EWS-FLI1 AS A MOLECULAR TARGET: SMALL MOLECULE INHIBITORS FOR A DISORDERED PROTEIN

A Dissertation
subscribed to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Tumor Biology

By

Julie Samantha Barber-Rotenberg, B. S.

Washington, DC
April 05, 2012
Copyright 2012 by Julie Samantha Barber-Rotenberg
All rights reserved
Ewing sarcoma family of tumors (ESFT) consists of highly malignant tumors of the bone and soft tissue. Ninety-five percent of cases contain a balanced t(11;22) or t(21;22) rearrangement, combining the amino-terminus of EWS to the carboxy-terminus of FLI1 or ERG, which contain the highly conserved ets DNA binding domain. As the EWS-FLI1 protein is found only in ESFT cells and its expression is required for the oncogenic phenotype, it presents a promising molecular target for anti-cancer therapies. EWS-FLI1 is a hydrophobic disordered protein with unknown three-dimensional structure, precluding standard structure-based drug design. RNA Helicase A (RHA) enhances EWS-FLI1 driven oncogenesis and interruption of this protein-protein complex validates this interaction as a unique therapeutic target. Surface plasmon resonance screening identified compounds that bind to EWS-FLI1, including a lead compound that induces apoptosis in ESFT cells and reduces the growth of xenografts. Our compound, YK-4-279, has a chiral center and can be separated into enantiomers, only one of which is able to specifically target the protein-protein interaction. This work is significant for its identification of a single enantiomer effect upon a protein-protein interaction suggesting that small molecule targeting of intrinsically disordered proteins can be highly specific. Given the challenges of drug design targeted to EWS-FLI1, we proposed that characterization of the physical interaction points
between EWS-FLI1 and RHA would allow us to better alter the lead compound to block this protein-protein interaction. While full length EWS-FLI1 is able to pull down RHA, fragments of the protein are not. Although we can successfully crosslink EWS-FLI1 and RHA, we have yet to identify what region of EWS-FLI1 is involved. We are able to show specific regions of order and disorder of EWS-FLI1, which may lead to the identification of the binding site for YK-4-279. The development of higher-throughput methods for testing small molecules that bind to or inhibit EWS-FLI1 function will allow us to further investigate protein structure and function. These data are a contribution to the future development of small molecules in an era where novel approaches to cancer therapy are critical for improving patient care.
ACKNOWLEDGEMENTS

This research would not have been possible without the help of the following individuals:

- My mentor, Dr. Jeffrey Toretsky, for all his support and guidance in helping me bring together a very multidisciplinary project that took many unexpected twists and turns along the way. Most importantly, I’d like to thank him for sharing his unrelenting enthusiasm and passion for science.

- My committee members, Dr. Elliott Crooke, Dr. Aykut Uren, Dr. Steven Metallo, and Dr. Milton Brown for all their advice and suggestions, and bringing many new techniques into my research.

- Dr. Verda Erkizan, Dr. Silke Schlottmann, Dr. Saravana Selvanathan, Dr. Ogan Abaan, and the rest of the current and past members of the Toretsky/Uren Laboratory for both technical support in the lab and invaluable friendships outside the lab. Even when none of my experiments were working, you gave me a reason to look forward to coming into lab everyday.

- Dr. Yali Kong and Perrer Tosso for the synthesis of analogues and providing a constant supply of YK-4-279 for experiments.

- My friends and family for their support, patience, and encouragement during this long process.

- My husband, Dr. Scott Rotenberg, for the hundreds of hours spent driving up and down I-95 on the weekend so that we could both continue our education in separate locations. After years of saying “just a few more months in DC...” I finally mean it.
# TABLE OF CONTENTS

Chapter I: INTRODUCTION ........................................................................................................... 1

A. Transcription Factors ............................................................................................................. 2

B. Ewing Sarcoma ....................................................................................................................... 4

C. EWS-FLI1 ............................................................................................................................. 5

D. RNA Helicase A ..................................................................................................................... 8

E. Disordered Proteins ............................................................................................................... 11
   1. General Overview .............................................................................................................. 11
   2. Functions of Disordered Proteins ................................................................................... 13
   3. Disorder in Transcription Factors ................................................................................... 15
   4. Disordered Proteins in Cancer ......................................................................................... 16
   5. EWS-FLI1 is a Disordered Protein ................................................................................... 17

F. Targeting Transcription Factors & Disordered Proteins ..................................................... 20

G. Stereoselectivity ................................................................................................................... 22

H. Rational and Hypothesis ..................................................................................................... 23

Chapter II: Materials and Methods .......................................................................................... 27

A. Cell Lines ............................................................................................................................ 28

B. Small Molecules and Peptides ........................................................................................... 28

C. Cellular Proliferation Assays ............................................................................................. 28

D. Caspase-3 Activity .............................................................................................................. 29

E. Surface Plasmon Resonance ............................................................................................... 29

F. Fluorescent Polarization ...................................................................................................... 29
G. ELISA ........................................................................................................................................ 30
H. Nuclear Prep Immunoprecipitation ......................................................................................... 30
I. Preparation of Whole Cell Lysate ............................................................................................ 31
J. Immunoblotting ...................................................................................................................... 31
K. Luciferase Reporter Assay ...................................................................................................... 32
L. Quantitative Real Time PCR .................................................................................................. 32
M. Animal Studies ..................................................................................................................... 33
N. GST-EWS-FLI1 Fragment Pull-Down .................................................................................... 33
O. Protein Purification ................................................................................................................. 34
P. Protein Crosslinking ................................................................................................................. 35
Q. Protein Melting ........................................................................................................................ 35
R. Circular Dichroism .................................................................................................................. 35
S. Native State Proteolysis .......................................................................................................... 36
T. AlphaScreen ............................................................................................................................ 36

Chapter III: Results .................................................................................................................... 37

A. Targeting EWS-FLI1 ............................................................................................................. 38
   1. NSC635437 is the lead compound for targeting EWS-FLI1.............................................. 38
   2. Optimized derivative YK-4-279 has higher specificity and potency................................. 41
   3. YK-4-279 structure is similar to the E9R RHA peptide ..................................................... 45
   4. YK-4-279 is able to displace E9R from EWS-FLI1............................................................. 46
   5. (S)-YK-4-279 disrupts binding of RHA with EWS-FLI1..................................................... 49
   6. EWS-FLI1 functional activity is reduced by only one enantiomer .................................... 52

vii
7. YK-4-279 demonstrates enantiospecific cellular effects ................................................................. 54
8. Racemic YK-4-279 inhibits in vivo growth of ESFT ................................................................. 57
B. Characterization of interaction between EWS-FLI1 and RHA............................................. 60
  1. Full-length EWS-FLI1 is required to pull down RHA............................................................ 61
  2. RHA 630-1020 fragment can be purified for use in binding experiments......................... 63
  3. EWS-FLI1 purification improvements result in increased purity........................................ 68
  4. Recombinant EWS-FLI1 is dependent on imidazole in the buffer...................................... 70
  5. Imidazole prevents structural analysis by circular dichroism.............................................. 73
  6. EWS-FLI1 and RHA can be crosslinked................................................................................... 76
  7. Native state proteolysis shows EWS-FLI1 regions of order and disorder............................ 78
C. Assay development for small molecule probes of structure and function.......................... 83
  1. Fluorescence Polarization........................................................................................................ 84
  2. Protein Melting Curves......................................................................................................... 85
  3. AlphaScreen.......................................................................................................................... 86
Chapter IV: Discussion.................................................................................................................. 90
  A. Single enantiomer of YK-4-279 demonstrates specificity in targeting EWS-FLI1.......... 91
  B. EWS-FLI1 binds to RHA but region of interaction remains unknown............................ 96
  C. Small molecule assay for structure-function relationship............................................... 100
  D. Proposed models of interaction between EWS-FLI1 and RHA...................................... 101
Chapter VI: References............................................................................................................... 107
LIST OF FIGURES

