1	179043
NT_187547.1	195710
NW_003315920.1	82728
NW_003571036.1	226852
NT_187551.1	164558
NW_003315917.2	1612928
NW_003315918.1	101241
NT_187549.1	172708
NW_003315919.1	173459
NT_187546.1	131892
NT_187651.1	1144418
NT_187652.1	130957



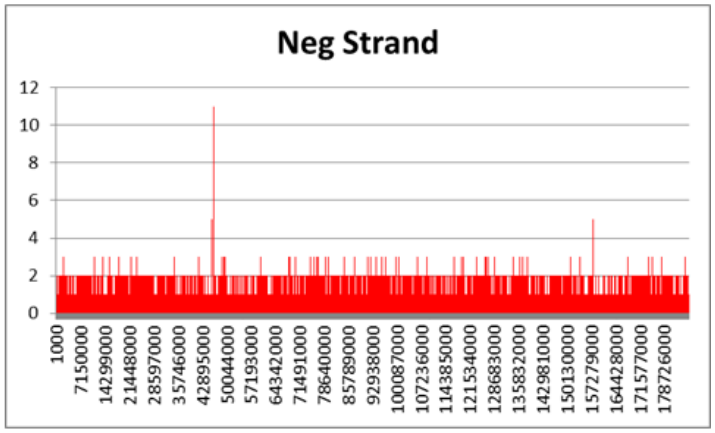


Figure C21: H. Sapiens Chromosome 5 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
205429612	1000	0.9999

Subsequence ID	Subsequence Length
NT_007592.16	58393888
NT_187199.1	1276046
NT_025741.16	110516045
NT_167244.2	4672374
NT_187555.1	175808
NT_187554.1	152148
NW_003315921.1	124736
NT_187556.1	870480
NT_187557.1	75005
NW_004166862.2	185823
NT_187552.1	197536
NT_187553.1	271782
NT_113891.3	4795265
NT_167245.2	4604811
NT_167246.2	4677643
NT_167247.2	4827813
NT_167248.2	4606388
NT_167249.2	4929269
NT_187692.1	76752

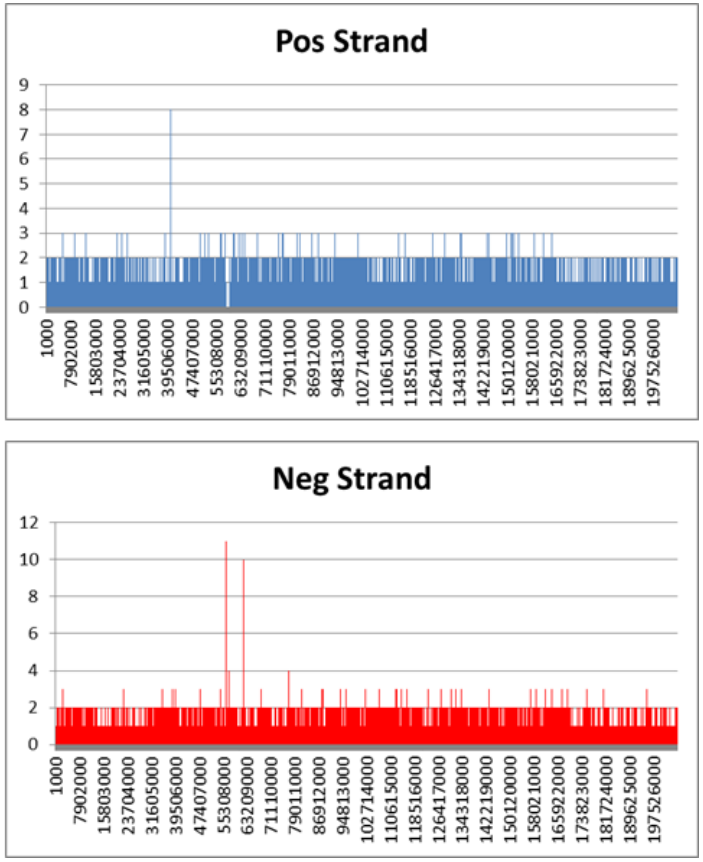
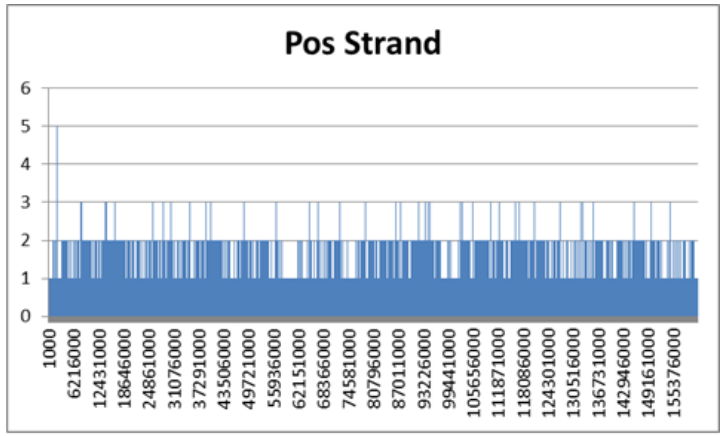


Figure C22: H. Sapiens Chromosome 6 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
161631176	1000	0.9999

Subsequence ID	Subsequence Length
NT_007819.18	58109653
NT_187200.1	2658581
NT_023603.6	449554
NT_187201.1	150232
NT_077528.3	878759
NT_007933.16	96829194
NT_187558.1	157952
NT_187561.1	209586
NT_187559.1	158166
NW_003315922.2	119183
NT_187562.1	1111570
NT_187564.1	271455
NT_187563.1	126434
NT_187560.1	209988
NT_187653.1	190869



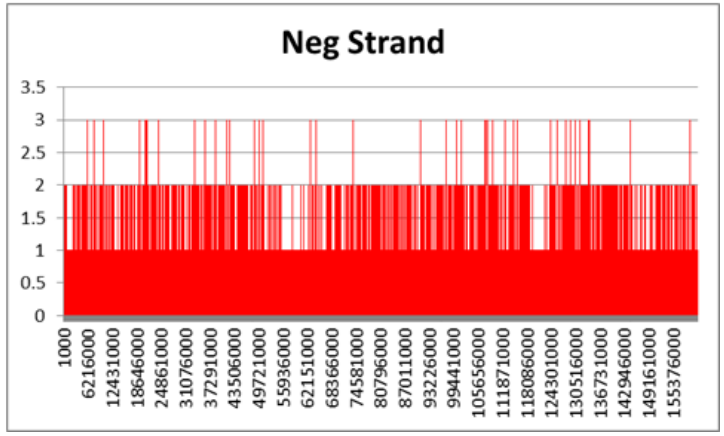


Figure C23: H. Sapiens Chromosome 7 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
149368040	1000	0.9999

Subsequence ID	Subsequence Length
NT_023736.18	7557127
NT_077531.5	4567218
NT_167187.2	31699399
NT_187202.1	1843521
NT_008183.20	39736957
NT_008046.17	59364414
NT_187572.1	145606
NT_187568.1	282736
NT_187565.1	292436
NT_187576.1	985506
NT_187570.1	300230
NT_187577.1	624492
NT_187566.1	141812
NT_187567.1	374415
NT_187574.1	133535
NT_187575.1	36640
NT_187573.1	158983
NT_187571.1	305841
NT_187569.1	132244
NT_187655.1	136959
NT_187654.1	318687
NT_187680.1	229282

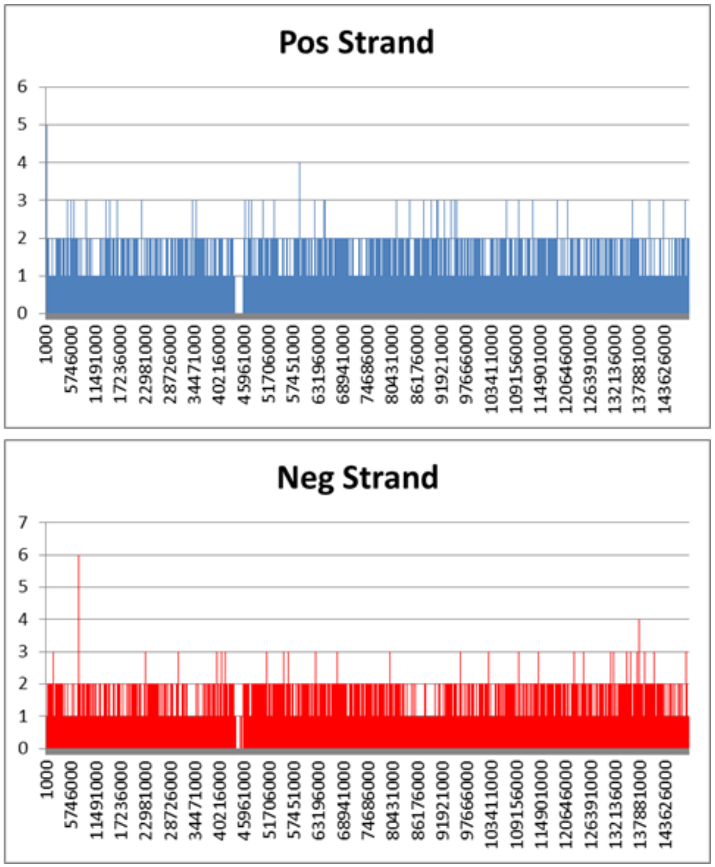


Figure C24: H. Sapiens Chromosome 8 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
123159567	1000	0.9999

Subsequence ID	Subsequence Length
NT_008413.19	43212167
NT_187203.1	153368
NT_187204.1	2128923
NT_113916.2	169874
NT_187205.1	41089
NT_187206.1	174366
NT_078052.5	178079
NT_187207.1	186842
NT_187208.1	216560
NT_187209.1	364370
NT_086755.3	499094
NT_187210.1	159539
NT_187211.1	483893
NT_078067.3	376183
NT_113961.2	246715
NT_078070.4	764920
NT_187212.1	195041
NT_187213.1	220068
NT_078055.5	746196
NT_187214.1	1329165
NT_008470.20	70114165
NT_187372.1	40062
NT_187373.1	38054
NT_187374.1	176845
NT_187375.1	39050
NW_003315928.1	162988
NW_003315929.1	71551
NW_003315930.1	171286
NW_003315931.1	60032
NT_187578.1	439082

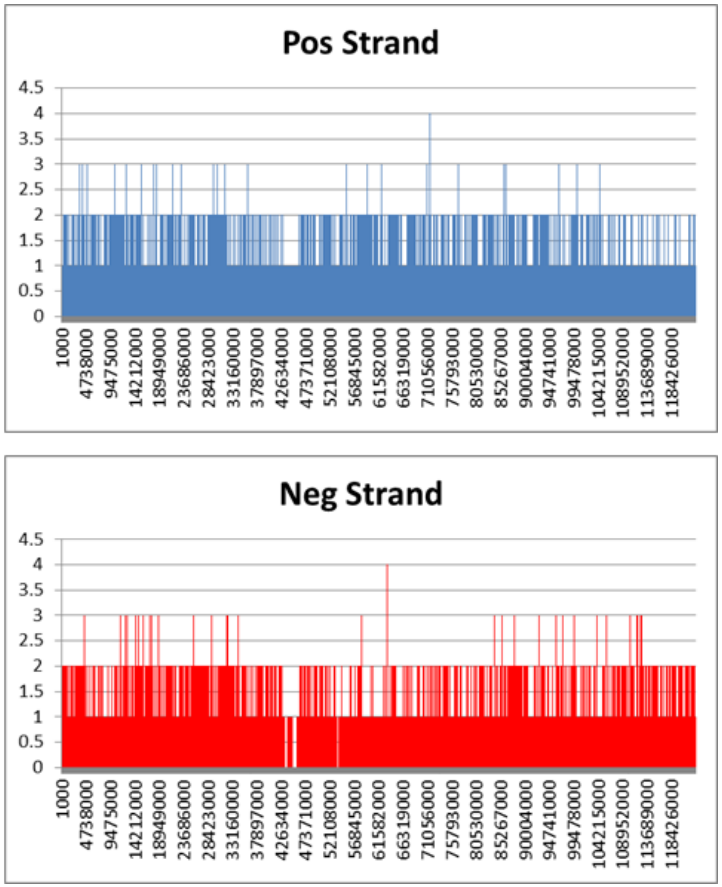


Figure C25: H. Sapiens Chromosome 9 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
134485989	1000	0.9999

Subsequence ID	Subsequence Length
NT_008705.17	39626682
NT_187215.1	249218
NT_187216.1	1561440
NT_187217.1	48180
NT_187218.1	47701
NT_030059.14	92093901
NW_003315934.1	179254
NT_187579.1	181496
NW_003315935.1	309802
NT_187580.1	188315

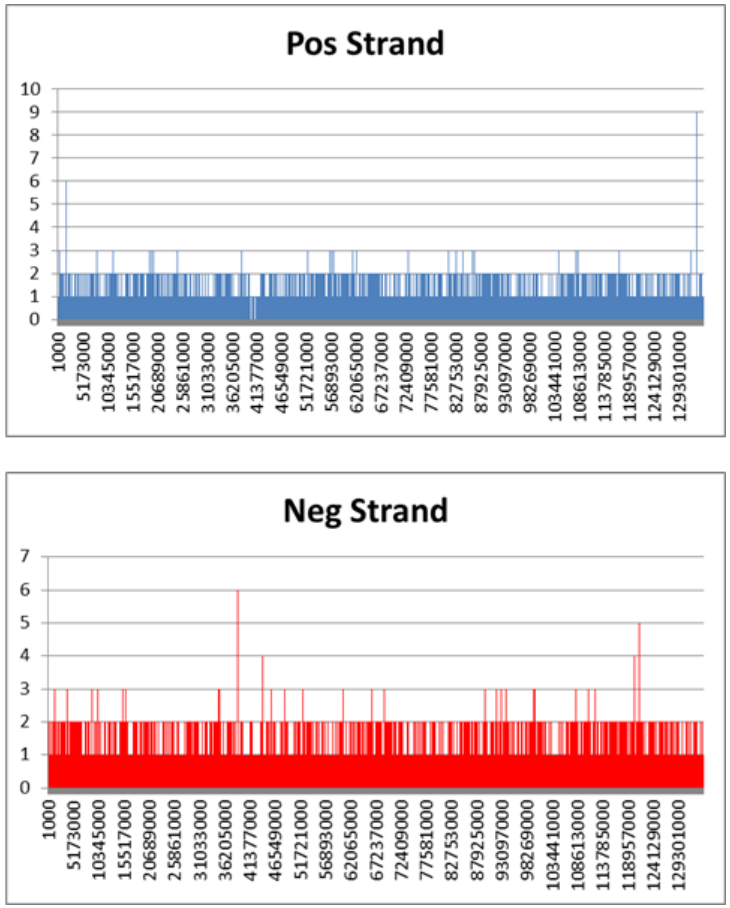


Figure C26: H. Sapiens Chromosome 10 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
136963861	1000	0.9999

Subsequence ID	Subsequence Length
NT_009237.19	50761348
NT_187219.1	11969
NT_187220.1	3251982
NT_187221.1	82575
NT_167190.2	33453128
NT_033899.9	47073726
NT_187376.1	100316
NT_187586.1	210133
NT_187584.1	177092
NT_187585.1	296895
NT_187583.1	204059
NW_003315936.1	154407
NW_003871073.1	200998
NW_003871074.1	191409
NT_187582.1	67707
NT_187581.1	186169
NT_187656.1	106711
NT_187657.1	214625
NT_187681.1	218612