Figure 1: EWS-FLI1 is an oncogenic fusion protein.................................................................7
Figure 2: RHA is overexpressed in many cancers........................................................................9
Figure 3: EWS-FLI1 binds in a unique region of RHA...............................................................11
Figure 4: Protein disorder predictors indicate that EWS-FLI1 is a disordered protein.............20
Figure 5: NSC635437 is the lead compound..............................................................................41
Figure 6: The derivative of YK-4-279 improves potency..............................................................42
Figure 7: YK-4-279 is a potent and specific inhibitor of ESFT in biological assays....................44
Figure 8: YK-4-279 displays similar structure to the E9R peptide...............................................45
Figure 9: E9R peptide directly binds to EWS-FLI1 .................................................................47
Figure 10: YK-4-279 is able to displace E9R peptide from EWS-FLI1........................................49
Figure 11: (S)-YK-4-279 is able to disrupt the binding between EWS-FLI1 and RHA.............51
Figure 12: (S)-YK-4-279 blocks transcriptional activity of EWS-FLI1........................................53
Figure 13: (S)-YK-4-279 reduces Cyclin D1b/D1a levels in A673i cells.......................................54
Figure 14: (S)-YK-4-279 is the active enantiomer in biological assays........................................56
Figure 15: YK-4-279 inhibits ESFT cell proliferation in vivo in a mouse xenograft model...........60
Figure 16: Full length GST-EWS-FLI1 pulls down endogenous RHA, but fragments do not....62
Figure 17: RHA 630-1020 fragment purification challenges.....................................................66
Figure 18: EWS-FLI1 pulls down purified recombinant RHA 630-1020......................................67
Figure 19: Purity of recombinant EWS-FLI1 is improved with increase in imidazole.............70
Figure 20: Dilution of EWS-FLI1 changes the profile of the protein melting curve..................72
Figure 21: Addition of urea improves EWS-FLI1 solubility without imidazole.........................73
Figure 22: Circular dichroism of EWS-FLI1 is hindered due to buffer constraints. .................. 75

Figure 23: EWS-FLI1 and RHA crosslink in the presence of BS3. ........................................ 77

Figure 24: EWS-FLI1 has a lower melting point than two ordered proteins. ......................... 79

Figure 25: Native state proteolysis identifies regions of EWS-FLI1 disorder. ..................... 82

Figure 26: Addition of YK-4-279 changes digestion profile of EWS-FLI1. ......................... 83

Figure 27: High-throughput FP does not yield a large dynamic range. ............................. 85

Figure 28: Addition of small molecule does not shift melting temperature of EWS-FLI1. .... 86

Figure 29: AlphaScreen provides inconsistent results and may be quenched by protein.... 88

Figure 30: ANCHOR predicts two sites of interaction on EWS-FLI1. ................................. 105
LIST OF TABLES

Table 1: Disordered protein predictors................................................................. 13
Table 2: IC\textsubscript{50} values in a panel of cells......................................................... 57
Table 3: Mass spectrometry analysis of crosslinked bands ........................................ 78
Chapter I: INTRODUCTION
A. Transcription Factors

Transcription factors are proteins that bind to specific DNA sequences, controlling the rate of transcription of DNA to mRNA[1]. Transcription factors function by promoting or blocking the recruitment of RNA polymerase II to specific genes, thereby activating or repressing the transcription of that particular gene. All transcription factors contain at least one structured DNA binding domain (DBD) that recognizes specific promoter sequences of DNA, known as the response element, adjacent to the genes they regulate[2]. One gene may be flanked by multiple binding sites for different transcription factors, requiring the binding of several transcription factors for expression of the gene[3]. Transcription factors bind to the response element in a sequence-specific manner using a combination of electrostatic and Van der Waals forces.

In many situations, a transcription factor may be regulated by only being synthesized in one particular tissue or cell type. This often occurs in signaling cascades, where a transcription factor is located downstream in a signaling cascade and allows for the upregulation or downregulation of other genes within the signaling network[4]. Transcription factors, acting as oncogenes and tumor suppressors, regulate the cell cycle and play a role in cell growth or apoptosis[5].

Transcription factors play an important role in disease, as mutations involving transcription factors are often associated with specific diseases and are thus an attractive target for therapeutic treatment. Cancer can arise from either the aberrant activation of genes encoding growth-promoting factors or from mutations resulting in the inactivation of tumor suppressors. Cancers driven by oncogenes that encode transcription factors
occur when the genes are expressed at significantly higher levels or encode for a protein with abnormal activity. Transcription factors are also involved in chromosomal rearrangements, particularly in leukemias, fusing an oncogene to part of another protein resulting in a novel protein with oncogenic activity[6].

The three most important classes of transcription factors involved in cancer are nuclear receptors, resident nuclear proteins, and latent cytoplasmic factors[3]. Nuclear receptors, such as estrogen receptors in breast cancer and androgen receptors in prostate cancer, regulate gene expression by binding to a small molecule ligand to induce a conformational change. Resident nuclear factors enter the nucleus upon synthesis where they bind to DNA constitutively and are activated by serine kinase cascades[3]. Oncogenic proteins affect cascades that end with the phosphorylation of these transcription factors, leading to altered transcriptional patterns. Latent cytoplasmic factors are usually activated by receptor-ligand interactions at the cell surface including phosphorylation by serine or tyrosine kinases. Upon activation, latent cytoplasmic factors translocate into the nucleus where they regulate transcription in conjunction with resident nuclear factors. Examples of cancers driven by transcription factors include mutations in p53 resulting in Li-Fraumeni syndrome, an autosomal dominant hereditary disorder increasing susceptibility to a wide range of cancers[7]; mutations to the STAT family which result in breast cancer[8]; mutations to HOX-family transcription factors, relevant in a wide variety of cancers; and the prolific oncogene c-Myc[9]. Additionally, transcription factor translocations, involving fusion proteins that result from chromosomal translocations, play an important role in tumorigenesis and account for 20% of human cancer morbidity[10].
B. Ewing Sarcoma

Ewing sarcoma family of tumors (ESFT), which comprise Ewing sarcoma, peripheral primitive neuroectodermal tumors, and Askin tumor, consists of highly malignant tumors of the bone and soft tissue. ESFT are highly aggressive undifferentiated tumors that can occur anywhere in the body, but are most frequently found in the femur[11]. Ewing sarcoma is the second most common primary bone tumor in the pediatric population, accounting for approximately 4% of pediatric malignancies[12]. Current therapy for ESFT patients is a five-drug regimen consisting of alternating cycles of doxorubicin, vincristine, and cyclophosphamide with etoposide and ifosfamide, often resulting in severe hematologic cytopenias, nausea/vomiting, and infections[13]. Even with a combined approach of surgery, radiotherapy, and chemotherapy, patients presenting with localized disease have a 70% five-year survival rate, while patients presenting with metastatic disease have only a 20% chance of disease free survival[14]. These clinical response rates have remained unchanged even after chemotherapy and bone marrow transplantation, indicating the need for novel approaches to improve survival.

The cell of origin of ESFT is still unknown, impeding basic understanding of the disease. Histologically, ESFT consists of small round cells containing a small cytoplasm with vesicular nuclei[15]. Currently, many believe that ESFTs are derived from neural crest progenitors. It has been observed that ESFTs express catechol acetyltransferase, which is an enzyme involved in neurotransmitter biosynthesis, and neuron-specific enolase. Certain ESFT cell lines also have the ability to form primitive dendrites and express neural-associated proteins, undergoing neural differentiation, when treated with differentiating
agents[16, 17]. A gene expression profiling study found that ESFT cells highly expressed genes commonly found in neural tissues or during neuronal differentiation, and clustered Ewing sarcoma with brain tissue[18]. However, other experiments indicate that ESFT may have mesenchymal stem cell features, as forced expression of the oncogenic fusion protein characteristic of ESFT inhibits adipogenic and osteogenic differentiation in murine marrow stromal cells[19] and impairs myogenic differentiation in C2C12 myoblasts[20]. The recognition of the 11;22 chromosomal translocation in ESFT helped to define Ewing tumors as a distinct tumor subgroup.

C. EWS-FLI1

Ninety-five percent of ESFT cases contain a balanced t(11;22) or t(21;22) rearrangement, combining the amino-terminus of EWS to the carboxy-terminus of FLI1 or ERG, both of which contain the highly conserved ets (erythroblastosis virus E26 transforming sequence gene) DNA binding domain[21] (Figure 1). Translocations have also been identified joining EWS to three other ETS family transcription factors, ETV1, ETV4, and FEV, forming an oncogenic fusion protein[22-24]. All ETS family transcription factors are defined by an 87 amino acid domain which is required for site-specific DNA-binding, and are thought to act by binding to the promoter of target genes resulting in transcriptional activation or repression. Chromosomal translocations may occur at various exon-intron boundaries, resulting in 12 different EWS-FLI1 fusions. The t(11;22) breakpoints are located within an 8kb region on EWS, but spread over 35kb on FLI1[21]. Type 1 translocations, the most common fusion, account for approximately 60% of patients
and bring together the EWS exon 7 with FLI1 exon 6. Type 2 translocations fuse together exon 7 of EWS with exon 5 of FLI1. Type 3 fusions, which are uncommon, occur between exon 10 of EWS and exon 6 of FLI1[21, 25]. Retrospective studies initially indicated that clinical outcome may be dependent on the specific fusion type expressed in the tumor, suggesting that the type 1 fusion results in an increased chance of survival compared to alternative fusions[26, 27]. However, more a more recent prospective study concluded that there was no significant prognostic value based on the fusion type[28].

EWS-FLI1 acts as an aberrant transcription factor, modifying the transcription of target genes, and thus is able to regulate downstream targets by both activating the cell cycle and repressing key anti-tumorigenic genes[29]. EWS-FLI1 directly increases expression of Id2[30], tenasin-C[31], MMP-1[32], PTPL1[33], and GLI1[34], and downregulates p21WAF/CIP1[35] and TGF-βRII[36]. Other gene targets, which may be indirect targets controlled by another direct target, include EAT-2[37], mE2C[38], manic fringe[39], c-myc[40], p57KIP[41], PIM-3[42], and PDGF-C[43]. EWS-FLI1 is also able to modulate RNA splicing by alteration of an E1A splice site and interaction with U1C, a member of the U1 small nuclear ribonucleoprotein-specific protein family[44, 45]. Functionally, EWS-FLI1 is able to inhibit RNA splicing in vivo, as demonstrated with an E1A splicing assay[45, 46].

All EWS/ETS proteins contain the ETS DNA-binding domain, and are able to bind DNA in a site-specific manner. The ETS domain is a variant of the winged helix-turn-helix motif[47], and is a target for protein-protein interactions that are mediated by co-regulatory proteins. Mutations within the binding domain of the fusion proteins result in
reduced or completely inhibited transformation potential[48], while deletion of EWS transcriptional activation domains similarly affects transformation potency[49]. Reduction in EWS-FLI1 by either antisense oligodeoxynucleotides[50, 51] or antisense RNA[52-54] results in a decrease of tumor cells in vitro and a regression of tumors in vivo. These treatments are associated with a cell cycle shift toward G0/G1. However, despite promising results in mice, these approaches currently lack translation to clinical therapy[51, 55]. These data indicate that the expression of EWS-FLI1 is required to maintain the oncogenic phenotype. Since the EWS-FLI1 protein is found only in ESFT cells and its expression is required for oncogenesis, it presents a promising molecular target for anti-cancer therapies.

Figure 1: EWS-FLI1 is an oncogenic fusion protein.
The t(11;22)(q24;12) balanced translocation combines the amino-terminal end of EWS with the carboxy-terminal region of the FLI1 protein. The resulting fusion protein contains both the transcriptional activating EWS domain and the ets binding domain which binds to DNA and modulates the transcription of target genes.
D. RNA Helicase A

RNA helicases are characterized by an amino acid sequence containing a centrally located “helicase domain” consisting of 8 motifs[56]. The two largest families of human RNA helicases are DDX and DHX, named after the DEAD-box and DEAH-box of the helicase domain, based on the single amino acid designation of motif II of the helicase domain[57]. RNA Helicase A (RHA), also known as DHX9 or nuclear DNA helicase II, is a highly conserved DEXH-box protein with a role in activating transcription[58-61], modulating RNA splicing[62-64], and binding to the nuclear pore complex[65]. It is both an RNA[66] and DNA helicase catalyzing the ATP-dependent unwinding of double-stranded RNA and DNA[67], and is required for normal gastrulation[68]. RHA was originally purified from nuclear complexes while studying the role of helicases in splicing[69, 70]. RHA complexes with RNA polymerase II (Pol II) and CREB-binding protein (CBP), and modulates transcription through these interactions[58]. RHA binds pre-mRNA and RNA[67] and is able to shuttle between the nucleus and the cytoplasm due to a bidirectional nuclear transport domain located in the C-terminus[71]. It is part of both the basal transcriptional apparatus[35] and the posttranscriptional RNA metabolism[72].

RHA is involved in cancers, including breast and lung. Similar to CBP, RHA links the C-terminus domain of breast cancer protein-1 (BRCA1) to Pol II[59], and dysregulation of expression of RHA may interfere with the ability of BRCA1 to act as a tumor suppressor[73]. RHA has also been shown to bind DNA in a sequence specific manner to the promoter sequences of the p16\textsuperscript{INK4a} tumor suppressor gene[74] and is recruited to the promoter of the multidrug resistance gene (MDR1), upregulating transcription of the MDR1 gene[75].
RHA may also play a role in lung cancer, as it is overexpressed in lung cancer cell lines and primary carcinomas in comparison to normal lung tissue[76].

![Figure 2: RHA is overexpressed in many cancers.](image)

An Oncomine search for RHA shows overexpression of the protein in multiple types of cancers. Here, expression is shown for (A) normal lung (blue) versus squamous lung cell carcinoma (red) and (B) normal pancreatic duct (blue) versus pancreatic ductal adenocarcinoma (red). In both cases, RHA is overexpressed in the abnormal cells compared to the normal tissue. Oncomine™ (Comprendia Bioscience) was used for analysis and visualization.

RHA directly binds to EWS-FLI1 to modulate transcription and anchorage-independent growth. RHA was identified as a partner protein of EWS-FLI1 based on its homology to peptides from the screening of an M13 phage library[77]. The 630-1020 amino acid fragment of GST-tagged RHA co-precipitates with endogenous EWS-FLI1.
Following identification of RHA as a partner protein, tissue microarray studies found RHA in 21/21 primary tumors and 19/19 metastatic Ewing tumors. A reduction in RHA levels using short hairpin RNA reduced viability of ESFT cells by 90%, while transfecting mouse embryonic fibroblasts with EWS-FLI1 and RHA induced anchorage independent colony growth. RHA function is necessary to enhance the transformation of EWS-FLI1, making it an ideal target for disruption of a protein-protein interaction. Most protein-protein interactions, including CBP[58], Pol II and BRCA1[59], and RNA-induced silencing complex components[78], occur in the amino-terminal region of RHA. EWS-FLI1 binds to a region in the distal portion of the helicase domain of RHA that is currently known only to interact with EGFR[79, 80], thus further highlighting this protein-protein interaction as a possible molecular target.

---

**Diagram:**

- **N-terminal region (~411 aa)**
  - dsRBD I/II
- **Helicase Core Domains (~355 aa)**
  - ATP binding/hydrolysis motif,
  - DNA binding motif, DEXH domain,
  - nucleic acid unwinding domain, RNA binding domain
- **C-terminal region (~503 aa)**
  - RGG-box

---

**Legend:**

- Dicer, TRBP, Ago2 1-272 aa
- RNA Pol II 282-380 aa
- p65 1-250 aa
- CBP 1-250 aa
- BRCA1 230-325 aa
- EWS-FLI1 Binding Site 630-1020 aa
- EGFR 623-1270 aa

---
Figure 3: EWS-FLI1 binds in a unique region of RHA
RHA was investigated as a partner protein of EWS-FLI1 based on its homology to peptides from the screening of an M13 phage library. The 630-1020 amino acid fragment of RHA binds to EWS-FLI1. RHA is known to interact with several other proteins, but most of these interactions occur in the amino-terminal region, including CBP, RNA polymerase II, BRCA1, p65, and others. Meanwhile, at this point, only EGFR has been identified to bind in the same region of EWS-FLI1, and therefore disrupting the interaction between RHA and EWS-FLI1 using a small molecular inhibitor may be a potential therapeutic target.

E. Disordered Proteins

1. General Overview

The structure-function paradigm of proteins states that the amino acid structure of a protein dictates its 3-D structure, which thus determines the function of the protein. The paradigm implies that a structure is necessary for protein function, so functional proteins must exist in a fixed structure in their native states. However, in 1978, the crystal structure of proteins with known functions was shown to have regions with missing electron density[81, 82]. The same year, the tail of histone H5 was shown to be disordered through NMR[83]. These were the first examples of functional proteins containing regions of disorder, disproving the structure-function paradigm.