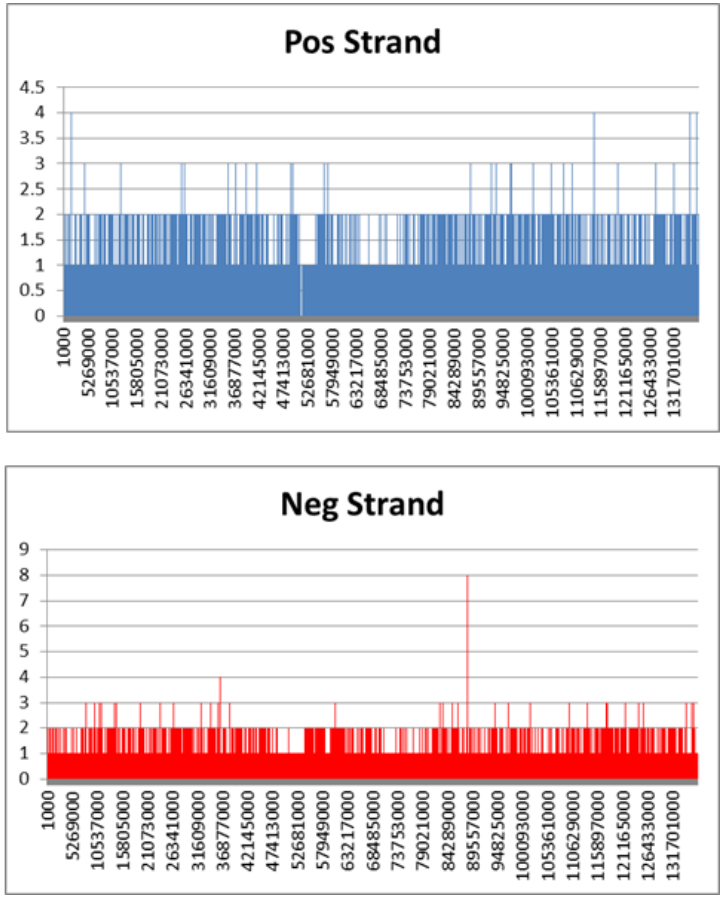


Figure C27: H. Sapiens Chromosome 11 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
135596594	1000	0.9999

Subsequence ID	Subsequence Length
NT_009759.17	7073650
NT_009714.18	27634757
NT_187222.1	47204
NT_187223.1	3474
NT_187224.1	11803
NT_187226.1	2349957
NT_029419.13	94988110
NT_024477.15	1040947
NW_003571049.1	167313
NW_003571050.1	408271
NT_187588.1	40090
NW_003315938.1	120804
NT_187587.1	238139
NW_003315939.2	169178
NW_003315941.1	138655
NW_003315942.2	152874
NT_187590.1	119498
NW_003315940.1	184319
NT_187589.1	76061
NT_187591.1	56134
NT_187658.1	572349

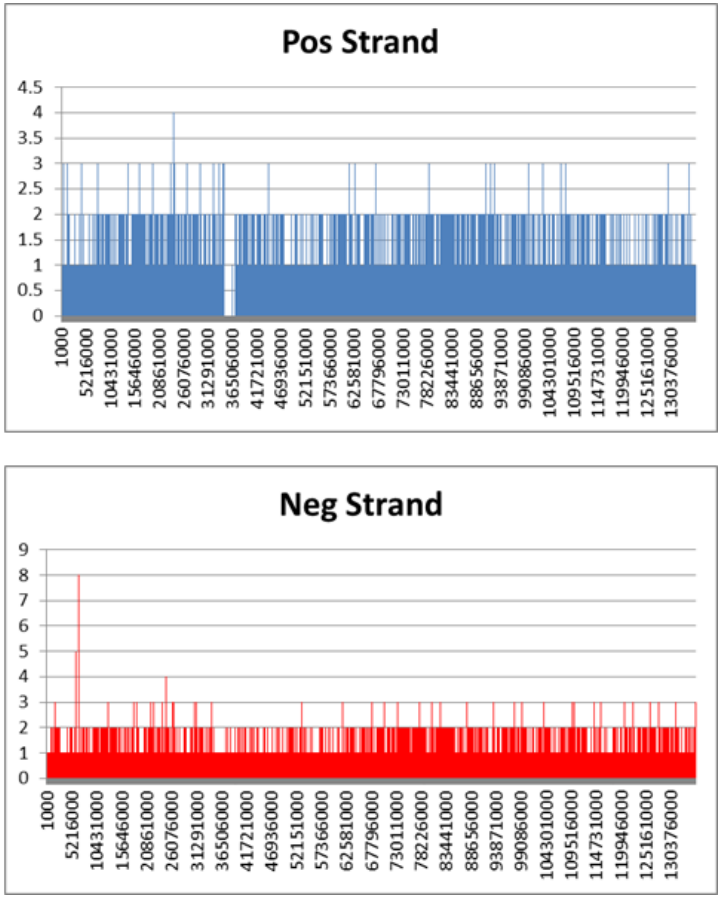
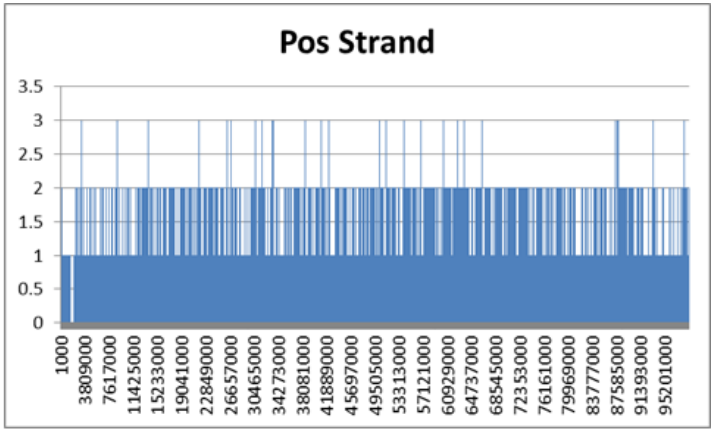


Figure C28: H. Sapiens Chromosome 12 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
99022284	1000	0.9999

Subsequence ID	Subsequence Length
NT_187227.1	22537
NT_187228.1	88022
NT_187229.1	54133
NT_187230.1	63535
NT_187231.1	20670
NT_187232.1	6670
NT_187233.1	3245
NT_187234.1	22561
NT_187235.1	1134211
NT_187236.1	340
NT_187237.1	340
NT_187238.1	1198
NT_187239.1	632586
NT_113923.1	186858
NT_024524.15	67794873
NT_009952.15	25540462
NT_027140.7	1829579
NT_024498.13	631308
NT_187594.1	191684
NT_187593.1	180306
NT_187597.1	103832
NT_187595.1	169134
NT_187592.1	306913
NT_187596.1	37287



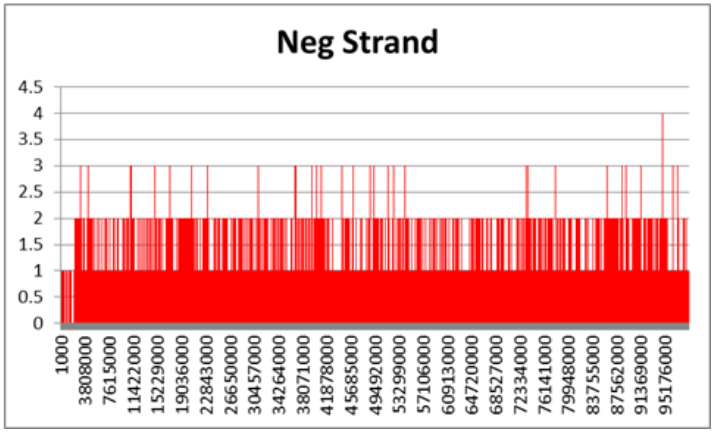


Figure C29: H. Sapiens Chromosome 13 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
95290211	1000	0.9999

Subsequence ID	Subsequence Length
NT_187240.1	22537
NT_187241.1	73893
NT_187242.1	43897
NT_187243.1	88022
NT_187244.1	54133
NT_187245.1	63535
NT_187246.1	20670
NT_187247.1	6670
NT_187248.1	3245
NT_187249.1	22561
NT_187250.1	4185
NT_187251.1	1134211
NT_187252.1	340
NT_187253.1	340
NT_187254.1	1198
NT_187255.1	632586
NT_026437.13	88660195
NT_113796.3	201709
NT_167219.1	211173
NT_187377.1	194050
NT_113888.1	191469
NT_187378.1	38115
NT_187379.1	39555
NT_187380.1	172810
NT_187381.1	43739
NT_187598.1	322166
NT_187601.1	1511111
NT_187599.1	180703
NT_187600.1	1351393

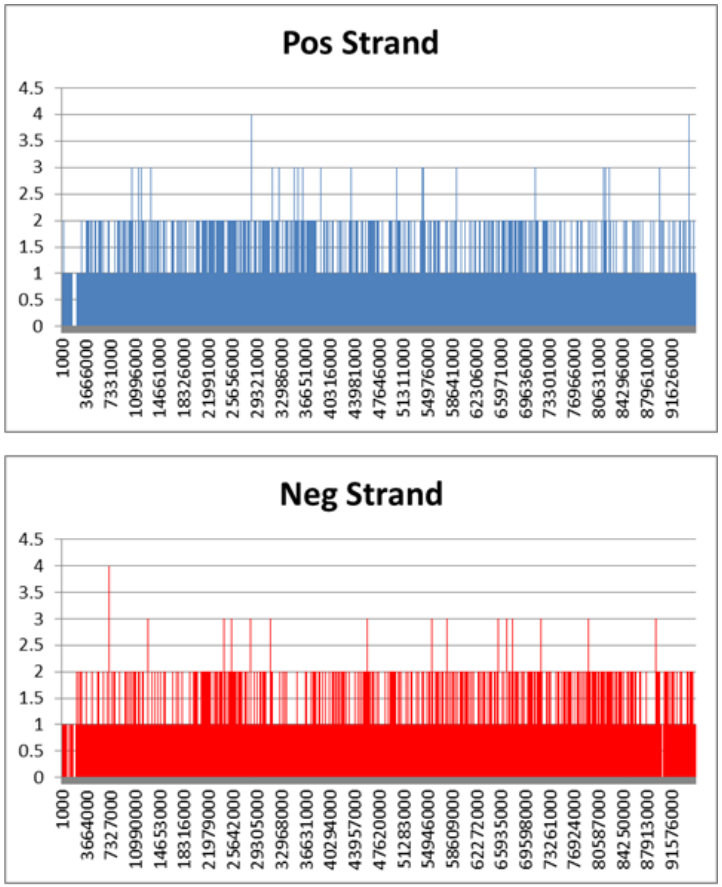


Figure C30: H. Sapiens Chromosome 14 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
93067467	1000	0.9999

Subsequence ID	Subsequence Length
NT_187256.1	83573
NT_187257.1	415278
NT_187258.1	855957
NT_187259.1	1370146
NT_037852.7	2532988
NT_078094.3	868632
NT_010194.18	78704315
NT_187382.1	448248
NT_187602.1	478999
NT_187604.1	263054
NT_187603.1	327382
NW_003315943.1	296527
NT_187605.1	244917
NW_003315944.2	388773
NT_187606.1	430880
NT_187659.1	196384
NT_187660.1	5161414

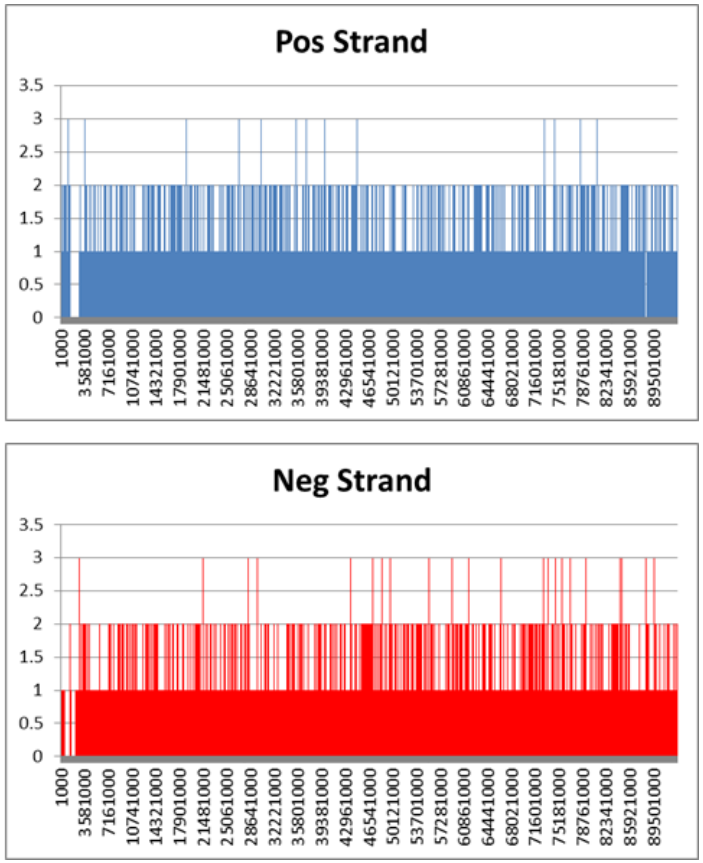


Figure C31: H. Sapiens Chromosome 15 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
87163670	1000	0.9999

Subsequence ID	Subsequence Length
NT_010393.17	18426486
NT_187260.1	15802843
NT_187261.1	182181
NT_024773.12	1689648
NT_187262.1	23302
NT_187263.1	3006
NT_187264.1	1928003
NT_187265.1	14913
NT_010498.16	43847663
NT_187383.1	1872759
NT_187610.1	134193
NT_187609.1	63982
NT_187608.1	232857
NT_187607.1	2659700
NW_003315945.1	192462
NW_003315946.1	89672

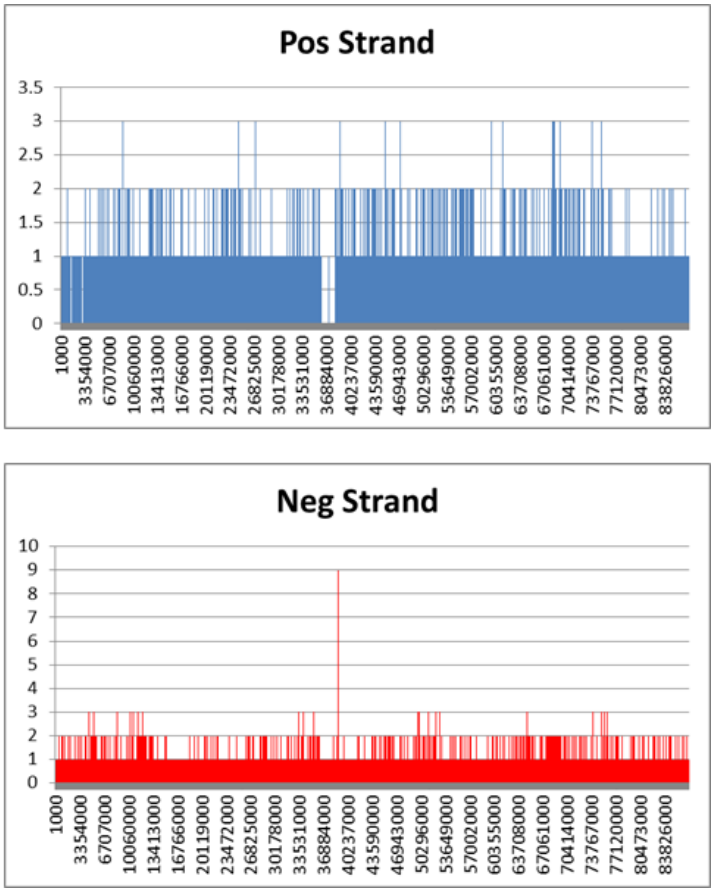


Figure C32: H. Sapiens Chromosome 16 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
92639257	1000	0.9999

Subsequence ID	Subsequence Length
NT_024972.9	388188
NT_010718.17	21503074
NT_024862.15	721618
NT_187266.1	381239
NT_187267.1	3371615
NT_187268.1	49431
NT_187269.1	263990
NT_187270.1	5626
NT_010783.16	54806562
NT_010663.16	1454899
NT_113930.2	185591
NT_187384.1	280839
NT_187385.1	112551
NW_003315952.3	375691
NT_187613.1	391357
NT_187611.1	196688
NT_187614.1	2877074
NW_003871091.1	278131
NW_003871092.1	70345
NW_003315953.2	133151
NT_167251.2	1821992
NW_003315954.1	223995
NT_187615.1	235827
NT_187616.1	108763
NW_003315955.1	90219
NT_187612.1	178921
NT_187662.1	137721
NT_187664.1	157099
NT_187661.1	325800
NW_003871093.1	88070
NT_187663.1	1423190

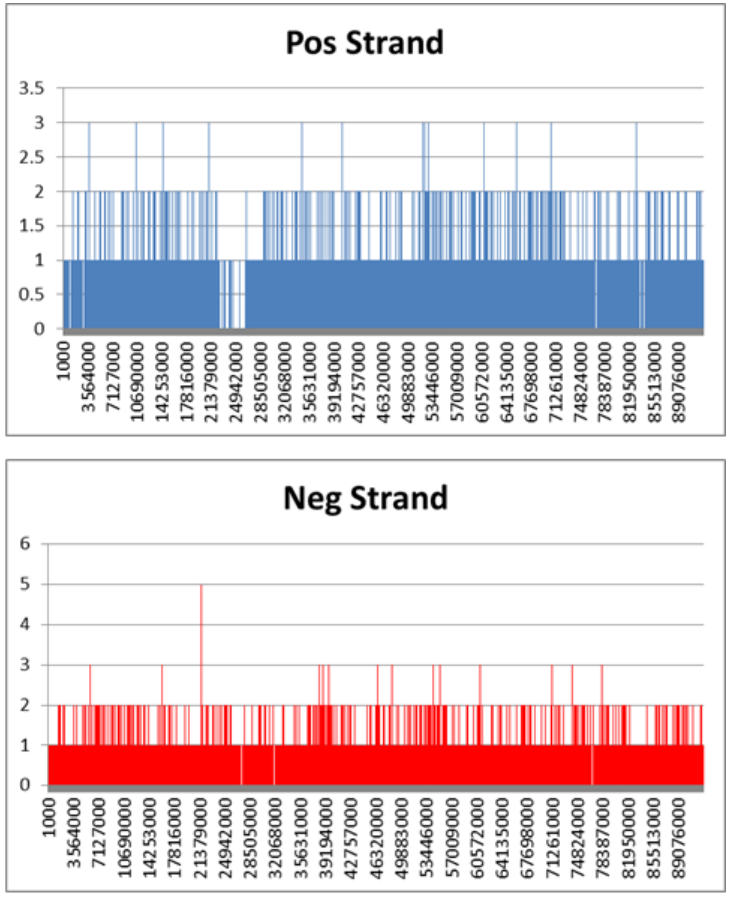
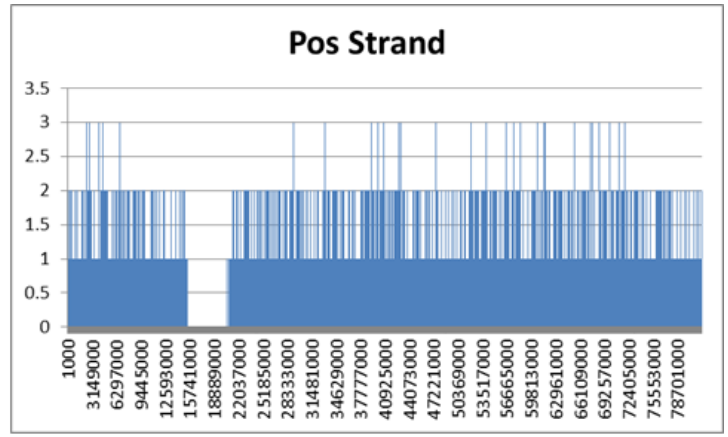


Figure C33: H. Sapiens Chromosome 17 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
81848839	1000	0.9999

Subsequence ID	Subsequence Length
NT_010859.15	15400899
NT_187271.1	319478
NT_187272.1	17278
NT_187273.1	4763584
NT_187274.1	21096
NT_187275.1	20412
NT_187276.1	93042
NT_187277.1	39636
NT_187278.1	76958
NT_187279.1	26514
NT_187280.1	21409
NT_010966.15	59352079
NT_187618.1	111737
NW_003315956.1	289831
NW_003315959.1	164789
NW_003315960.1	198278
NW_003315957.1	104552
NW_003315958.1	167950
NW_003315961.1	159547
NT_187617.1	167999
NT_187665.1	174061
NT_187666.1	157710



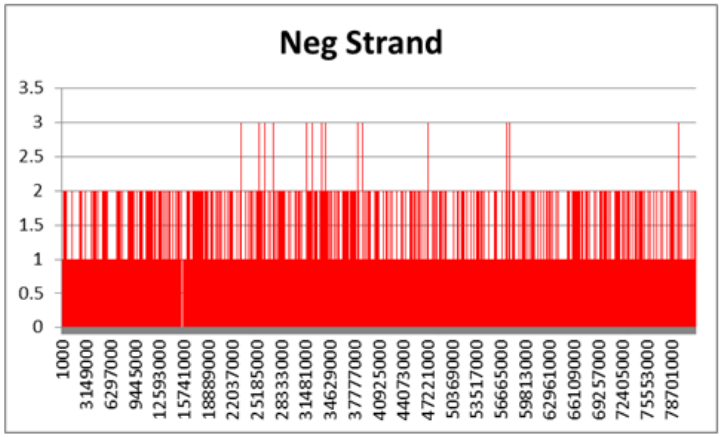


Figure C34: H. Sapiens Chromosome 18 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Subsequence ID	Subsequence Length
NT_011295.12	24388980
NT_187281.1	53672
NT_187282.1	338504
NT_187283.1	4434
NT_187284.1	12699
NT_187285.1	2282185
NT_011109.17	31366742
NT_187622.1	61734
NT_187621.1	52969
NW_003315962.1	385657
NW_003315964.2	170222
NW_003315965.1	188024
NW_003315963.1	155864
NT_187619.1	43156
NT_187620.1	233762
NW_003571054.1	987716
NT_187636.1	248807
NT_187637.1	170399
NT_187638.1	157053
NT_187639.1	171027
NT_187640.1	204239
NT_187641.1	209512
NT_187642.1	155532
NT_187643.1	170698
NT_187644.1	184499
NT_187645.1	170680
NW_003571055.2	729520
NT_187668.1	205194
NT_187669.1	170665
NT_187670.1	184516
NT_187671.1	190932
NT_187672.1	123111
NT_187673.1	170701
NT_187674.1	198005
NT_187675.1	282224
NT_187676.1	187935
NT_187677.1	189352
NW_003571056.2	1064304
NT_187683.1	186203
NT_187684.1	200773
NT_187685.1	170148
NT_187686.1	215732
NT_187687.1	170537
NT_113949.2	177381
NW_003571057.2	1091841
NW_003571058.2	1066390
NW_003571059.2	1002683
NW_003571060.1	987100
NW_003571061.2	796479
NT_187693.1	1066800

Total Length	Bin Size	Filter Value
73397292	1000	0.9999

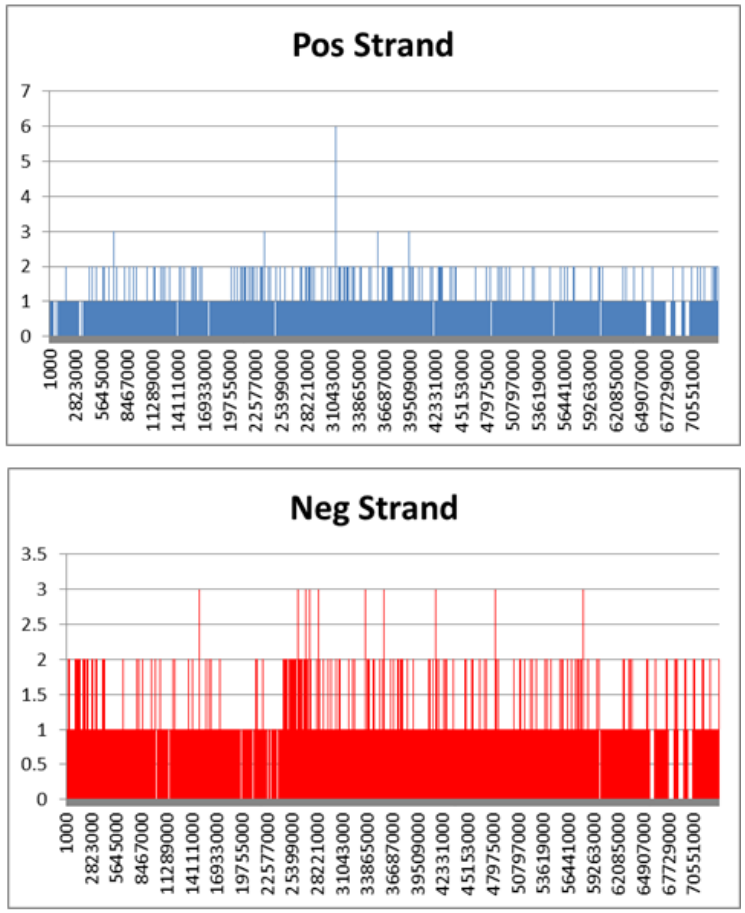


Figure C35: H. Sapiens Chromosome 19 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
64531272	1000	0.9999

Subsequence ID	Subsequence Length
NT_011387.9	26326232
NT_187286.1	150723
NT_187287.1	20990
NT_187288.1	1886394
NT_187289.1	14258
NT_187290.1	47956
NT_187291.1	89142
NT_187292.1	1713
NT_187293.1	80766
NT_113914.2	396719
NT_187294.1	78875
NT_187295.1	679982
NT_187296.1	32454
NT_187297.1	120944
NT_025215.5	336780
NT_187298.1	305821
NT_187299.1	189610
NT_011362.11	33282659
NW_003315966.2	128386
NT_187623.1	118774
NT_187625.1	58661
NT_187624.1	183433

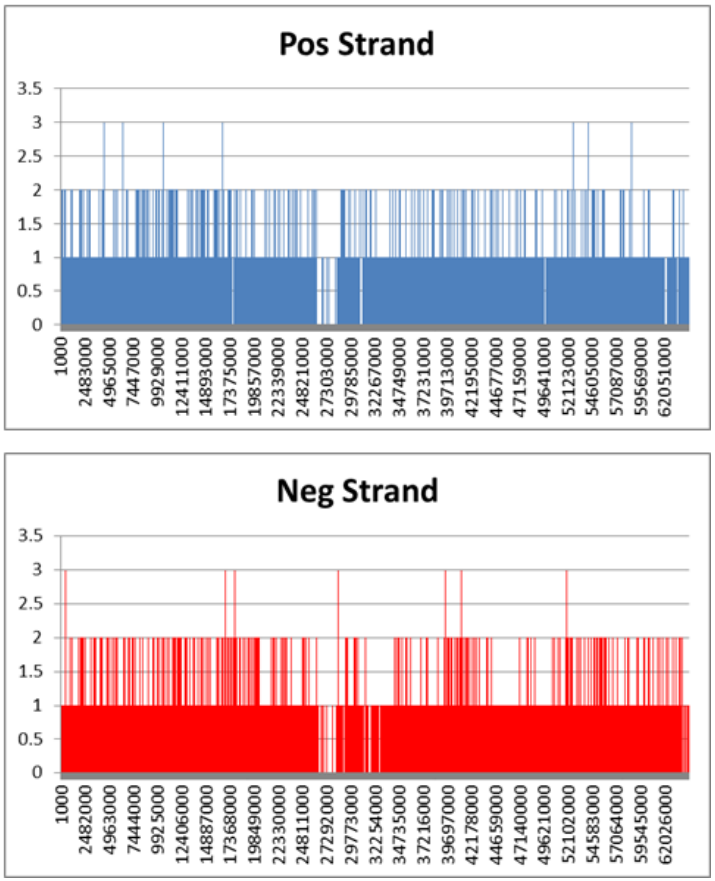


Figure C36: H. Sapiens Chromosome 20 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
40988574	1000	0.9999

Subsequence ID	Subsequence Length
NT_187300.1	156246
NT_187301.1	177312
NT_187302.1	5454
NT_187303.1	128584
NT_187304.1	118413
NT_187305.1	70584
NT_187306.1	194778
NT_187307.1	165887
NT_187308.1	152923
NT_187309.1	108904
NT_187310.1	145134
NT_187311.1	165308
NT_187312.1	128338
NT_187313.1	123025
NT_187314.1	142810
NT_187315.1	122046
NT_187316.1	134093
NT_187317.1	161132
NT_187318.1	161389
NT_113952.1	184355
NT_113954.1	129889
NT_113958.2	209483
NT_113953.1	131056
NT_187319.1	642725
NT_187320.1	4459
NT_029490.5	490233
NT_187321.1	22537
NT_187322.1	88022
NT_187323.1	54133
NT_187324.1	63535
NT_187325.1	20670
NT_187326.1	6670
NT_187327.1	3245
NT_187328.1	22561
NT_187329.1	1134211
NT_187330.1	340
NT_187331.1	340
NT_187332.1	1198
NT_187333.1	632586
NT_011512.12	33734175
NW_003315967.2	63917
NT_187628.1	166743
NT_187627.1	143900
NW_003315968.2	201197
NW_003315969.2	74653
NW_003315970.2	116689
NT_187626.1	82692

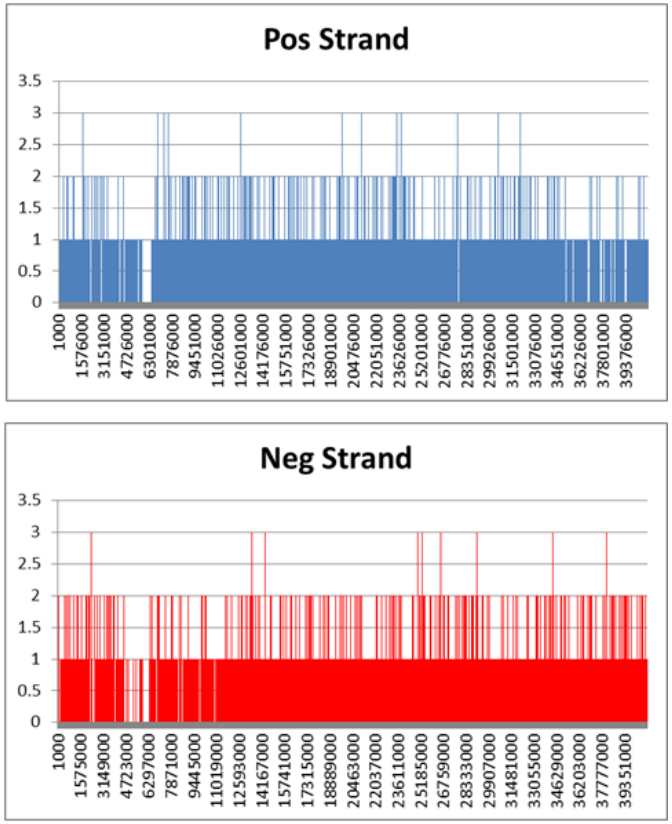
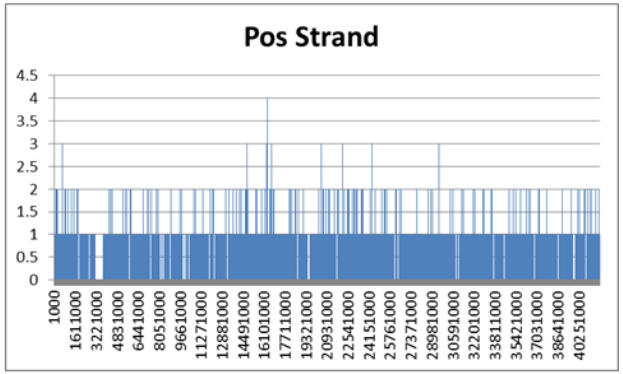


Figure C37: H. Sapiens Chromosome 21 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Subsequence ID	Subsequence Length
NT_167212.2	274643
NT_167238.1	39929
NT_167235.1	42152
NT_167231.2	52263
NT_167230.1	41934
NT_187334.1	167135
NT_167227.2	69281
NT_187335.1	83951
NT_167237.1	43341
NT_187336.1	202926
NT_187337.1	198033
NT_187338.1	163102
NT_187339.1	153040
NT_167229.1	34474
NT_167234.1	41933
NT_167239.1	36651
NT_187340.1	22537
NT_187341.1	43897
NT_187342.1	88022
NT_187343.1	54133
NT_187344.1	63535
NT_187345.1	20670
NT_187346.1	6670
NT_187347.1	3245
NT_187348.1	22561
NT_187349.1	4185
NT_187350.1	1134211
NT_187351.1	340
NT_187352.1	340
NT_187353.1	1198
NT_187354.1	632586
NT_187355.1	3084811
NT_187356.1	94384
NT_187357.1	176051
NT_011520.13	31264301
NT_011526.8	833103
NT_187386.1	150754
NT_187387.1	41543
NT_187388.1	179772
NT_187389.1	165050
NT_187390.1	42811
NT_187391.1	181920
NT_187392.1	103838
NT_187393.1	99375
NT_187394.1	73985
NT_187629.1	259914
NT_187632.1	186262
NT_187633.1	304135
NT_187630.1	263666
NT_187631.1	101331
NW_003315972.2	96924
NW_003315971.2	162811
NW_004504305.1	74013
NT_187682.1	176103