The Database of Protein Disorder (DisProt) defines intrinsically disordered proteins (IDPs) as “a protein that contains at least one experimentally determined disordered region”[84]. Proteins that lack disorder exist as dynamic ensembles with variance between atom positions and backbone Ramachandran angles that lack specific equilibrium values[85]. Since the first studies, dozens of examples of functional proteins with
disordered regions have been identified based on x-ray structure, NMR, or other biophysical techniques. Some of these proteins are fully disordered, while others contain both ordered and disordered regions. Structural disorder is found throughout eukaryotes, where estimates suggest that 5-15% of proteins are full disordered and 35-50% of proteins have at least one intrinsically disordered region of more than 30 amino acids[86]. The high natural abundance of IDPs suggests that they likely have important biological functions, and a search of Swiss-Prot functional keywords indicates that IDPs are involved in transcription and its regulation, signal transduction and cell cycle regulation, functioning of nucleic acid containing organelle, mRNA processing and splicing, and cytoskeleton organization[86-88]. Additionally, sites of posttranslational modification, including acetylation, hydroxylation, ubiquitination, methylation, and phosphorylation, are frequently associated with regions of intrinsic disorder[89].

Disordered proteins share similar characteristics that allow for prediction based on propensity-based characteristics (Table 1). IDPs lack bulky hydrophobic residues (Ile, Leu, Val) and aromatic amino acid residues (Trp, Tyr, and Phe), which form the hydrophobic core of a globular protein, along with Cys and Asn. Together, these amino acids are referred to as order-promoting residues. Alternatively, Ala, Arg, Gly, Gln, Ser, Pro, Glu, and Lys are over-represented in IDPs, and are called disorder-promoting amino acids[90, 91]. As a result of this amino acid composition, IDPs have a low overall mean hydropathy and high mean net charge at neural pH, resulting in extreme pI values[92]. The high net charge results in charge-charge repulsion, while the low hydrophobicity results in less driving force for protein compaction, which leads to a protein lacking fixed structure. Plotting this
mean net charge versus mean hydropathy distinctly separates ordered and disordered proteins into distinct regions[93]. From order-promoting to disorder-promoting, amino acid residues can be ranked as follows: Trp, Phe, Tyr, Ile, Met, Leu, Val, Asn, Cys, Thr, Ala, Gly, Arg, Asp, His, Gln, Lys, Ser, Glu, and Pro[94].

<table>
<thead>
<tr>
<th>Predictor</th>
<th>What is Predicted</th>
<th>Based On</th>
</tr>
</thead>
<tbody>
<tr>
<td>PONDR</td>
<td>Regions that are not rigid</td>
<td>AA composition, flexibility, hydropathy</td>
</tr>
<tr>
<td>Disopred2</td>
<td>Lack of regular secondary structure</td>
<td>Classifiers trained on PSI-BLAST profiles</td>
</tr>
<tr>
<td>Globplot</td>
<td>High propensity for globularity</td>
<td>Russell/Linding scale of disorder</td>
</tr>
<tr>
<td>Disembl</td>
<td>Lack of regular secondary structure</td>
<td>Neural networks from X-ray structure data</td>
</tr>
<tr>
<td>FoldIndex</td>
<td>Low hydrophobicity and high net charge</td>
<td>Charge/hydropathy</td>
</tr>
<tr>
<td>PreLink</td>
<td>Regions expected to be unstructured in all conditions regardless of binding partner</td>
<td>Compositional bias and low hydrophobic cluster content</td>
</tr>
<tr>
<td>IUPred</td>
<td>Regions lacking 3D-structure under native conditions</td>
<td>Energy resulting from inter-residue interactions, estimated from local amino acid composition</td>
</tr>
<tr>
<td>RONN</td>
<td>Regions lacking 3D-structure under native conditions</td>
<td>Bio-basis function neural network trained on disordered proteins</td>
</tr>
</tbody>
</table>

Table 1: Disordered protein predictors.
Many online disordered protein predictors determine regions of order or disorder based on the amino acid sequence. Eight examples are listed, along with what structural features are predicted and how those predictions are made.

2. Functions of Disordered Proteins

Regions of intrinsic disorder are thought to provide many functional advantages, including an increase in surface area for protein-protein interactions, structural plasticity
allowing for interactions with several targets, the ability to fold upon binding, and accessible posttranslational modification sites. Splicing and regulation are also considered to be very important functions of IDPs[95]. Because disordered regions can bind with high specificity and low affinity, interactions with other proteins are easily dispersed, allowing for the ability to both turn a signal on and turn it back off[90]. Furthermore, IDPs are able to bind with many other structurally diverse proteins and ligands, allowing them to function as a hub in a signaling network. Regions of intrinsic disorder allow a protein to overcome the steric restrictions found in a structured protein, which allows for larger interaction surfaces in protein-protein and protein-ligand complexes than those with rigid partners.

Alternative splicing occurs when two or more mature mRNAs are produced from a single precursor pre-mRNA, joining exons together and excluding the introns[96, 97]. This allows for the production of multiple proteins from one gene, as 40-60% of genes generate proteins using alternative splicing[98]. The loss of an exon sequence from a structured protein could result in a change in protein folding, leading to a loss of function. However, most splice variants are found to fold into the same overall structure, with only slight changes to the structure that could be functionally important[99, 100]. This is possible because alternatively spliced regions of mRNA often code for intrinsically disordered regions of proteins, resulting in multiple splice variants that do not have an overall effect on protein structure.

IDPs can be classified by their function into two groups: entropic chains and proteins involved in molecular recognition[101, 102]. Those proteins involved in
molecular recognition often have the ability to fold and become partially ordered upon binding to their targets. This binding can be permanent or transient[103]. Many disordered proteins, both natively unfolded and those with regions of intrinsic disorder, undergo disorder-to-order transitions upon binding to specific targets[92]. IDPs often bind their partners via short molecular recognition features (MoRFs)[104], around 20 residues in length, which allow for this rapid and reversible interaction through either multiple interaction sites or regions specific to distinct partners[101, 102]. IDPs can fold into different structures following binding to target proteins with different functional outcomes[105]. Examples of coupled folding and binding in an IDP include a cation binding to alpha-synuclein[106], the SNARE complex formation from the disordered components Snc1 and Sec9[107], and the interaction of intrinsically disordered caldesmon with its binding partner, calmodulin[108]. The large decrease in conformational entropy, which happens in the disorder-to-order transition, uncouples specificity from binding strength, allowing highly specific interactions to be reversible as required for inducible responses in signaling and regulation.

3. Disorder in Transcription Factors

Localized regions of transcription factors, such as the DBD, are abundant in the Protein Data Bank, but there are no 3D structures of whole eukaryotic transcription factors[109]. This absence suggests that transcription factors may contain large regions of intrinsic disorder. Computational analysis indicates that over 80% of transcription factors contain regions of disorder, mostly in the activation regions[110], and up to 49% of the
entire sequence of transcription factors contains regions of intrinsic disorder[109]. Many transcription factors have been shown to contain MoRFs[111] which gain an ordered structure following binding to a partner. Additionally, disordered regions are hypothesized to contribute to the nuclear localization of transcription factors, as nuclear localization signals are typically located outside of structured domains[109].

Disorder is particularly useful in transcription factors, as one segment can bind to many partners and adopt different conformations upon binding to each segment. The p53 protein is a transcription factor located at the center of a signaling network where it regulates over 150 genes[112]. p53 has multiple regions of disorder, and four different partners bind to the same segment on the C-terminus end. Depending on which partner binds, four different disorder-to-order transitions can occur, resulting in an alpha helix, a beta-sheet, or two different coils[113].

4. Disordered Proteins in Cancer

Intrinsic disorder is found within many proteins associated with various diseases, likely due to the importance of signaling in many disorders. The D², or “disorder in disorders” concept, was introduced because of the commonality of IDPs in many diseases[114]. Cancer-associated proteins are predicted to contain 1.7-times more regions of long disorder than general eukaryotic proteins and 6.1-times more regions of long disorder than ordered proteins in the Protein Data Bank[95].

Loss of p53 function is considered to be a major factor in cancer development[115]. p53 must interact with many other proteins, many of which are transcription factors or
activators/inhibitors of p53 transactivation activity. Binding with so many different proteins is possible because the terminal domains of p53 are intrinsically disordered, while the DNA binding domain has a fixed structure[113]. Over 70% of the interactions occur in regions of intrinsic disorder[113], indicating the importance of disorder in modulating interactions with other proteins.