Total Length	Bin Size	Filter Value
41859775	1000	0.9999



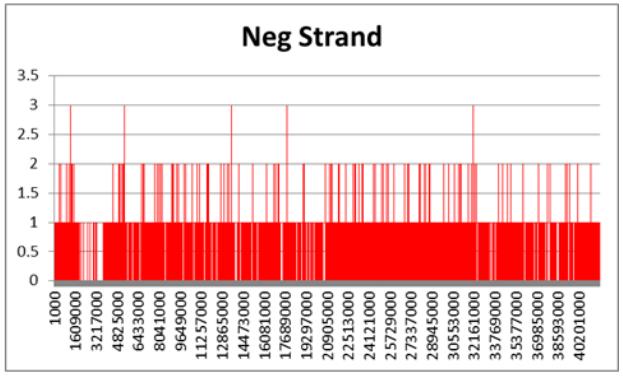


Figure C38: H. Sapiens Chromosome 22 Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
156065968	1000	0.9999

Subsequence ID	Subsequence Length
NT_187358.1	1939345
NT_167197.2	34966268
NT_079573.5	12943127
NT_011630.15	8276615
NT_187359.1	3806963
NT_011651.18	51818656
NT_028405.13	2226581
NT_011786.17	27830040
NT_011681.17	11555289
NT_187634.1	284869
NT_187635.1	144206
NT_187667.1	274009

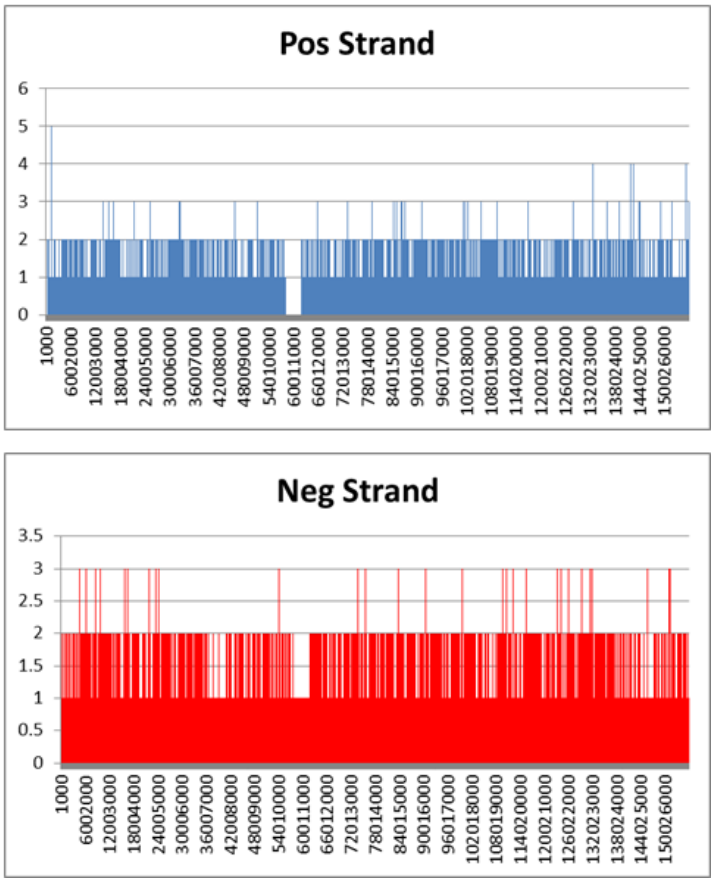
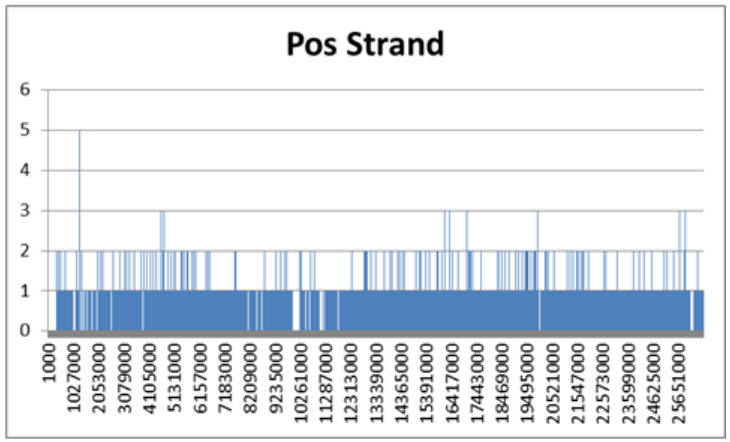


Figure C39: H. Sapiens Chromosome X Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Total Length	Bin Size	Filter Value
26695006	1000	0.9999

Subsequence ID	Subsequence Length
NT_167201.2	1939345
NT_167205.2	648485
NT_011896.10	6276129
NT_086998.2	296105
NT_011878.9	813231
NT_187360.1	227095
NT_087001.2	100153
NT_113819.2	848710
NT_011875.13	10146379
NT_011903.12	4867933
NT_025975.3	98295
NT_091573.1	66393
NT_167206.2	329513
NT_187395.1	37240



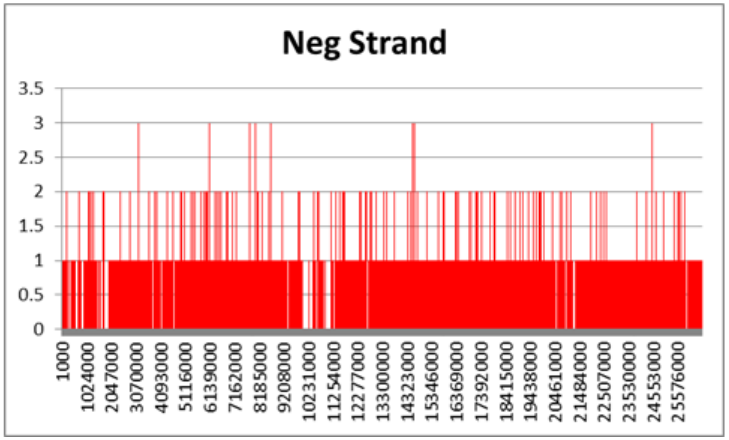


Figure C40: H. Sapiens Chromosome Y Hit Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

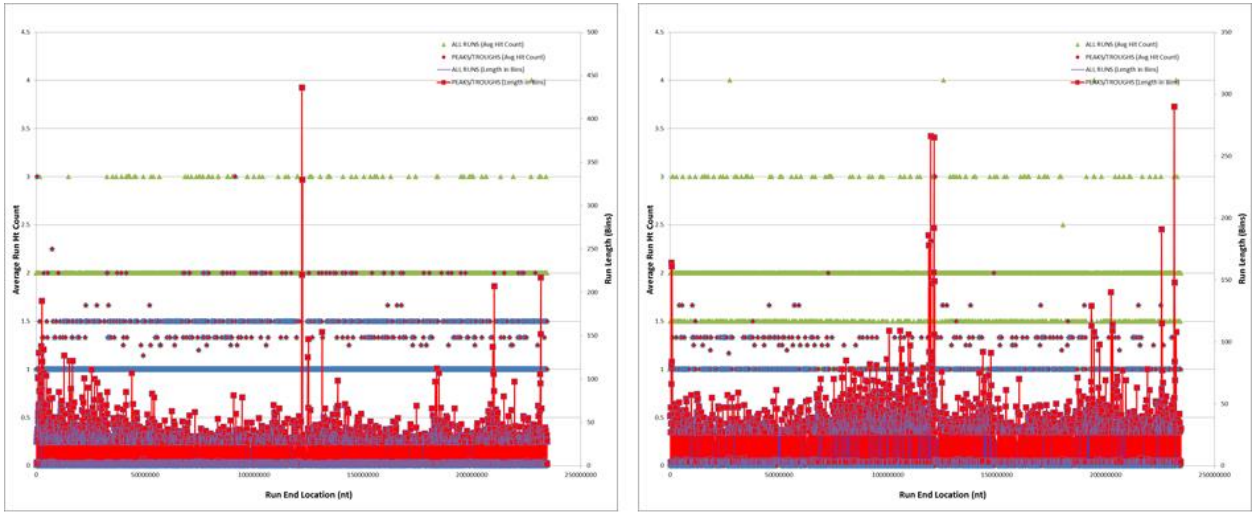
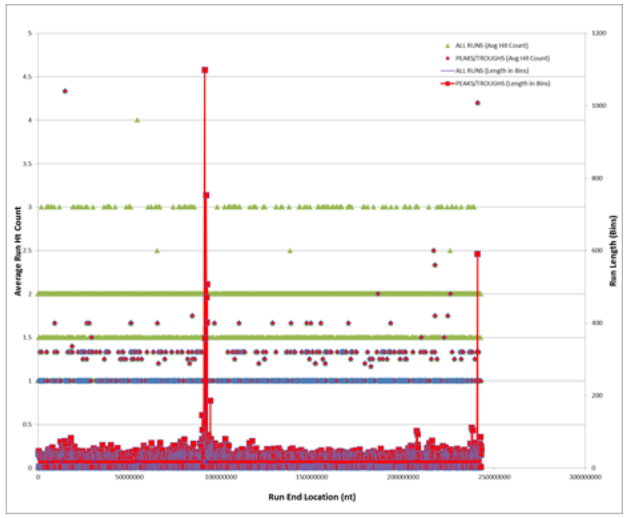


Figure C41: H. Sapiens Chromosome 1 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Chromosome 2 Positive Strand



Chromosome 2 Negative Strand

Figure C42: H. Sapiens Chromosome 2 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

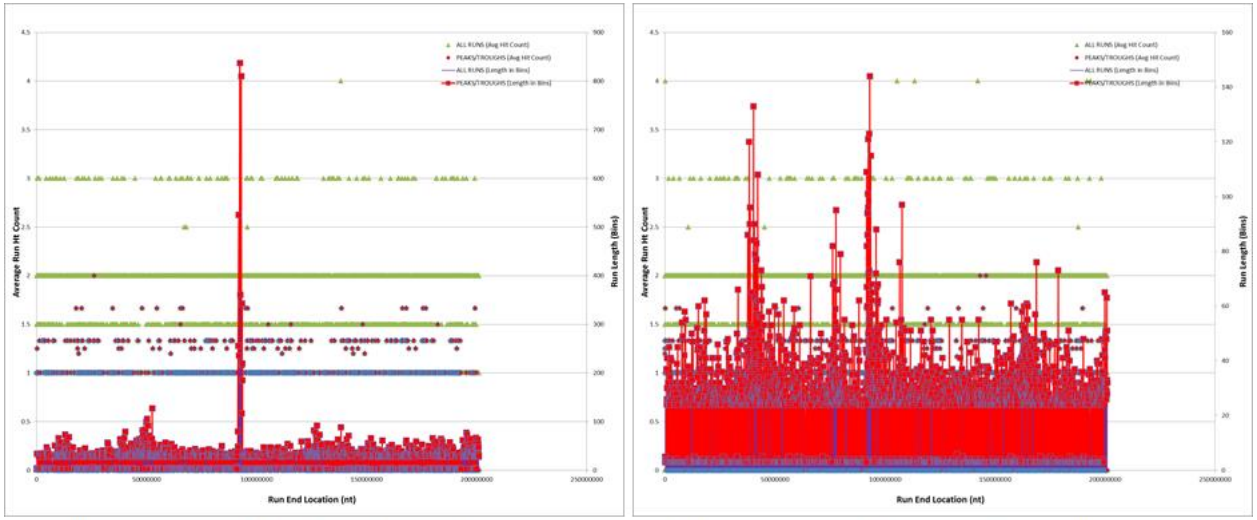
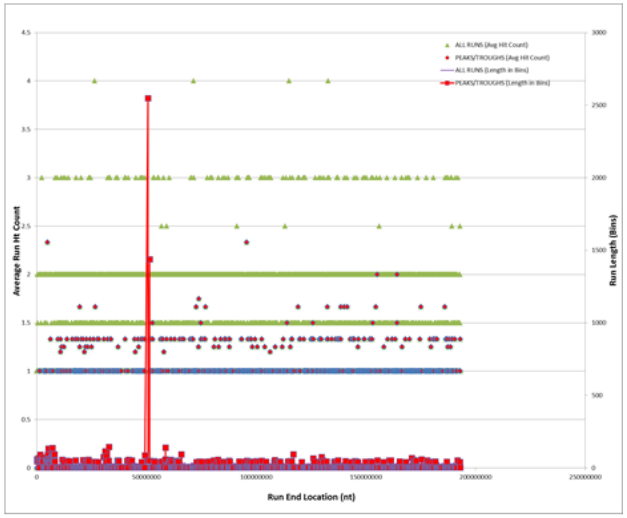


Figure C43: H. Sapiens Chromosome 3 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

Chromosome 4 Positive Strand



Chromosome 4 Negative Strand

Figure C44: H. Sapiens Chromosome 4 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

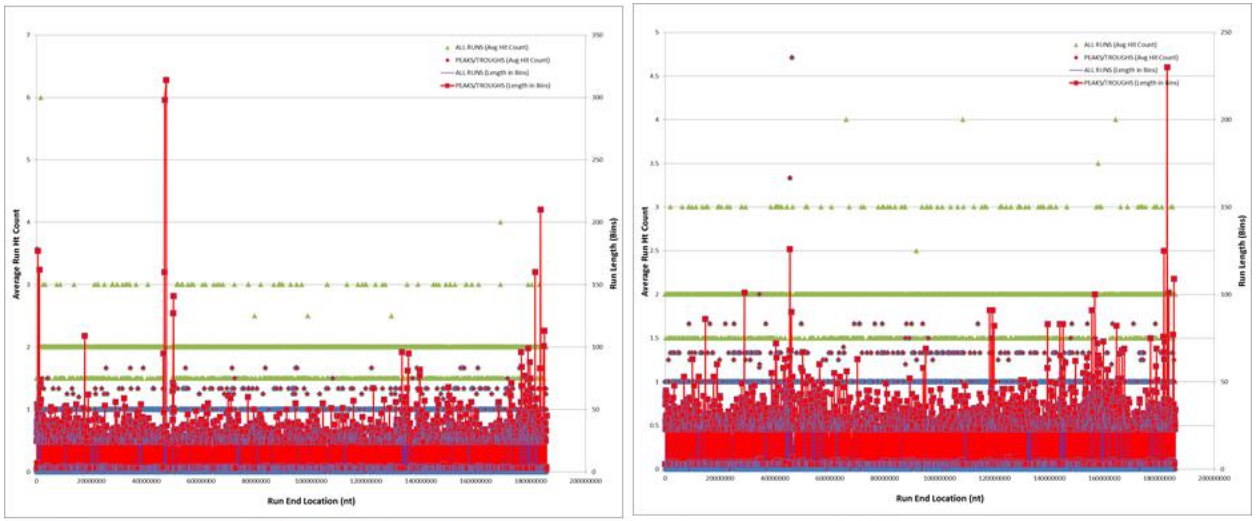


Figure C45: H. Sapiens Chromosome 5 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

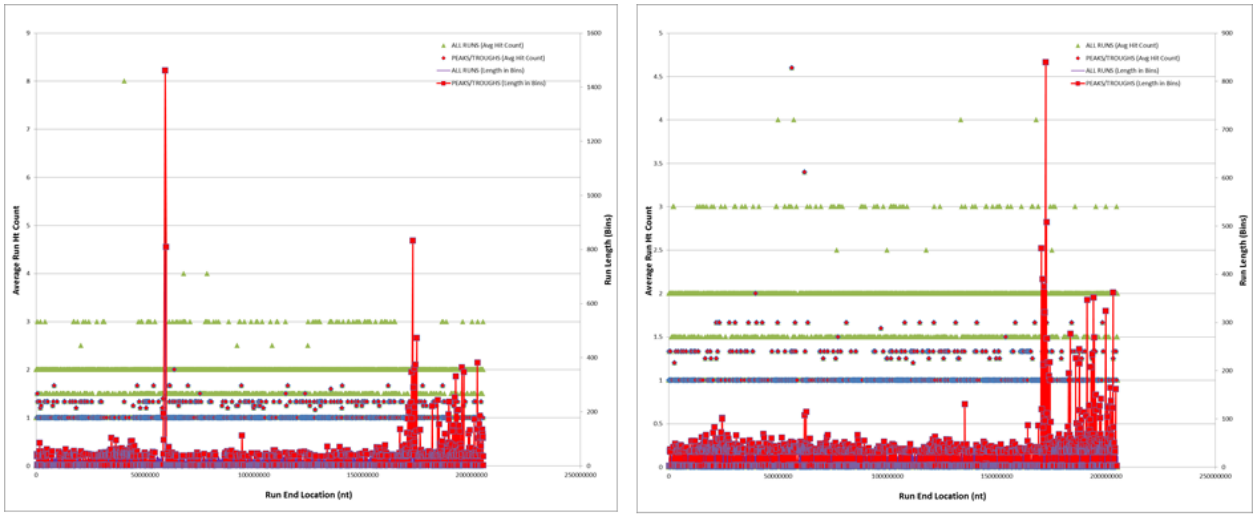


Figure C46: H. Sapiens Chromosome 6 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

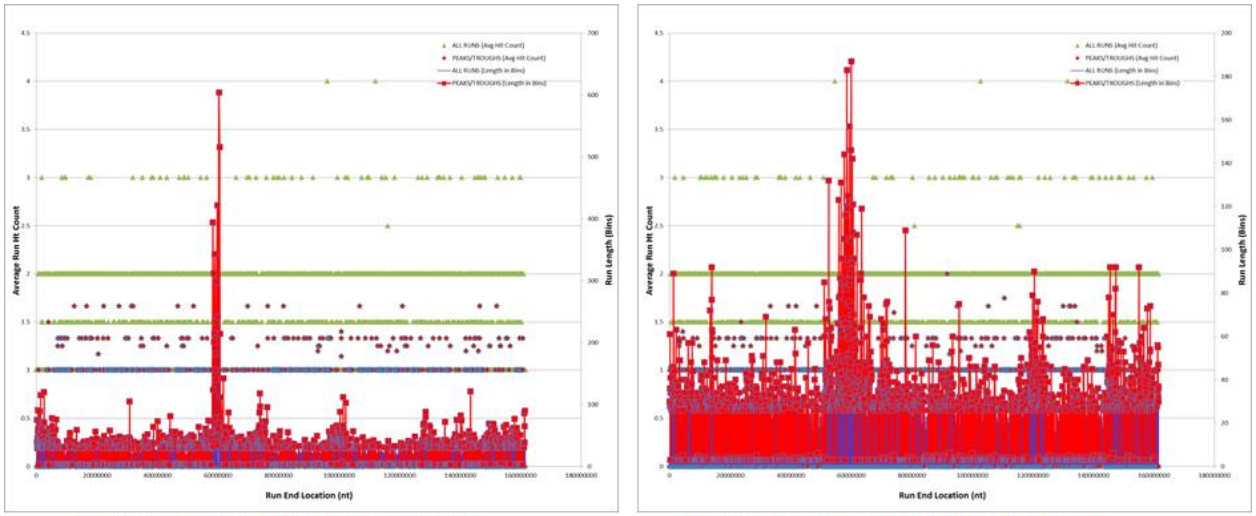


Figure C47: H. Sapiens Chromosome 7 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

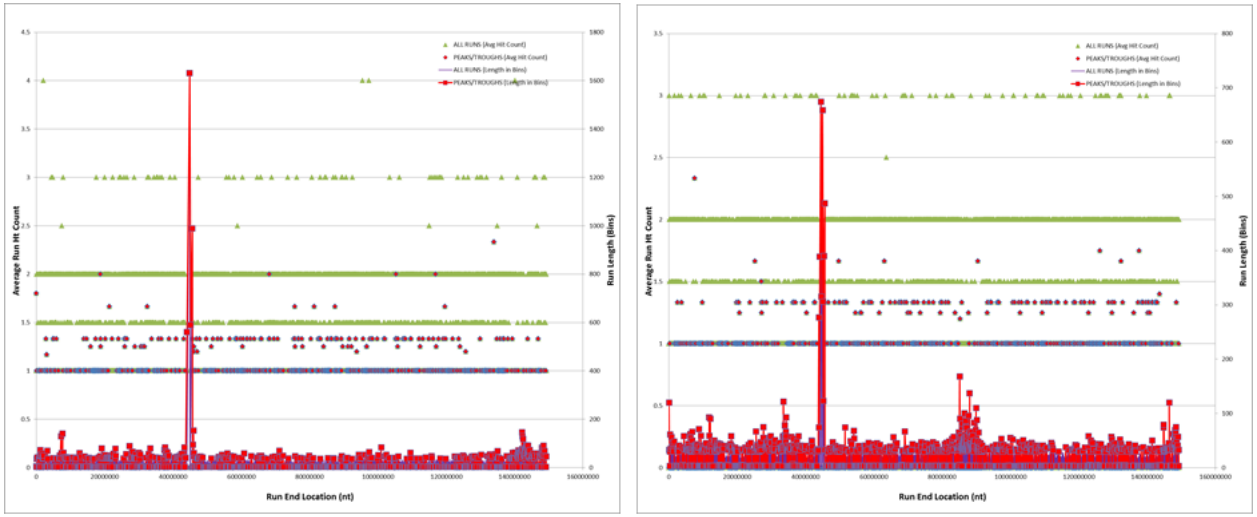


Figure C48: H. Sapiens Chromosome 8 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

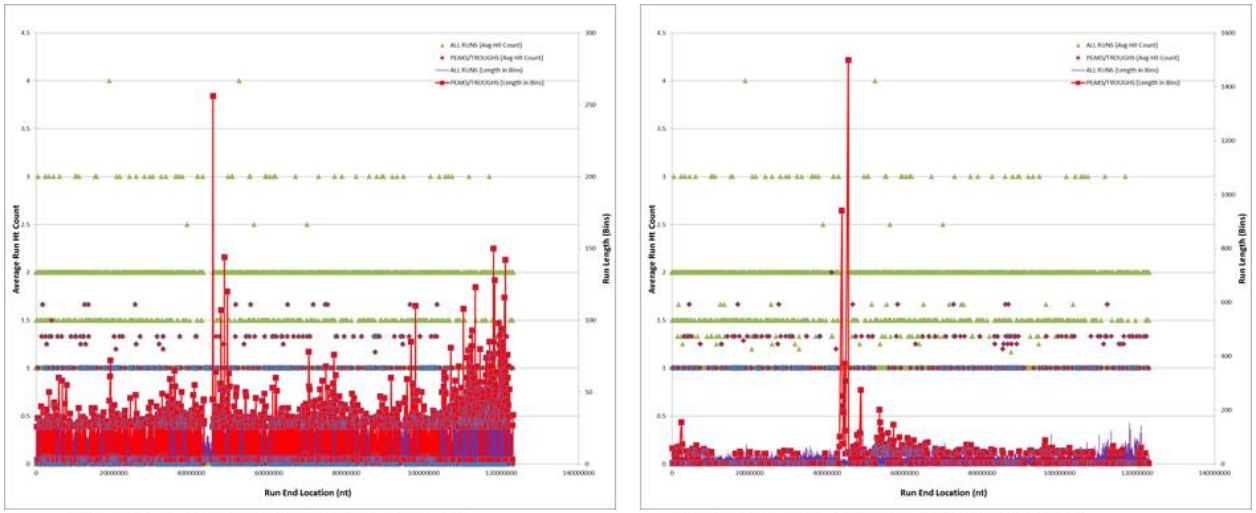


Figure C49: H. Sapiens Chromosome 9 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

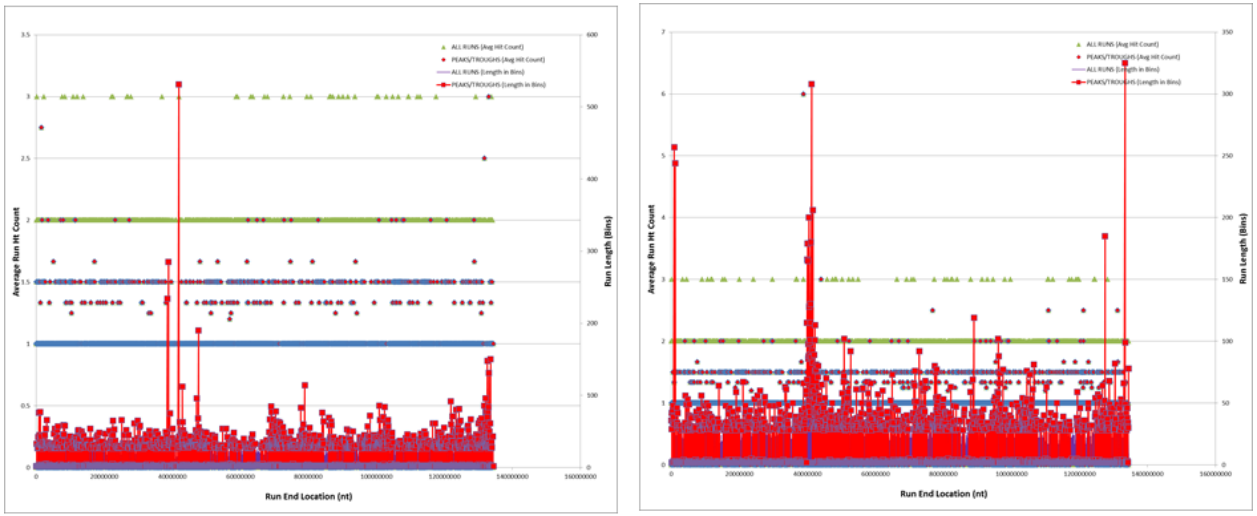


Figure C50: H. Sapiens Chromosome 10 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

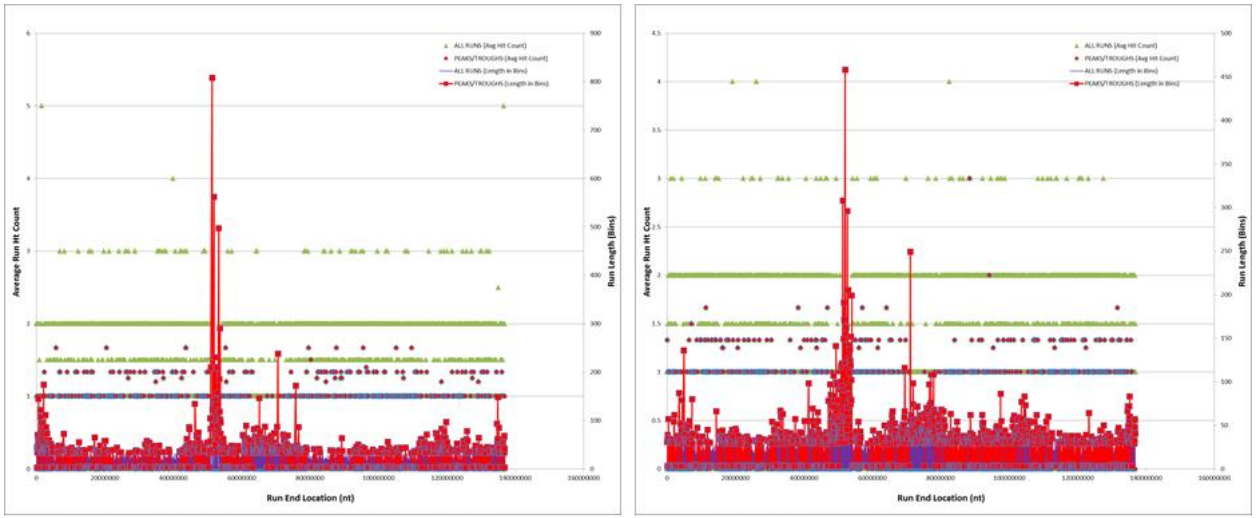


Figure C51: H. Sapiens Chromosome 11 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

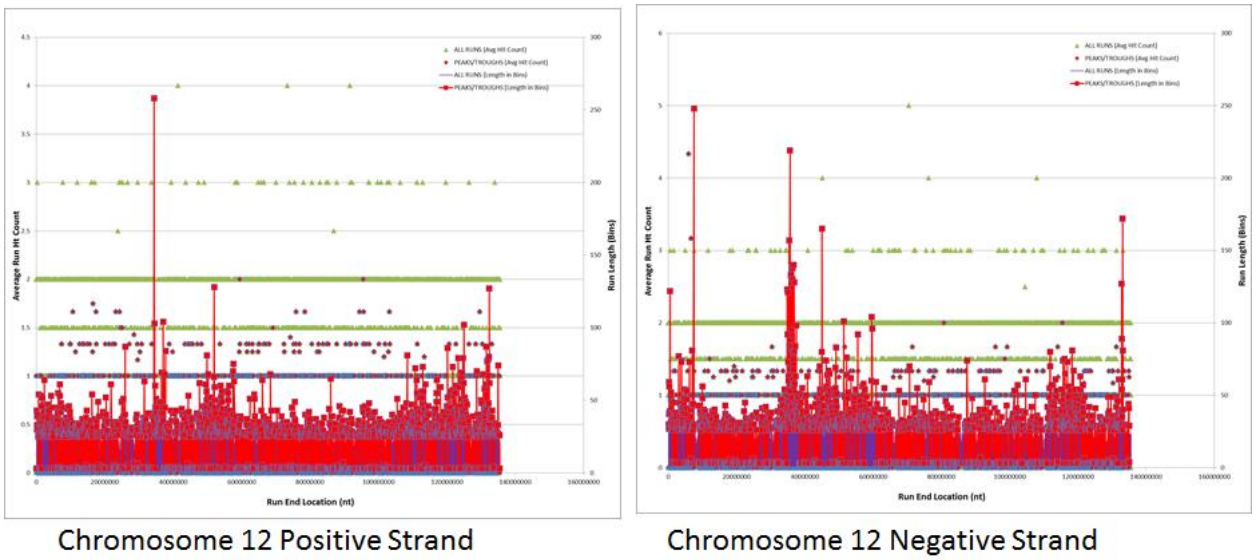


Figure C52: H. Sapiens Chromosome 12 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

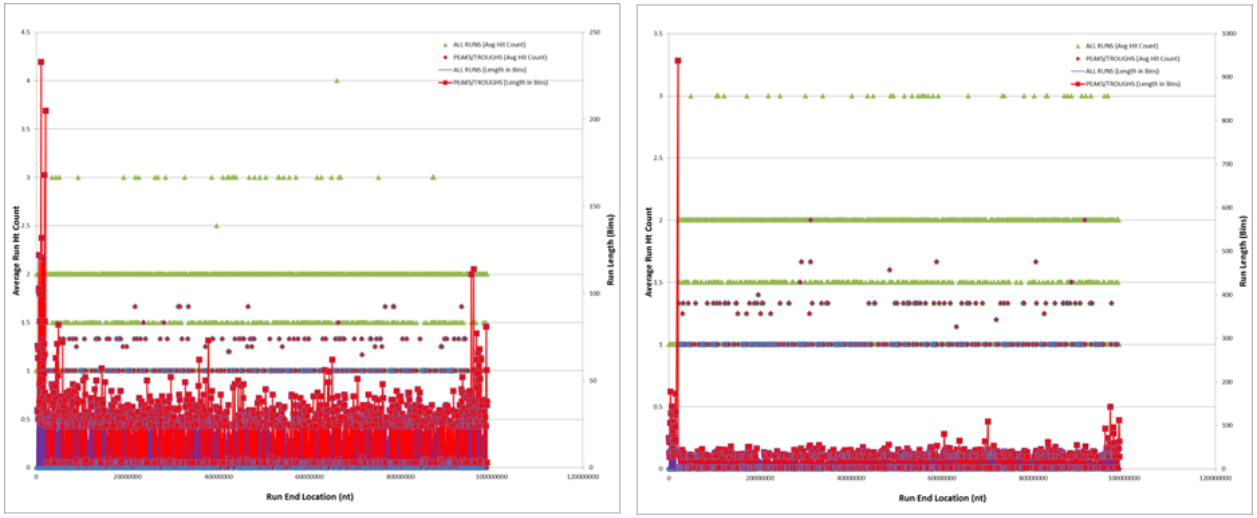


Figure C53: H. Sapiens Chromosome 13 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

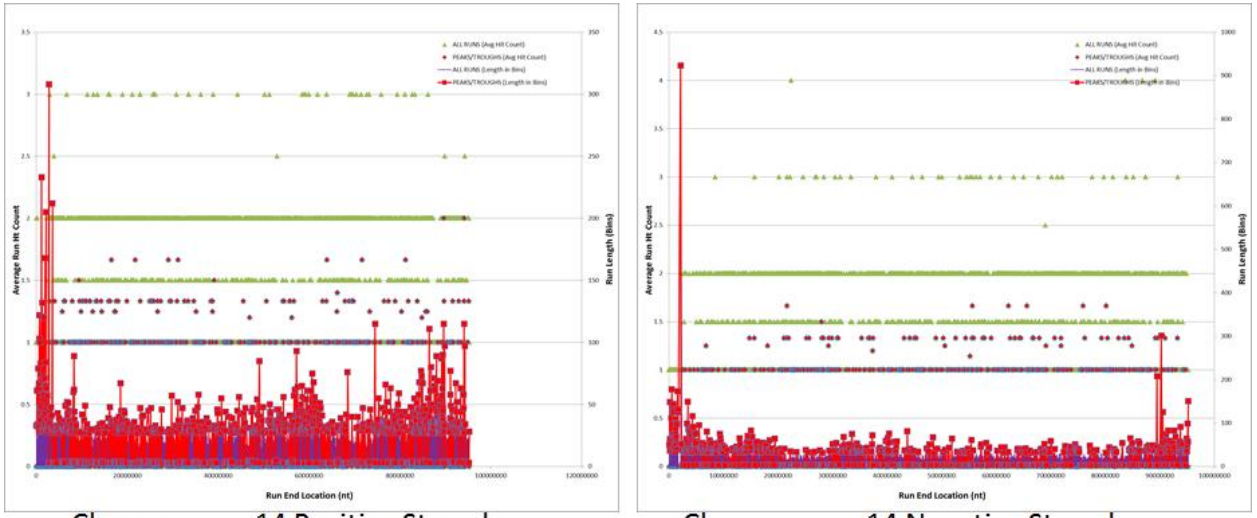


Figure C54: H. Sapiens Chromosome 14 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