Chromosomal translocations, such as EWS-FLI1, tend to have few structural domains that are connected by long uncharacterized regions of disorder. An analysis recently determined that fusion protein partners are, on average, twice as disordered as all human proteins combined[116]. These translocation breakpoints generally avoid globular domains, which allows the resultant fusion proteins to function and often gain oncogenic fusions despite losing a large region of the original protein.

5. EWS-FLI1 is a Disordered Protein

The low overall hydrophobicity of the sequence of EWS-FLI1 indicates that the protein is likely disordered under native conditions. In addition to amino acid composition, a disordered protein can be recognized using the following equation:

\[
\langle H \rangle \leq \langle H \rangle_b = \frac{\langle R \rangle + 1.15}{2.79}
\]

where \(\langle H \rangle\) and \(\langle R \rangle\) are the mean hydrophobicity and mean net charge for a given protein and \(\langle H \rangle_b\) is the boundary mean hydrophobicity, below which a region with a given \(\langle R \rangle\) is predicted to be disordered[92]. The values for EWS-FLI1 are \(\langle H \rangle = 0.39\), \(\langle R \rangle = 0.08\), and \(\langle H \rangle_b = 0.42\), which indicates that EWS-FLI1 should be disordered under native conditions[117]. Disordered protein predictors also indicate that EWS-FLI1 is more than
75% disordered[118] (Figure 3), which is further confirmed through circular dichroism of refolded recombinant protein with spectra exhibiting characteristics of an unfolded conformation[117]. EWS-FLI1 does contain the structured ets DNA binding domain on the FLI1 region, but the highly disordered N-terminal domain is responsible for almost all protein-protein interactions, likely including those that lead to oncogenesis[118]. As EWS-FLI1 lacks direct enzymatic activity, it is thought to interact with other proteins on the basis of its intrinsic disorder[119].
**Figure 4: Protein disorder predictors indicate that EWS-FLI1 is a disordered protein.** Importing the sequence of EWS-FLI1 to four different disordered protein predictors all indicate that EWS-FLI1 has significant disorder in the EWS region and the C-terminal end of the FLI1 region. (A) Globplot predicts the propensity for globularity based on the Russell/Linding scale of disorder. (B) DisEMBL predicts lack of regular secondary structure based on neural networks from crystallography data. (C) IUPred predicts regions lack 3D structure under native conditions based on energy estimated from inter-residue interactions. (D) FoldIndex predicts low hydrophobicity and high net charge.

**F. Targeting Transcription Factors & Disordered Proteins**

Many transcription factors have an increased activity in many cancers and may be the most direct therapeutic target, as there are fewer players involved than upstream signaling enzymes[9]. These disordered transcription factors are involved in recognition, regulation, and signaling, and the ability to disrupt the action of these proteins through the interruption of necessary binding partners is extremely attractive for drug design[120, 121]. Drugs such as tamoxifen and bicalutamide, which are used in the treatments of breast and prostate cancers respectively, directly target the nuclear receptor class of transcription factors[122]. However, transcription factors outside the nuclear receptor family are widely considered to be ‘undruggable’ since they lack ligand-binding domains or intrinsic enzymatic activities[123]. Unlike targeting an enzyme at the ATP binding site, development of a therapeutic target for a transcription factor requires very specific disruption of a DNA-protein or protein-protein interaction[9].

IDPs often have a great potential for binding to small molecules due to higher induced-fit sampling properties and the potential for multiple binding sites to small
molecules[124]. As a result of this promiscuous binding, finding initial compounds for protein-protein interactions is more likely. Disordered proteins frequently bind their partners in MoRF regions, which become ordered after binding[111, 125, 126]. To disrupt interactions involving IDPs, one can either target the binding site on a structured partner, similar to traditional drug discovery approaches, or target a possible binding site in a region of disorder, especially if both binding partners lack fixed structure. Drugs targeting disordered regions may induce a structure that prevents the IDP from binding to its partner. Interactions using this disorder-to-order transition resulting in folding after binding have a very low binding free energy per unit area of interaction, and therefore would be relatively easy to block with small molecules[85].

Several IDPs have already been targeted for drug discovery, such as the kinase and phosphorylation sites, which are located within areas of intrinsic disorder[127]. The c-Myc oncoprotein is overexpressed in most human cancers and regulates many genes involved in growth and differentiation. c-Myc can be targeted by small molecules, which bind to the disordered region of c-Myc and unfold the Myc-Max dimer by binding to the helix-loop-helix-leucine zipper region[128-130]. These small molecules are able to inhibit the Myc-Max interaction by forcing the target protein from a structured heterodimer into a disordered form. Another example is the discovery of small, drug-like molecules that block Mdm2’s interaction with p53 by binding to the groove of Mdm2. Nutlin-2, a small molecule, mimics the crucial residues of the p53 fragment when bound to Mdm2[131]. Small molecules have also been developed to modulate CBP and C/EBPα, which inhibit transcription or mimic the functions of transcriptional activation domains[132]. A novel
strategy to quench EWS-FLI1 function would block protein interactions with EWS-FLI1 using inhibitory small molecules.

**G. Stereoselectivity**

A racemic mixture contains a 1:1 ratio of two enantiomers. Each enantiomer rotates plane-polarized light in an opposite direction; either right hand (d) or left hand (l), while the racemic mixture does not rotate polarized light. Stereopure drugs containing only one enantiomer can often reduce the total dose of drug given and minimize any toxicity resulting from the inactive enantiomer[133]. A stereoselective drug contains one therapeutically active enantiomer, while the other is inactive and considered an impurity[134].

Differences in enantiomers can range from harmless, such as estrone with an inactive (-) form, to penicillamine, which is extremely toxic in the L-form[135]. When dosed in the (R)-form, thalidomide acts as a sedative, but treatment with the (S)-enantiomer is highly teratogenic. However, chiral inversion of thalidomide occurs within the body, and both the (S)- and (R)-forms are interconverted with both oral and intravenous dosing[136]. Metabolic differences between enantiomers are also common. In propranolol, the (-) isomer is 100 times more active than the (+) isomer, but the oral clearance of the (+) enantiomer has been reported to be up to 50% higher. This stereospecific metabolism means that at equal plasma concentrations, the racemic mixture is 2 to 3 fold more potent when dosed orally in comparison to IV injection[137]. In addition to a difference in activity, enantiomers can differ in action, sometimes even
behaving antagonistically[138]. In the case of timolol, each enantiomer individually treats a separate condition: the (R)-enantiomer exerts localized β-adrenoreceptor action for the treatment of glaucoma, while the (S)-enantiomer has potent β-adrenergic antagonist activity and is used as a β-blocker in hypertension[139, 140].

In 1992, the FDA revised its policy on the registration of new drugs, and now requires separate pharmacological, pharmacokinetic, and toxicological profiles for each enantiomer in a racemic mixture[141]. As a result of these changes requiring additional toxicology studies of racemic compounds, single enantiomer drugs have since dominated the pharmaceuticals approved in the United States[142].

H. Rational and Hypothesis

Disruption of protein-protein interactions using small molecules is a rapidly growing field. Target-oriented synthesis allows for the development of compounds based on the 3D structure of the protein, but the lack of structural data available for EWS-FLI1 prevents this form of drug synthesis. Due to its considerable motion in solution, there is no crystallographic data available and the protein cannot be concentrated to the levels required for NMR. While structural information is available on EWS and FLI1, the changes in charge distribution indicate that the structure of the fusion protein is likely to vary widely from its wild-type partner proteins.

Since EWS-FLI1 is tumor-specific and necessary for the survival of ESFT, it is an ideal molecular target. Previous studies indicate that RHA enhances EWS-FLI1-driven oncogenesis, and interruption of this protein-protein complex by small molecules may be a
therapeutic target. Three main aims were established to help achieve the goal of targeting a therapy to EWS-FLI1. First, we hypothesized that the oncogenic function of EWS-FLI1 can be eliminated or disrupted through the use of small molecule inhibitors targeting the critical protein interaction with RHA. Second, we set out to characterize the points of interaction between EWS-FLI1 and RHA. Finally, we postulated that the structure and function of EWS-FLI1 could be investigated through the development of functional assays to study the biochemical and biophysical interactions of the protein with small molecules. As more information is obtained regarding the structure of EWS-FLI1 and its interaction with RHA, we hoped to use target oriented synthesis to modify the structure of the lead compound to better inhibit the binding of RHA to EWS-FLI1.