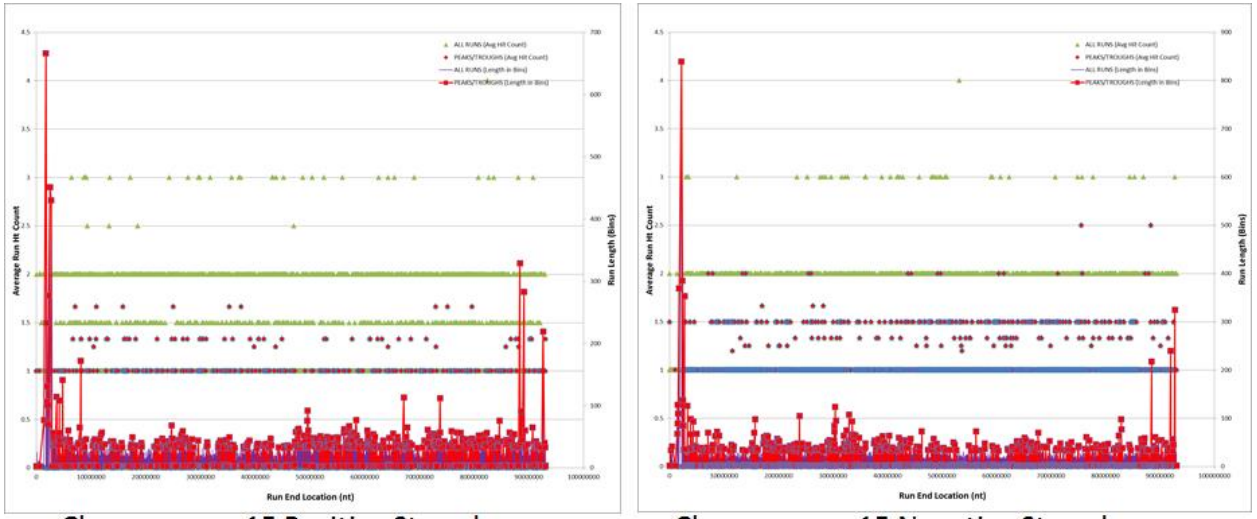


Figure C55: H. Sapiens Chromosome 15 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

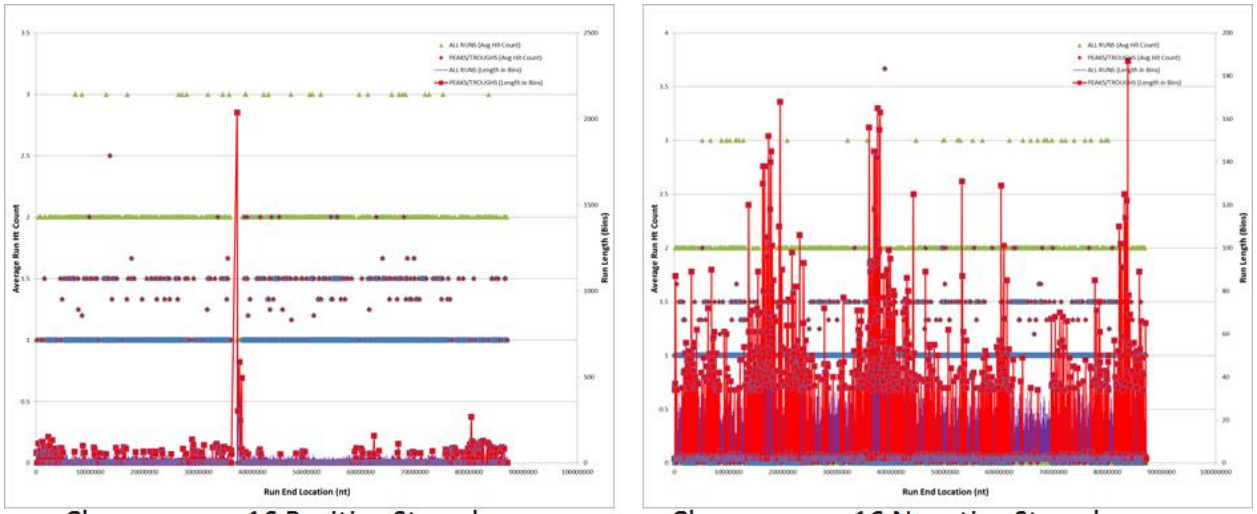


Figure C56: H. Sapiens Chromosome 16 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

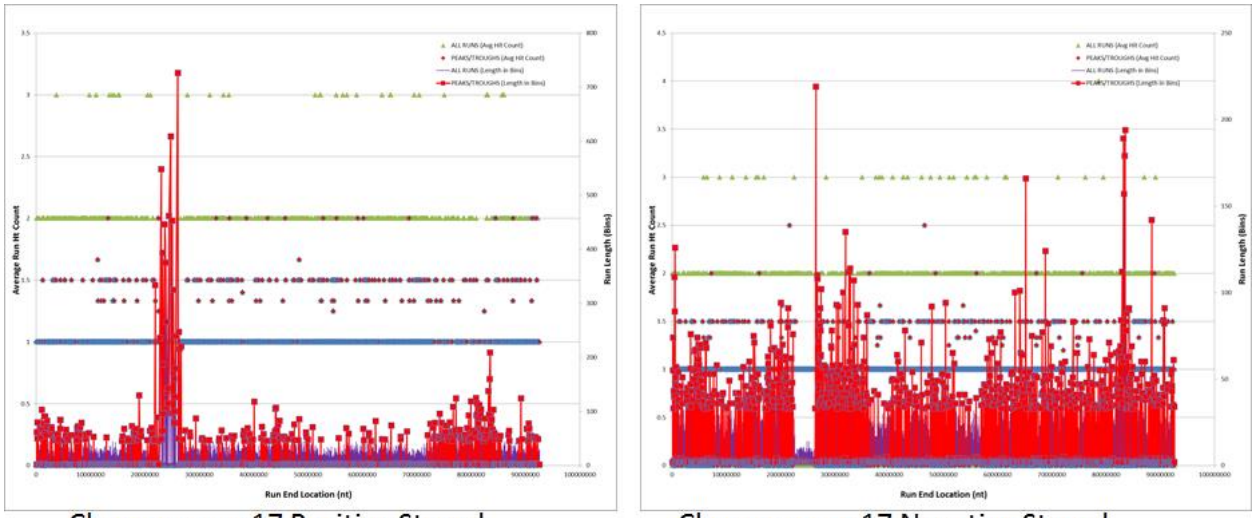


Figure C57: H. Sapiens Chromosome 17 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

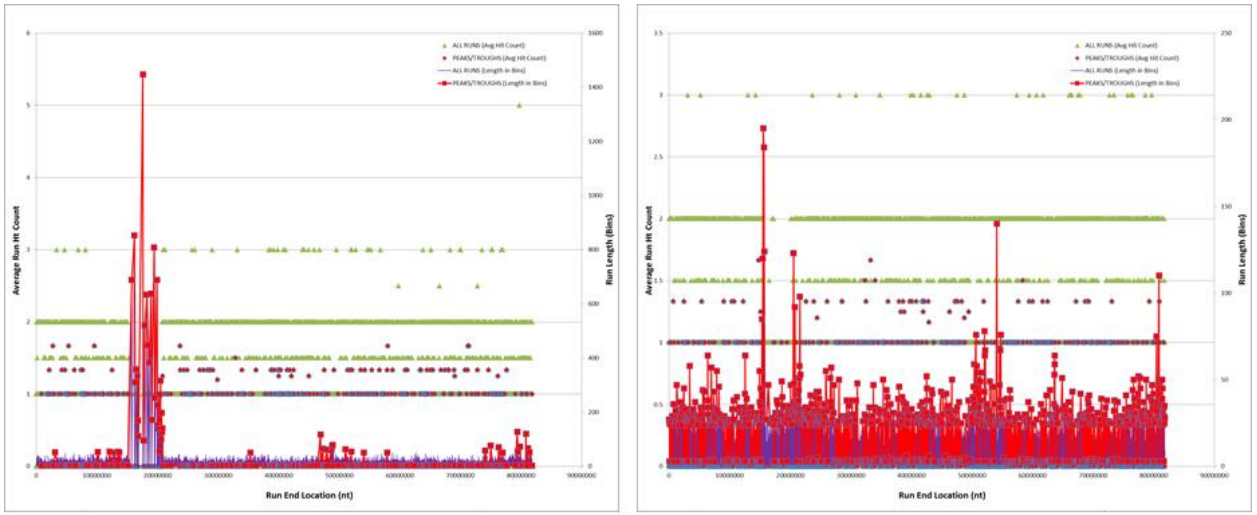


Figure C58: H. Sapiens Chromosome 18 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

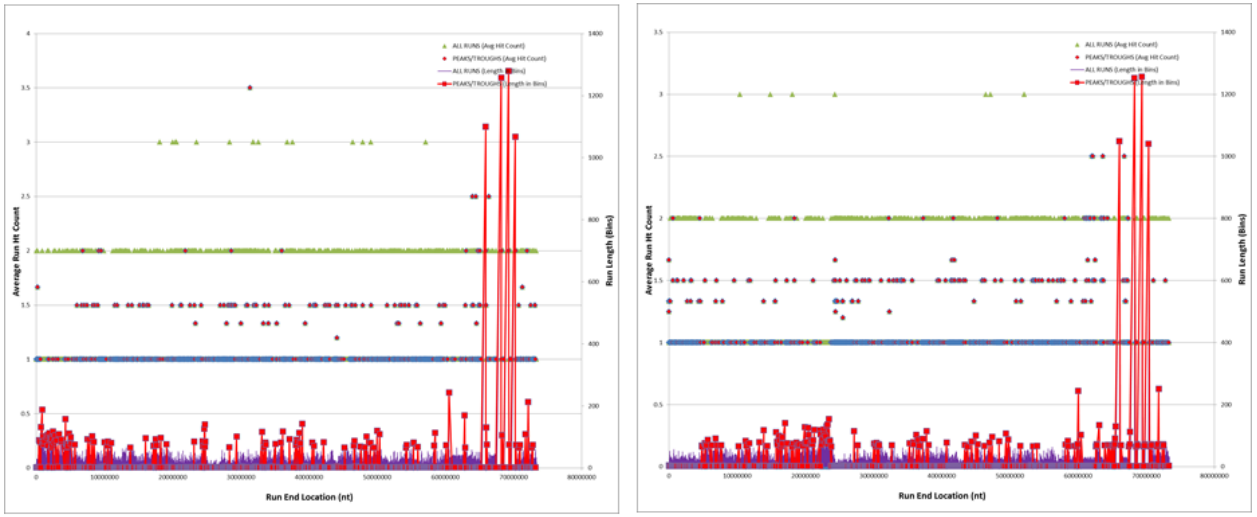


Figure C59: H. Sapiens Chromosome 19 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

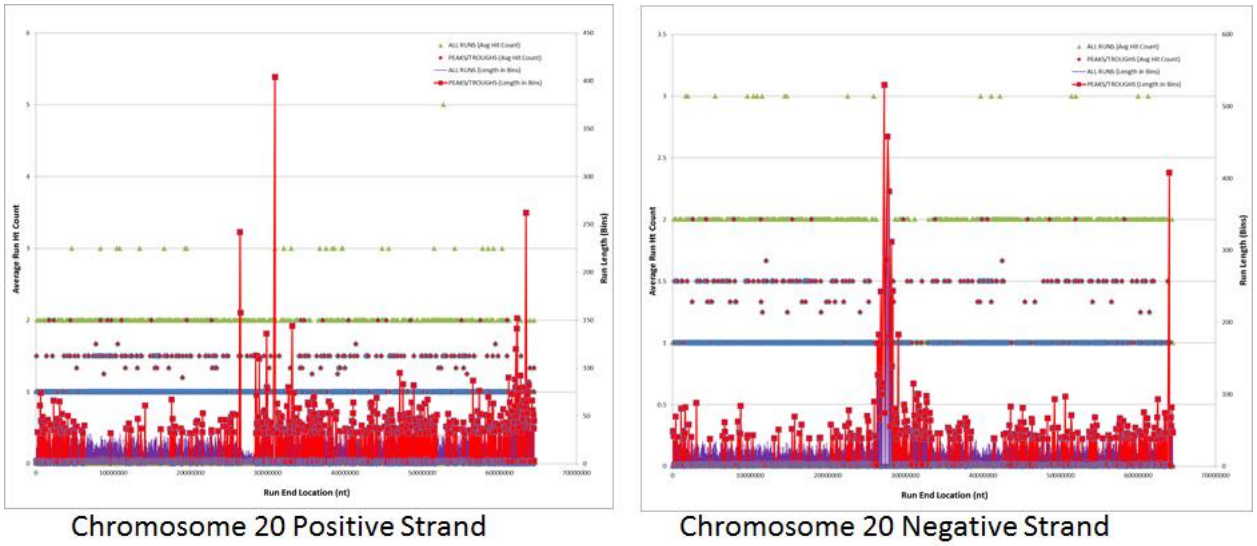


Figure C60: H. Sapiens Chromosome 20 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

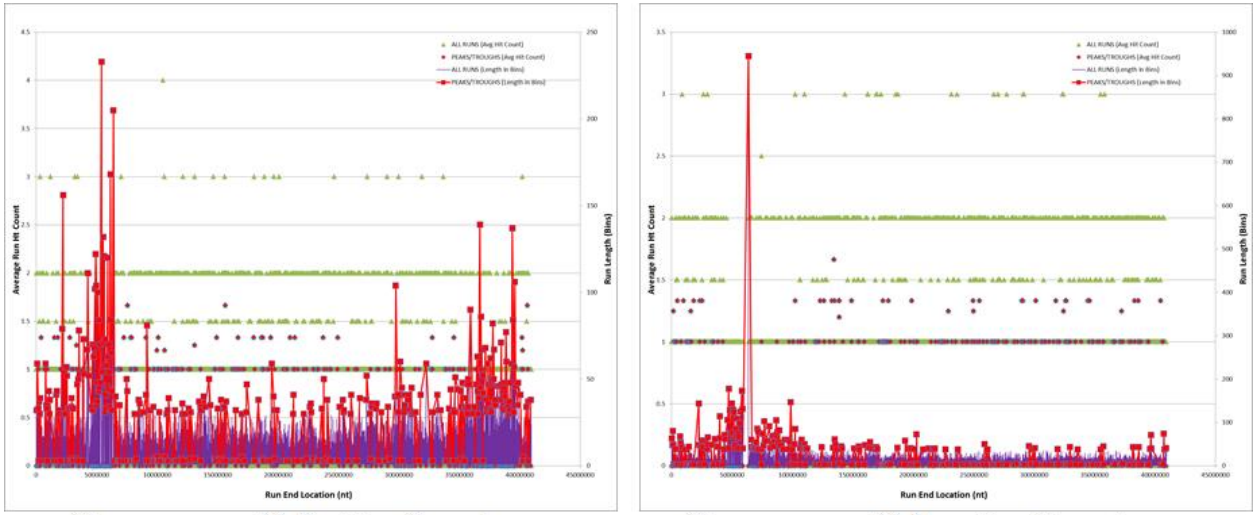


Figure C61: H. Sapiens Chromosome 21 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