The first aim, established in the Toretsky laboratory, was successfully accomplished. The team identified a small molecule that is cytotoxic to ESFT cells and is able to disrupt the binding between EWS-FLI1 and RHA. My thesis work was critical to the establishment of the mechanism of YK-4-279 through validation experiments that indirectly lead to the conclusion that YK-4-279 was binding to EWS-FLI1 and causing the disruption of RHA. My hypothesis extended this work based on the knowledge that YK-4-279 has a chiral center and I show that there is a significant difference in activity between the two enantiomers. The active enantiomer is able to disrupt binding between EWS-FLI1 and RHA in an immunoprecipitation assay and blocks the transcriptional activity of EWS-FLI1, while the inactive enantiomer cannot. Enantiomeric effects were also established in cytotoxicity assays and caspase assays, where up to a log-fold difference is seen between the active enantiomer and the racemic molecule. This work is significant for its identification of a
single enantiomer effect upon a protein interaction, suggesting that small molecule targeting of intrinsically disordered proteins can be specific.

While it is known that EWS-FLI1 and RHA interact, little is known about the structural relationship between the two proteins. I hoped to identify the region of EWS-FLI1 to which RHA binds, as this knowledge will enhance our understanding of the functional relationship between the two proteins. The identification of physical interaction points could further lead to the revision of our small molecule to fit into a binding pocket or better disrupt the RHA/EWS-FLI1 interaction. Working towards this aim required an improvement in protein purification of recombinant EWS-FLI1 and RHA, which lead to successful crosslinking of RHA and EWS-FLI1. Unfortunately, we have not yet been able to appropriately interpret the resulting data following digestion of the crosslinked fragment. Instead, I am currently attempting to determine the region of EWS-FLI1 to which our small molecule binds through native state proteolysis. I am able to show regions of order and disorder of EWS-FLI1, and hope to use this assay to determine the region of the protein that is protected from digestion by the small molecule.

Lastly, I aimed to develop a high-throughput method for testing small molecules and their ability to bind to or inhibit EWS-FLI1. This assay would allow for both rapid screening of newly developed analogues and further allow us to investigate the structure and function of EWS-FLI1 by determining which chemical moieties alter binding to the protein. Additional structural information could be obtained from determining which small molecules are also able to bind to wild-type EWS or FLI1 in addition to EWS-FLI1. While we initially identified the small molecule lead compound through a surface plasmon
resonance screen, not all the derivative small molecules bound well using Biacore. Initially, fluorescence polarization was used to measure RHA peptide displacement from EWS-FLI1, but the reaction is both time consuming and reagent limiting, so we evaluated new strategies using a 96-well plate format with the goal of high-throughput data collection. Techniques examined include fluorescence polarization, protein melting curves, and AlphaScreen. Optimization of a luciferase assay and ELISA have recently provided some success as a higher-throughput method for determining small molecule function, and data can help predict the synthesis of future analogues. Data obtained from the developed assay will be used as input to a quantitative -structure activity relationship (QSAR) method. This QSAR method will predict biological activity from chemical structure and the knowledge gained will aid in the optimization and design of future analogues. These data are a contribution to the future development of our small molecules in an era where novel approaches to cancer therapy are critical to improving patient care.
Chapter II: Materials and Methods
A. Cell Lines

TC32, TC71, RDES, MMH-ES-1, STA-ET 7.2, A4573, COLOPL, CHP100, ES925, and PC3 cells were grown in RPMI with 10% fetal bovine serum (FBS) and 1% HEPES. SKES cells were grown in McCoy’s 5A medium with 15% FBS. MCF7, MDA-MB-231, PANC1, ASPC1, SKNAS, SKLMS, HEK293, and Cos7 cells were grown in DMEM with 10% FBS. A673 cells were grown in DMEM with 10% FBS and 1% sodium pyruvate. A673-inducible EWS-FLI1 shRNA (A673i) cells were grown in DMEM with 10% FBS, 1% sodium pyruvate, 1% Pen-strep, 20 µg/ml blastocidine, and 200 µg/ml zeocin.

B. Small Molecules and Peptides

The initial surface plasmon resonance screen test compounds, including NSC635437, from the Developmental Therapeutics Program of the National Cancer Institute, National Institutes of Health. The initial series of 14 derivatives were synthesized by Yali Kong. All small molecules were dissolved in 100% DMSO. The E9R peptide (PPPLDAVIE) and FITC-E9R were obtained from Bio-Synthesis.

C. Cellular Proliferation Assays

Cells were grown in colorless media at a plating density of 5,000 – 15,000 cells/well, depending on cell line, in a 96-well plate. Small molecule or vehicle alone (DMSO) were added to cells in appropriate growth media the day after plating. After three days, viable cells were quantified using either MTT or WST-1 (Roche) according to the manufacturer’s
protocol. IC$_{50}$ values were calculated by sigmoidal dose-response curve fit using Prism Graphpad 4.0.

**D. Caspase-3 Activity**

Cells were plated at 100,000 – 200,000 cells/well, depending on cell line, in a 12 well plate. Small molecule was added the following day. 18 hours after addition of small molecule, lysates were collected using the manufacturer’s protocol for the AC-DEVD-AMC substrate (BD Bioscience Pharmingen). Caspase-3 substrate was incubated for 2 hours with protein lysate and fluorescence from cleaved substrate was measured in a fluorometer. Fluorescent signal was then normalized to protein lysate via BCA assay.

**E. Surface Plasmon Resonance**

Recombinant EWS-FLI1 was bound on the surface of a CM5 chip to a concentration of 16,000 RU. DNA oligonucleotides were used for quality control before each run to confirm appropriate binding to EWS-FLI1. Running buffers used for the displacement assay were HBSP buffer + 0.1% DMSO and HBSP buffer + 0.1% DMSO + 10µM YK-4—279. E9R peptide was diluted with the appropriate running buffer for each condition. Experiments were performed on a Biacore T100.

**F. Fluorescent Polarization**

For measuring the binding affinity between FITC-E9R and recombinant EWS-FLI1, a fluorescent polarization assay was established. Increasing concentrations of FITC-E9R
were added to a fixed concentration of EWS-FLI1 (4.8µM) to obtain a saturated binding curve. The assay was performed in 20 mM Tris, 500 mM NaCl, and 0.67 M imidazole, pH 7.4. FP was analyzed in a Photon Technology International QuantaMaster fluorometer equipped with polymer sheet polarizers at an excitation wavelength of 495 nm and an emission wavelength of 517 nm. To determine the effect of YK-4-279, competition binding assays were performed. Increasing concentrations of YK-4-279 were added to a fixed concentration of EWS-FLI1 (4.8 µM) and FITC-E9R (3.2 µM, as determined from saturated binding curve) with the same buffer and instrumental settings as described above.

G. ELISA

EWS-FLI1 was used to coat the surface of a 96 well plate (MaxiSorb) at a concentration of 150 ng/well in 100 µL buffer containing 20 mM Tris, 500 mM sodium chloride, and 1 M imidazole. Following overnight incubation, the plate was blocked with 4% BSA in PBS and washed with PBS + 0.1% Tween-20. Small molecules were added at concentrations of 1 µM, 3 µM, 10 µM, and 30 µM, followed by RHA (300 ng/well) and allowed to incubate overnight. Following washing, incubation with RHA antibody (Everest) and anti-goat secondary, protein binding was detected using a TMB Peroxidase EIA Substrate Kit (Bio-Rad) per the manufacturer’s instructions.

H. Nuclear Prep Immunoprecipitation

TC32 were grown to ~70% confluency and cells were treated for 15 hours with small molecule. Nuclear lysate was collected using the Active Motif Magnetic Co-IP Kit
(Active Motif). Protein concentration was determined using bicinchoninic acid protein assay for each lysate (Pierce) and lysate was bound to 1 μg of antibody overnight at 4 °C on a rotating axis with addition of same concentration of small molecule as used for 15 hour treatment. Magnetic beads (Active Motif) were added to the lysates and tumbled for 2 hours at 4 °C.

I. Preparation of Whole Cell Lysate

Cells were grown to ~70% confluency, washed with cold PBS, and lysed using an appropriate volume of phospholysis buffer (50 mM HEPES pH 7.9, 100 mM NaCl, 4 mM NaPP, 10 mM NaF, 1% Triton-X) with the addition of phosphatase inhibitors (2 mM vanadate, 1 mM PMSF, 2 μg/ml aprotonin, and 2 μg/ml leupeptin) on ice. Lysed cells were scraped and pelleted in a microfuge at 12,000 rpm for 10 minutes at 4 °C.

J. Immunoblotting

Protein lysates were resolved using PAGE and transferred to a polyvinylidene difluoride membrane (Millipore). Membranes were blocked in 5% nonfat dry milk in TTBS (20mM Tris-HCl, 150mM NaCl, 0.5% Tween 20) for 2 hours. Dilutions for primary antibodies were: FLI1 (Santa Cruz Biotechnologies) 1:1000, RHA (Abcam) 1:1000, and anti-actin-HRP (C-11, Santa Cruz Biotechnologies) 1:1000 in either TTBS + 0.5% BSA or 5% nonfat dry milk for 2 hours or overnight. Membranes were washed 4 times in 1x TTBS. Horseradish peroxidase-linked anti-rabbit or anti-mouse secondary antibody (GE Healthcare) was added for 1 hour. Blots were again washed 4 times in 1X TTBS and
detection was carried out using Millipore Immobilon Western Chemiluminescent HRP Substrate per the manufacturer's instructions (Millipore Corp.). Chemiluminescence was detected using a Fujifilm LAS-3000 imaging system. Densitometry values were obtained using either Multigauge software (FUJIFILM) or ImageJ software.

**K. Luciferase Reporter Assay**

EWS-FLI1 activity was assessed using an NR0B1 luciferase construct containing 25 EWS-FLI1 binding sites. Cos7 cells were transiently transfected with the NR0B1 luciferase reporter[143] and full length EWS-FLI1 or empty vector control (Clneo) with Fugene-6 (Roche) according to the manufacturer's protocol. Two hours after transfection, cells were treated with 0.3, 1, and 3 μM of YK-4-279, (S)-YK-4-279, and (R)-YK-4-279. Luciferase activity was measured 16 hours after drug treatment. All luciferase assays were performed using a luciferase assay kit according to the manufacturer's protocol (Promega).

**L. Quantitative Real Time PCR**

A673i cells were treated with 1 μM small molecule for 8 hours. Positive control cells were treated with tetracycline for 72 hours to knock down expression of EWS-FLI1. Total RNA was extracted using an RNeasy Micro Kit (Qiagen) and reverse transcribed using the Transcriptor First Strand cDNA Synthesis Kit (Roche) per the manufacturer's instructions. qPCR was performed using an Eppendorf Mastercycler Realplex with FastStart Universal SYBR Green Master (ROX) (Roche) with primers for 18S, CyclinD1a, and CyclinD1b. Data
were analyzed for expression relative to 18S using the comparative Ct method. Data from four separate experiments performed in duplicate were averaged.

**M. Animal Studies**

Animals were injected with either 1 million TC71 or 2 million A4573 or SKES cells in 100 µL of Hank’s Buffered Saline Solution (HBSS). Injections were done orthotopically into the gastrocnemius muscle of 4-week old severe combined immunodeficient-beige (SCID) mice (Charles River). Following the development of a palpable tumor, mice were randomized into treatment groups receiving daily intraperitoneal injections of vehicle (100% DMSO) or small molecule (YK-4-279, (S)-YK-4-279, or (R)-YK-4-279) at a dose of 200 mg/kg in 20 µL total volume. Tumor size was measured daily and volume was calculated using the formula \( V = \frac{4}{3}\pi \times \left( \frac{D}{2} \right)^2 \times \left( \frac{d}{6} \right) \) where D is the longest diameter of the tumor and \( d_2 \) is the shorter diameter. Mice were weighed once per week. Animals were euthanized when tumor size reached 1.6 cm³ or following weight loss greater than 10% of total body weight. The Georgetown University Institutional Animal Care and Use Committee approved all animal studies.

**N. GST-EWS-FLI1 Fragment Pull-Down**

The EWS-FLI1 fragments, corresponding to amino acids 1 – 246 (EWS fragment), 281 – 497 (FLI1 fragment), and 245 – 281 (breakpoint fragment) were cloned into pET-42(a) vectors, which contain a GST tag, along with full-length EWS-FLI1 and CBP. Fragments were transformed into BL21 codon-plus bacteria. Bacterial lysates were
tumbled with glutathione beads and then mixed with HEK293 lysate as a source of RHA. Following overnight tumbling, the beads were washed and run on SDS-PAGE. Blotting was done first for RHA and then for GST to ensure all constructs were present at the expected size. The experiment was repeated 5 times.

O. Protein Purification

EWS-FLI1 and RHA 630-1020 were grown in BL21 cells to an O.D. of 0.5 – 0.6 and induced with 1mM IPTG. Following an additional 2 (RHA) – 4 (EWS-FLI1) hours of shaking at 25 °C, protein was pelleted by centrifugation. The inclusion bodies were treated with BugBuster (Novagen) per manufacturer’s instructions. The resulting pellet was denatured in binding buffer containing 20 mM Tris-HCl, 500 mM NaCl, 6 M guanadinium, and 50 mM imidazole, pH 7.4, and filtered through a 0.22 micron filter. Protein was loaded on a HiTrap Chelating HP Column (GE) that is pre-charged with 0.1 M NiSO₄. The purification process was done on an AKTA explorer or AKTA purifier system. Protein was loaded on the column at 0.5 mL/minute and washed with buffer containing 20 mM Tris-HCl, 500 mM NaCl, 8 M urea, and 50 mM imidazole. Next, the protein on the column was refolded using a slow gradient to a buffer containing 20 mM Tris-HCl, 500 mM NaCl, and 50 mM imidazole at 0.5 mL/minute. Elution from the column is done in a buffer containing 2 M imidazole, and the protein elutes in fractions equivalent to 1.2 M imidazole.
P. Protein Crosslinking

Recombinant EWS-FLI1 and RHA were dialyzed in a 6 step process from 1 M imidazole to 1M urea (20 mM phosphate, 500 mM NaCl) with one hour per step. Crosslinking was completed using bis[sulfosuccinimidyl] substrate that was prepared immediately before use at a 25 mM concentration. After a 30 minute incubation at room temperature, the reaction was quenched using 50 mM Tris-HCl pH 7.5 and incubated for 15 minutes at room temperature. Reactions were run on SDS-PAGE and coomassie stained. Bands of interest were isolated from the gel, digested with trypsin overnight, and analyzed using LC MS/MS via QSTAR.

Q. Protein Melting

Protein and SYPRO orange dye (Sigma-Aldrich) concentrations were determined through a titration experiment for optimal melting. A melting curve analysis was run from 20 °C to 85 °C over 20 minutes. Experiments were performed in 96 well plates. Data was collected and analyzed on a LightCycler 480 Real-Time PCR system (Roche) or a Mastercycler (Eppendorf).

R. Circular Dichroism

Circular dichroism was studied using EWS-FLI1 in 20 mM sodium phosphate, 500 mM NaCl, 1 M urea, pH 7.4 in a 1 cm cell. Data is averaged from 5 or more independent scans. Scans were performed on a Jasco J-715 spectropolarimeter from 250 – 180 nm with a bandwidth of 5.0 nm, a response of 4 seconds, and a scan speed of 10 nm/minute.
**S. Native State Proteolysis**

Proteases (proteinase K, trypsin, and chymotrypsin) were prepared at a 0.5 mg/ml enzyme stock and diluted to 5 ng/µL working concentration. The buffer for proteinase K was 20 mM HEPES pH 6.8, 200 mM KCl, 1 mM CaCl2, and 1 mM DTT. The buffer for trypsin was 20 mM Tris pH 8.0, 200 mM KCl, 1 mM CaCl2, and 1 mM DTT. The buffer for chymotrypsin was 20 mM Tris pH 7.5, 200 mM KCl, 1 mM CaCl2, and 1 mM DTT. EWS-FLI1 (60 µg/reaction) was added to each protease. At each time point, aliquots were removed from the protein-enzyme mixture and added to an equal volume of 10% TCA to precipitate the protein. Samples were centrifuged and washed three times with acetone to remove the TCA salts. Pellets were air dried and stored at room temperature. Immediately before running a gel, pellets were resuspended in sample buffer and run on a 15% SDS-PAGE gel. The gel was coomassie stained and appropriate bands were excised and sent for analysis by mass spectrometry.