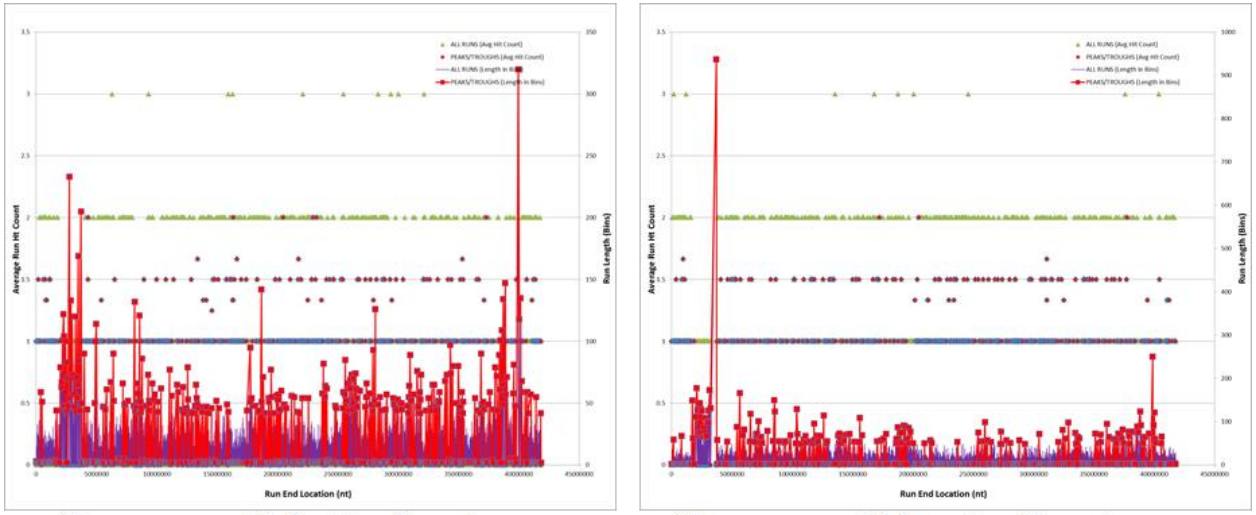


Figure C62: H. Sapiens Chromosome 22 Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

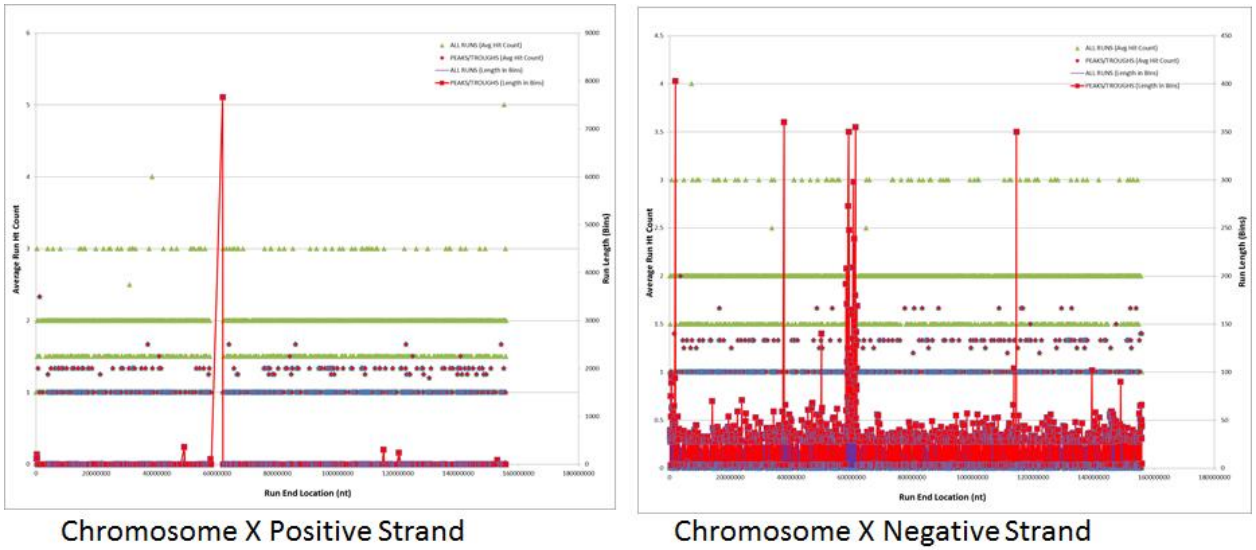


Figure C63: H. Sapiens Chromosome X Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

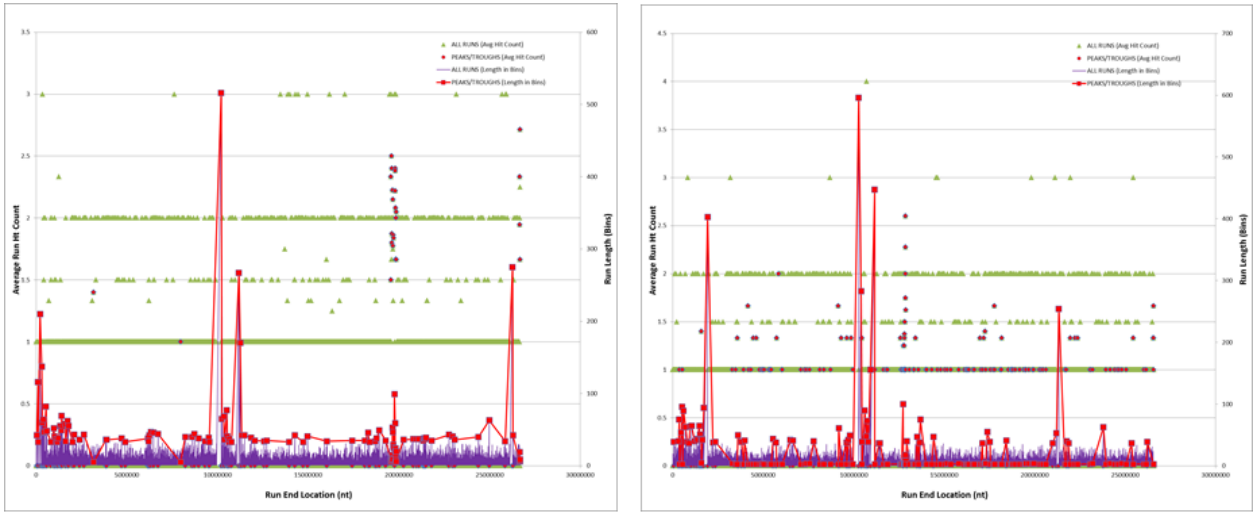


Figure C64: H. Sapiens Chromosome Y Run Distributions for Both Strands. The bin size used was 1000nt. Results are for all hits with total scores >= 0.99. The loci ID numbers are shown. See Chapter 5.

References

CHAPTER 1 REFERENCES

1. Crick,F. (1970) Central Dogma of Molecular Biology. *Nature*, **227**, 561–563.

2. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.

3. Saini,N., Zhang,Y., Usdin,K. and Lobachev,K.S. (2013) When secondary comes first – The importance of non-canonical DNA structures. *Biochimie*, **95**, 117–123.

4. Burge,S., Parkinson,G., Hazel,P., Todd,A. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.

5. Federiconi,F., Ausili,P., Fragneto,G., Ferrero,C. and Mariani,P. (2005) Locating counterions in guanosine quadruplexes: A contrast-variation neutron diffraction experiment in condensed hexagonal phase. *J. Phys. Chem. B*, **109**, 11037–11045.

6. Lipps,H. and Rhodes,D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.

7. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.

8. Paramasivan,S., Rujan,I. and Bolton,P.H. (2007) Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods*, **43**, 324–331.

9. Vorlíčková,M., Kejnovská,I., Sagi,J., Renčiuk,D., Bednářová,K., Motlová,J. and Kypr,J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64–75.

10. Randazzo,A., Spada,G.P. and da Silva,M.W. (2013) Circular Dichroism of Quadruplex Structures. *Quadruplex Nucleic Acids*, **330**, 67–86.

11. Lane,A., Chaires,J., Gray,R. and Trent,J. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.

12. Qin,Y. and Hurley,L. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **90**, 1149–1171.

13. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.

14. Mukundan,V.T. and Phan,A.T. (2013) Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.

15. Martadinata,H. and Phan,A.T. (2014) Formation of a Stacked Dimeric G-Quadruplex Containing Bulges by the 5′-Terminal Region of Human Telomerase RNA (hTERC). *Biochemistry*, **53**, 1595–1600.

16. Brooks,T.A., Kendrick,S. and Hurley,L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.*, **277**, 3459–3469.

17. Sun,D. and Hurley,L.H. (2009) The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J Med Chem*, **52**, 2863–2874.

18. Benabou,S., Ferreira,R., Aviñó,A., González,C., Lyonnais,S., Solà,M., Eritja,R., Jaumot,J. and Gargallo,R. (2014) Solution equilibria of cytosine- and guanine-rich sequences near the promoter region of the n-myc gene that contain stable hairpins within lateral loops. *Biochim. Biophys. Acta BBA - Gen. Subj.*, **1840**, 41–52.

19. Bardin,C. and Leroy,J.L. (2008) The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.*, **36**, 477–488.

20. Gellert,M., Lipsett,M. and Davies,D. (1962) Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.*, **48**, 2013–&.

21. Wu,Y. and Brosh Jr.,R.M. (2010) G-quadruplex nucleic acids and human disease. *FEBS J.*, **277**, 3470–3488.

22. Hurley,L.H. (2002) DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer*, **2**, 188–200.

23. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 2105.

24. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 8572–8577.

25. Cahoon,L.A. and Seifert,H.S. (2009) An Alternative DNA Structure Is Necessary for Pilin Antigenic Variation in Neisseria gonorrhoeae. *Science*, **325**, 764–767.

26. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.

27. Biffi,G., Di Antonio,M., Tannahill,D. and Balasubramanian,S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.*, **6**, 75–80.

28. Masuda-Sasa,T., Polaczek,P., Peng,X.P., Chen,L. and Campbell,J.L. (2008) Processing of G4 DNA by Dna2 helicase/nuclease and replication protein a (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. *J. Biol. Chem.*, **283**, 24359–24373.

29. Sanders,C.M. (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem. J.*, **430**, 119–128.

30. Paeschke,K., Bochman,M.L., Garcia,P.D., Cejka,P., Friedman,K.L., Kowalczykowski,S.C. and Zakian,V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.

31. Ribeyre,C., Lopes,J., Boule,J.-B., Piazza,A., Guedin,A., Zakian,V.A., Mergny,J.-L. and Nicolas,A. (2009) The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. *Plos Genet.*, **5**, e1000475.

32. Chakraborty,P. and Grosse,F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair*, **10**, 654–665.

33. Creacy,S.D., Routh,E.D., Iwamoto,F., Nagamine,Y., Akman,S.A. and Vaughn,J.P. (2008) G4 Resolvase 1 Binds Both DNA and RNA Tetramolecular Quadruplex with High Affinity and Is the Major Source of Tetramolecular Quadruplex G4-DNA and G4-RNA Resolving Activity in HeLa Cell Lysates. *J. Biol. Chem.*, **283**, 34626 –34634.

34. Vaughn,J.P., Creacy,S.D., Routh,E.D., Joyner-Butt,C., Jenkins,G.S., Pauli,S., Nagamine,Y. and Akman,S.A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **280**, 38117–38120.

35. Lattmann,S., Giri,B., Vaughn,J.P., Akman,S.A. and Nagamine,Y. (2010) Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.*, **38**, 6219 –6233.

36. London,T.B.C., Barber,L.J., Mosedale,G., Kelly,G.P., Balasubramanian,S., Hickson,I.D., Boulton,S.J. and Hiom,K. (2008) FANCJ Is a Structure-specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts. *J. Biol. Chem.*, **283**, 36132–36139.

37. Wu,Y., Rawtani,N., Thazhathveetil,A.K., Kenny,M.K., Seidman,M.M. and Brosh,R.M. (2008) Human Replication Protein A Melts a DNA Triple Helix Structure in a Potent and Specific Manner. *Biochemistry*, **47**, 5068–5077.

38. Wu,Y., Sommers,J.A., Khan,I., de Winter,J.P. and Brosh,R.M. (2012) Biochemical Characterization of Warsaw Breakage Syndrome Helicase. *J. Biol. Chem.*, **287**, 1007–1021.

39. Mohaghegh,P., Karow,J.K., Brosh Jr,R.M., Bohr,V.A. and Hickson,I.D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.*, **29**, 2843 –2849.

40. Li,J.L., Harrison,R.J., Reszka,A.P., Brosh,R.M., Bohr,V.A., Neidle,S. and Hickson,I.D. (2001) Inhibition of the Bloom's and Werner's syndrome helicases by G-quadruplex interacting ligands. *Biochemistry*, **40**, 15194–15202.

41. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.

42. Huber,M.D., Duquette,M.L., Shiels,J.C. and Maizels,N. (2006) A conserved G4 DNA binding domain in RecQ family helicases. *J. Mol. Biol.*, **358**, 1071–1080.

43. Popuri,V., Bachrati,C.Z., Muzzolini,L., Mosedale,G., Costantini,S., Giacomini,E., Hickson,I.D. and Vindigni,A. (2008) The Human RecQ Helicases, BLM and RECQ1, Display Distinct DNA Substrate Specificities. *J. Biol. Chem.*, **283**, 17766 –17776.

44. Fry,M. and Loeb,L.A. (1999) Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)(n). *J. Biol. Chem.*, **274**, 12797–12802.

45. Han,H., Bennett,R.J. and Hurley,L.H. (1999) PIPER, a G4 DNA-interactive compound, prevents the unwinding of G-quadruplex DNA by yeast Sgs1 helicase. *Clin. Cancer Res.*, **5**, 3852S–3852S.

46. Han,H.Y., Bennett,R.J. and Hurley,L.H. (2000) Inhibition of unwinding of G-quadruplex structures by Sgs1 helicase in the presence of N,N ’-bis[2-(1-piperidino)ethyl]-3,4,9,10-perylenetetracarboxylic diimide, a G-quadruplex-interactive ligand. *Biochemistry*, **39**, 9311–9316.

47. Han,H., Langley,D.R., Rangan,A. and Hurley,L.H. (2001) Selective Interactions of Cationic Porphyrins with G-Quadruplex Structures. *J. Am. Chem. Soc.*, **123**, 8902–8913.

48. Wu,X. and Maizels,N. (2001) Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Res.*, **29**, 1765–1771.

49. Postberg,J., Tsytlonok,M., Sparvoli,D., Rhodes,D. and Lipps,H.J. (2012) A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication. *Gene*, **497**, 147–154.

50. Tuesuwan,B., Kern,J.T., Thomas,P.W., Rodriguez,M., Li,J., David,W.M. and Kerwin,S.M. (2008) Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents. *Biochemistry*, **47**, 1896–1909.

51. Chen,J., Sperry,J., Ip,N. and Brimble,M. (2011) Natural products targeting telomere maintenance. *MedChemComm*, **2**, 229–245.

52. Wei,P.-C., Wang,Z.-F., Lo,W.-T., Su,M.-I., Shew,J.-Y., Chang,T.-C. and Lee,W.-H. (2012) A cis-element with mixed G-quadruplex structure of NPGPx promoter is essential for nucleolin-mediated transactivation on non-targeting siRNA stress. *Nucleic Acids Res.*

53. Cian,A.D. and Mergny,J.-L. (2007) Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. *Nucleic Acids Res.*, **35**, 2483–2493.

54. Han,F.X., Wheelhouse,R.T. and Hurley,L.H. (1999) Interactions of TMPyP4 and TMPyP2 with Quadruplex DNA. Structural Basis for the Differential Effects on Telomerase Inhibition. *J. Am. Chem. Soc.*, **121**, 3561–3570.

55. Vialas,C., Pratviel,G., Claparols,C. and Meunier,B. (1998) Efficient Oxidation of 2‘-Deoxyguanosine by Mn-TMPyP/KHSO5 to Imidazolone dIz without Formation of 8-Oxo-dG. *J Am Chem Soc*, **120**, 11548–11553.

56. Kovaleva,O.A., Shchyolkina,A.K., Mamaeva,O.K., Ol'shevskaya,V.A., Makarenkov,A.V., Semeikin,A.S., Shtil,A.A., Borisova,O.F. and Kaluzhny,D.N. (2013) Complexes of antiparallel telomeric G-quadruplex d(TTAGGG)4 with carboxymethyl tetracationic porphyrins. *Mol. Biol.*, **47**, 453–460.

57. Nagesh,N., Sharma,V.K., Ganesh Kumar,A. and Lewis,E.A. (2009) Effect of Ionic Strength on Porphyrin Drugs Interaction with Quadruplex DNA Formed by the Promoter Region of C-myc and Bcl2 Oncogenes. *J. Nucleic Acids*, **2010**.

58. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.

59. Granotier,C., Pennarun,G., Riou,L., Hoffschir,F., Gauthier,L.R., De Cian,A., Gomez,D., Mandine,E., Riou,J.-F., Mergny,J.-L., *et al.* (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.*, **33**, 4182–4190.

60. Kern,J.T., Thomas,P.W. and Kerwin,S.M. (2002) The Relationship between Ligand Aggregation and G-Quadruplex DNA Selectivity in a Series of 3,4,9,10-Perylenetetracarboxylic Acid Diimides. *Biochemistry*, **41**, 11379–11389.

61. D'Ambrosio,D., Reichenbach,P., Micheli,E., Alvino,A., Franceschin,M., Savino,M. and Lingner,J. (2012) Specific binding of telomeric G-quadruplexes by hydrosoluble perylene derivatives inhibits repeat addition processivity of human telomerase. *Biochimie*, **94**, 854–863.

62. Schoonover,M. and Kerwin,S.M. (2012) G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*, **20**, 6904–6918.

63. Moore,M.J.B., Cuenca,F., Searcey,M. and Neidle,S. (2006) Synthesis of distamycin A polyamides targeting G-quadruplex DNA. *Org Biomol Chem*, **4**, 3479–3488.

64. Bidzinska,J., Cimino-Reale,G., Zaffaroni,N. and Folini,M. (2013) G-Quadruplex Structures in the Human Genome as Novel Therapeutic Targets. *Molecules*, **18**, 12368–12395.

65. Imai,K. and Takaoka,A. (2006) Comparing antibody and small-molecule therapies for cancer. *Nat. Rev. Cancer*, **6**, 714–727.

66. Kolb,H.C., Finn,M.G. and Sharpless,K.B. (2001) Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.*, **40**, 2004–2021.

CHAPTER 2 REFERENCES

1. Granotier,C., Pennarun,G., Riou,L., Hoffschir,F., Gauthier,L.R., De Cian,A., Gomez,D., Mandine,E., Riou,J.-F., Mergny,J.-L., et al. (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.*, **33**, 4182–4190.

2. Rodriguez,R., Miller,K.M., Forment,J.V., Bradshaw,C.R., Nikan,M., Britton,S., Oelschlaegel,T., Xhemalce,B., Balasubramanian,S. and Jackson,S.P. (2012) Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.*, **8**, 301–310.

3. Müller,S., Sanders,D.A., Antonio,M.D., Matsis,S., Riou,J.-F., Rodriguez,R. and Balasubramanian,S. (2012) Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org. Biomol. Chem.*, **10**, 6537–6546.

4. Pasaogullari,N., Icil,H. and Demuth,M. (2006) Symmetrical and unsymmetrical perylene diimides: Their synthesis, photophysical and electrochemical properties. *Dyes Pigments*, **69**, 118–127.

5. Schoonover,M. and Kerwin,S.M. G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*

6. Tuesuwan,B., Kern,J.T., Thomas,P.W., Rodriguez,M., Li,J., David,W.M. and Kerwin,S.M. (2008) Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents. *Biochemistry*, **47**, 1896–1909.

7. Kern,J.T., Thomas,P.W. and Kerwin,S.M. (2002) The Relationship between Ligand Aggregation and G-Quadruplex DNA Selectivity in a Series of 3,4,9,10-Perylenetetracarboxylic Acid Diimides. *Biochemistry*, **41**, 12568–12568.

8. Rostovtsev,V.V., Green,L.G., Fokin,V.V. and Sharpless,K.B. (2002) A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chem.-Int. Ed.*, **41**, 2596.

9. Cuquerella,M.C., Lhiaubet-Vallet,V., Cadet,J. and Miranda,M.A. (2012) Benzophenone Photosensitized DNA Damage. *Acc. Chem. Res.*, **45**, 1558–1570.

10. Breslin,D.T. and Schuster,G.B. (1996) Anthraquinone Photonucleases: Mechanisms for GG-Selective and Nonselective Cleavage of Double-Stranded DNA. *J. Am. Chem. Soc.*, **118**, 2311–2319.

11. Aveline,B.M., Matsugo,S. and Redmond,R.W. (1997) Photochemical Mechanisms Responsible for the Versatile Application of Naphthalimides and Naphthaldiimides in Biological Systems. *J. Am. Chem. Soc.*, **119**, 11785–11795.

12. Dorazco-González,A., Höpfl,H., Medrano,F. and Yatsimirsky,A.K. (2010) Recognition of Anions and Neutral Guests by Dicationic Pyridine-2,6-dicarboxamide Receptors. *J. Org. Chem*., **75**, 2259–2273.

13. McPhee,M.M. and Kerwin,S.M. (1996) Synthesis and Metal Ion Binding Studies of Enediyne-Containing Crown Ethers. *J. Org. Chem.*, **61**, 9385–9393.

14. Triple-bond Isomerizations: 2- to 9-decyn-1-ol (1988) *Org. Synth.*, **66**, 127.

15. Rajput,A. and Mukherjee,R. (2013) Coordination chemistry with pyridine/pyrazine amide ligands. Some noteworthy results. *Coord. Chem. Rev.*, **257**, 350–368.

16. Shao,Y., Sheng,X., Li,Y., Jia,Z.-L., Zhang,J.-J., Liu,F. and Lu,G.-Y. (2008) DNA Binding and Cleaving Activity of the New Cleft Molecule N,N′-Bis(guanidinoethyl)-2,6-pyridinedicarboxamide in the Absence or in the Presence of Copper(II). *Bioconjug. Chem.*, **19**, 1840–1848.

17. Lee,L.V., Mitchell,M.L., Huang,S.-J., Fokin,V.V., Sharpless,K.B. and Wong,C.-H. (2003) A Potent and Highly Selective Inhibitor of Human α-1,3-Fucosyltransferase via Click Chemistry. *J. Am. Chem. Soc.*, **125**, 9588–9589.

18. Chenot,E.-D., Rodriguez-Dominguez,J.C., Hannewald,P., Comel,A. and Kirsch,G. (2008) Synthesis of potentially photoactivatable cournarin derivatives via a 1,3-dipolar cycloaddition. *J. Heterocycl. Chem.*, **45**, 1429–1435.

19. Li,H., Hao,M., Wang,L., Liang,W. and Chen,K. (2009) Preparation of Mono Boc-Protected Unsymmetrical Diamines. *Org. Prep. Proced. Int*., **41**, 301–307.

20. Almeida,M., Grehn,L. and Ragnarsson,U. (1988) Selective Protection of Polyamines - Synthesis of Model Compounds and Spermidine Derivatives. *J. Chem. Soc.-Perkin Trans. 1*.

21. Huang,H., Che,Y. and Zang,L. (2010) Direct synthesis of highly pure perylene tetracarboxylic monoimide. *Tetrahedron Lett.*, **51**, 6651–6653.

22. Amiralaei,S., Uzun,D. and Icil,H. (2008) Chiral substituent containing perylene monoanhydride monoimide and its highly soluble symmetrical diimide: synthesis, photophysics and electrochemistry from dilute solution to solid state. *Photochem. Photobiol. Sci.*, **7**, 936–947.

23. Asir,S., Demir,A.S. and Icil,H. (2010) The synthesis of novel, unsymmetrically substituted, chiral naphthalene and perylene diimides: Photophysical, electrochemical, chiroptical and intramolecular charge transfer properties. *Dyes Pigments*, **84**, 1–13.

24. Langhals,H., Krotz,O., Polborn,K. and Mayer,P. (2005) A Novel Fluorescent Dye with Strong, Anisotropic Solid-State Fluorescence, Small Stokes Shift, and High Photostability. *Angew. Chem. Int. Ed.*, **44**, 2427–2428.

25. Rajasingh,P., Cohen,R., Shirman,E., Shimon,L.J.W. and Rybtchinski,B. (2007) Selective bromination of perylene diimides under mild conditions. *J. Org. Chem.*, **72**, 5973–5979.

26. Langhals,H. and Ismael,R. (1998) Cyclophanes as Model Compounds for Permanent, Dynamic Aggregates – Induced Chirality with Strong CD Effects. *Eur. J. Org. Chem.*, **1998**, 1915–1917.

CHAPTER 3 REFERENCES

1. Kern,J.T. and Kerwin,S.M. (2002) The aggregation and G-quadruplex DNA selectivity of charged 3,4,9,10-perylenetetracarboxylic acid diimides. *Bioorg. Med. Chem. Lett.*, **12**, 3395–3398.

2. Tuesuwan,B., Kern,J.T., Thomas,P.W., Rodriguez,M., Li,J., David,W.M. and Kerwin,S.M. (2008) Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents. *Biochemistry*, **47**, 1896–1909.

3. Schoonover,M. and Kerwin,S.M. G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*

4. Ren,J.S. and Chaires,J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.

5. Kerwin,S. (2000) G-quadruplex DNA as a target for drug design. *Curr. Pharm. Des.*, **6**, 441–471.

6. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.

7. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.

8. Nicoludis,J.M., Miller,S.T., Jeffrey,P.D., Barrett,S.P., Rablen,P.R., Lawton,T.J. and Yatsunyk,L.A. (2012) Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.*

9. González,V. and Hurley,L.H. (2010) The c-MYC NHE III1: Function and Regulation. *Annu. Rev. Pharmacol. Toxicol.*, **50**, 111–129.

10. Ambrus,A., Chen,D., Dai,J.X., Jones,R.A. and Yang,D.Z. (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.

11. Schoonover,M. and Kerwin,S.M. (2011) A Fluorescence-Based G-Quadruplex DNA Cleavage Assay. In Sheardy,R.D., Winkle,S.A. (eds), *Frontiers in Nucleic Acids. Amer Chemical Soc, Washington*, **1082**, pp. 13–32.

12. Kern,J.T., Thomas,P.W. and Kerwin,S.M. (2002) The Relationship between Ligand Aggregation and G-Quadruplex DNA Selectivity in a Series of 3,4,9,10-Perylenetetracarboxylic Acid Diimides. *Biochemistry*, **41**, 11379–11389.

13. Aoki,K., Ishikawa,Y., Oyama,M., Tomisugi,Y. and Uno,T. (2003) Self-Aggregation Inhibits the Photonuclease Activity of Porphyrins. *Chem. Pharm. Bull. (Tokyo)*, **51**, 899–903.

CHAPTER 4 REFERENCES

1. Ren,J.S. and Chaires,J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.

2. Kerwin,S. (2000) G-quadruplex DNA as a target for drug design. *Curr. Pharm. Des.*, **6**, 441–471.

3. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.

4. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.

5. Nicoludis,J.M., Miller,S.T., Jeffrey,P.D., Barrett,S.P., Rablen,P.R., Lawton,T.J. and Yatsunyk,L.A. (2012) Optimized End-Stacking Provides Specificity of N-Methyl Mesoporphyrin IX for Human Telomeric G-Quadruplex DNA. *J. Am. Chem. Soc.*, **134**, 20446–20456.

6. Kern,J.T. and Kerwin,S.M. (2002) The aggregation and G-quadruplex DNA selectivity of charged 3,4,9,10-perylenetetracarboxylic acid diimides. *Bioorg. Med. Chem. Lett.*, **12**, 3395–3398.

7. Tuesuwan,B., Kern,J.T., Thomas,P.W., Rodriguez,M., Li,J., David,W.M. and Kerwin,S.M. (2008) Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents. *Biochemistry*, **47**, 1896–1909.

8. Schoonover,M. and Kerwin,S.M. (2012) G-quadruplex DNA cleavage preference and identification of a perylene diimide G-quadruplex photocleavage agent using a rapid fluorescent assay. *Bioorg. Med. Chem.*, **20**, 6904–6918.

9. Parris,C.N. and Seidman,M.M. (1992) A signature element distinguishes sibling and independent mutations in a shuttle vector plasmid. *Gene*, **117**, 1–5.

10. Han,F., Wheelhouse,R. and Hurley,L. (1999) Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. *J. Am. Chem. Soc.*, **121**, 3561–3570.

11. Han,H., Langley,D.R., Rangan,A. and Hurley,L.H. (2001) Selective Interactions of Cationic Porphyrins with G-Quadruplex Structures. *J. Am. Chem. Soc.*, **123**, 8902–8913.

12. Razin,A., Sedat,J.W. and Sinsheimer,R.L. (1970) Structure of the DNA of bacteriophage φX174: VII. Methylation. *J. Mol. Biol.*, **53**, 251–259.

13. Warren,R.A.J. (1980) Modified Bases in Bacteriophage DNAs. *Annu. Rev. Microbiol.*, **34**, 137–158.

14. Gommers-Ampt,J.H. and Borst,P. (1995) Hypermodified bases in DNA. *FASEB J.*, **9**, 1034–1042.

15. Paramasivan,S., Rujan,I. and Bolton,P.H. (2007) Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods*, **43**, 324–331.

16. Kypr,J., Kejnovská,I., Renčiuk,D. and Vorlíčková,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.

17. Vorlíčková,M., Kejnovská,I., Sagi,J., Renčiuk,D., Bednářová,K., Motlová,J. and Kypr,J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64–75.

18. Randazzo,A., Spada,G.P. and da Silva,M.W. (2013) Circular Dichroism of Quadruplex Structures. *Quadruplex Nucleic Acids*, **330**, 67–86.

CHAPTER 5 REFERENCES

1. Bardin,C. and Leroy,J.L. (2008) The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.*, **36**, 477–488.

2. Stadlbauer,P., Trantirek,L., Cheatham,T.E., Koca,J. and Sponer,J. (2014) Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie*, **105C**, 22–35.

3. Rajendran,A., Endo,M., Hidaka,K. and Sugiyama,H. (2014) Direct and Single-Molecule Visualization of the Solution-State Structures of G-Hairpin and G-Triplex Intermediates. *Angew. Chem.-Int. Ed.*, **53**, 4107–4112.

4. Cerofolini,L., Amato,J., Giachetti,A., Limongelli,V., Novellino,E., Parrinello,M., Fragai,M., Randazzo,A. and Luchinat,C. (2014) G-triplex structure and formation propensity. *Nucleic Acids Res.*, **42**, 13393–13404.

5. Kikin,O., D'Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–W682.

6. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.

7. Cock,P.J.A., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B., et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinforma. Oxf. Engl.*, **25**, 1422–1423.

8. Mukundan,V.T. and Phan,A.T. (2013) Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.

APPENDIX REFERENCES

1. Parris,C.N. and Seidman,M.M. (1992) A signature element distinguishes sibling and independent mutations in a shuttle vector plasmid. *Gene*, **117**, 1–5.

2. Dorazco-González,A., Höpfl,H., Medrano,F. and Yatsimirsky,A.K. (2010) Recognition of Anions and Neutral Guests by Dicationic Pyridine-2,6-dicarboxamide Receptors. *J. Org. Chem.*, **75**, 2259–2273.

3. Schmidt,B., Jiricek,J., Titz,A., Ye,G. and Parang,K. (2004) Copper dipicolinates as peptidomimetic ligands for the Src SH2 domain. *Bioorg. Med. Chem. Lett.*, **14**, 4203–4206.

4. Lee,L.V., Mitchell,M.L., Huang,S.-J., Fokin,V.V., Sharpless,K.B. and Wong,C.-H. (2003) A Potent and Highly Selective Inhibitor of Human α-1,3-Fucosyltransferase via Click Chemistry. *J. Am. Chem. Soc.*, **125**, 9588–9589.

5. Kadin,S.B. (1981) United States Patent: 4247699 - Process for making 2-amino-4-hydroxyquinolines.

6. Müller,S., Sanders,D.A., Antonio,M.D., Matsis,S., Riou,J.-F., Rodriguez,R. and Balasubramanian,S. (2012) Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org. Biomol. Chem.*, **10**, 6537–6546.

7. Zhao,H. and Thurkauf,A. (2001) A Practical and Convenient Synthesis of Methyl 5-Formyl-3-Methoxybenzoate. *Synth. Commun.*, **31**, 1921–1926.

8. McPhee,M.M. and Kerwin,S.M. (1996) Synthesis and Metal Ion Binding Studies of Enediyne-Containing Crown Ethers. *J. Org. Chem.*, **61**, 9385–9393.

9. Triple-bond Isomerizations: 2- to 9-decyn-1-ol (1988) *Org. Synth.*, **66**, 127.

10. Almeida,M., Grehn,L. and Ragnarsson,U. (1988) Selective Protection of Polyamines - Synthesis of Model Compounds and Spermidine Derivatives. *J. Chem. Soc.-Perkin Trans. 1*

11. Langhals,H., Krotz,O., Polborn,K. and Mayer,P. (2005) A Novel Fluorescent Dye with Strong, Anisotropic Solid-State Fluorescence, Small Stokes Shift, and High Photostability. *Angew. Chem. Int. Ed.*, **44**, 2427–2428.

12. Rajasingh,P., Cohen,R., Shirman,E., Shimon,L.J.W. and Rybtchinski,B. (2007) Selective bromination of perylene diimides under mild conditions. *J. Org. Chem.*, **72**, 5973–5979.