**T. AlphaScreen**

AlphaScreen (Perkin-Elmer) assay used nickel chelate donor beads and anti-FITC acceptor beads. EWS-FLI1, FITC-E9R peptide, and small molecule were added to a 96-well plate and incubated for 30 minutes at room temperature. Anti-FITC acceptor beads were added, followed by an hour-long incubation, and then addition of the nickel chelate donor beads. The plate was read an EnVision Multilabel Reader.
Chapter III: Results
A. Targeting EWS-FLI1

Oncogenic fusion proteins, such as EWS-FLI1, are excellent therapeutic targets as they are only located within the tumor. However, there are currently no agents targeted toward transcription factors, which are often considered to be ‘undruggable.’ A considerable body of evidence is accruing that refutes this claim based upon the intrinsic disorder of transcription factors. Our previous studies show that RHA enhances the oncogenesis of EWS-FLI1, and we are able to interrupt this protein-protein complex by small molecule inhibitors. We show that there is a significant difference in activity between the two enantiomers of our chiral compound in binding assays, transcriptional assays, and cellular assays. Our findings indicate that only one enantiomer of our small molecule is able to specifically target a protein-protein interaction.

1. NSC635437 is the lead compound for targeting EWS-FLI1

Due to the lack of structural data of EWS-FLI1, target oriented synthesis of small molecules was not an option. Instead, plates of small molecules from the NCI Diversity Library were screened using surface plasmon resonance (SPR). Following initial single concentration binding screening of 300 compounds, 20 compounds were analyzed for binding kinetics. The resulting kinetic data, along with cell cytotoxicity data from Ewing cells, were presented to a team of medicinal chemists. Each possible compound was ranked based on binding affinity to EWS-FLI1, cytotoxicity data, and ‘druggability’. From this, NSC635437 was selected as the lead compound for binding to EWS-FLI1 with a $K_d$ of 2.34 µM and a structure indicative of favorable drug-like properties. To determine the
potency of the compound and its specificity towards Ewing cells, NSC635437 was tested in a panel of Ewing sarcoma cell lines (TC71, A673, and TC32) (Figure 5A) and non-Ewing cell lines (RD, SKLMS, and SKNAS) (Figure 5B). Small molecule was added every other day for six days and cellular proliferation was measured using MTT or WST. Data was normalized to a vehicle (DMSO) control. The cytotoxicity panel indicates that the small molecule NSC635437 had an IC50 between 10 and 20 µM in the three ESFT cells tested, and was relatively specific for EWS-FLI1 containing cells, as no cytotoxicity was seen in two of the three non-Ewing cell lines.
Figure 5: NSC635437 is the lead compound.
NSC635437 was tested in a panel of (A) Ewing sarcoma cells and (B) non-Ewing cell lines. Small molecule was added in micromolar log concentrations ranging from 0.5\(\mu\)M to 30 \(\mu\)M on days 1, 3, and 5. Cellular proliferation was measured on day 6 using either an MTT or WST assay. The data shown represents the mean ± standard deviation of the mean from one experiment with 3 replicates for each dosage.

2. Optimized derivative YK-4-279 has higher specificity and potency

After identification of NSC635437, fourteen analogues were synthesized by Yali Kong using an aromatic optimization strategy. All analogues were tested in TC32 (ESFT) cells to determine the IC\(_{50}\) values of the small molecules in comparison to the lead compound, now called YK-4-275. A significant improvement was seen between YK-4-275 and YK-4-279 over the other synthesized analogues (Figure 6A). YK-4-279, which has a para-oxymethyl group in place of the para-chlorine of the lead compound (Figure 6B), reduced the IC\(_{50}\) from 20 \(\mu\)M to 0.9 \(\mu\)M in TC32 cells (Figure 6C). In comparison to non-transformed HEK293 cells, YK-4-279 demonstrated log-fold more potency in the ESFT-containing cell line (Figure 5D). From these studies, we determined that YK-4-279 was the best targeted of the derivatives toward EWS-FLI1.
Figure 6: The derivative of YK-4-279 improves potency.

The lead compound, NSC635437, was synthesized as YK-4-279 along with fourteen analogues with variations to the phenyl ring. (A) ESFT cells (TC32) were treated for 5 days with analogues and viability was determined using MTT. Cells were normalized to DMSO control. Values are shown as percentage of cell survival over the vehicle alone control of one experiment with three replicates for each treatment. Prism v4.0 was used to calculate a nonlinear sigmoidal dose-response curve fit. (B) The most potent and specific analogue is YK-4-279, which substitutes a para-oxyethyl group for the para-chlorine on YK-4-275. (C) TC32 cells were treated with YK-4-275 or YK-4-279 for 72 hours. YK-4-279 reduces the IC50 from 20 µM to 0.9 µM. (D) To show selectivity to EWS-FLI1, TC32 cells were compared to non-transformed (HEK293) cells. YK-4-279 is approximately log-fold more toxic to the ESFT cells. The data shown represents the mean ± standard deviation of the mean from one experiment with three replicates for each treatment.
Next, we treated a panel of five ESFT cells with YK-4-279 and showed IC$_{50}$ values between 0.5 µM and 2 µM, while non-Ewing and non-transformed cells had IC$_{50}$ values in excess of 25 µM (Figure 7A). In order to test for apoptotic activity as an insight into the mechanism of cell death, treated cells were assayed for the activation of caspase-3 and cleavage of AMC-DEVD. The caspase-3 assay was chosen as it allows for a higher-throughput assay compared to other traditional methods for testing apoptosis, such as DNA fragmentation assays or annexin V staining. Following a 24 hour treatment with YK-4-279 or doxorubicin (Figure 7B), TC32 cells were shown to undergo apoptosis. The 10 µM YK-4-279 treated cells induced more apoptosis than doxorubicin (1 µM), which is currently used in the treatment of ESFT patients. We next treated a panel of cells with YK-4-279 and saw caspase-3 activity increased from 2.5- to 4.5-fold in four ESFT cell lines (TC32, A4573, TC71, and ES925 cells), but less than a 1.5-fold caspase increase in all the non-ESFT or non-transformed cell lines tested (Figure 7C). Caspase data was normalized to protein quantity and then compared to a vehicle-treated control. Together, these results indicate both the potency of YK-4-279 and the specificity for cells containing EWS-FLI1.
Figure 7: YK-4-279 is a potent and specific inhibitor of ESFT in biological assays.

(A) ESFT and non-ESFT cells were treated with YK-4-279 for 3 days in culture to establish the IC$_{50}$ using WST. YK-4-279 is more than log-fold more toxic to ESFT cells. Data shown represents averages of at least 3 experiments performed in triplicate. (B) Apoptosis of TC32 cells is shown by activation of caspase-3 and cleavage of AMC-DEVD following a 24 hour treatment with YK-4-279 or doxorubicin. Data shown represents the mean ± standard deviation from one experiment with 3 replicates for each treatment. (C) YK-4-279 induces caspase-3 activity in 4 ESFT cell lines, but not in 5 non-ESFT or 3 non-transformed cell lines. Data represents the averages of at least 3 experiments ± standard deviation performed in duplicate.
3. YK-4-279 structure is similar to the E9R RHA peptide

Previous experiments identified a peptide (E9R) that binds to recombinant EWS-FLI1 and corresponds to amino acids 823 to 832 of the proximal HA2 region of RHA[77]. The peptide, PPPLDAVIEA, prevents RHA from binding to EWS-FLI1 and significantly decreases cell viability in ESFT cells. A collaborator on the drug discovery project, Sivanesan Dakshanamurthy, modeled the E9R peptide and YK-4-279, and noticed an overlap between the two compounds (Figure 8). A three-dimensional model of the E9R peptide shows that the PPP portion (in purple) of the PPPLDAVIEA motif mimics YK-4-279 (green).

![Figure 8: YK-4-279 displays similar structure to the E9R peptide.](image)

YK-4-279 (green) mimics the PPP region of PPPLDAVIEA, the E9R peptide corresponding to amino acids 823 to 832 of RHA (in purple). Modeling done by Sivanesan Dakshanamurthy.
4. YK-4-279 is able to displace E9R from EWS-FLI1

After establishing the specificity and potency in ESFT cells, we next looked at the ability of YK-4-279 to interrupt the binding between EWS-FLI1 and RHA. To test this in a cell free assay, we used the E9R peptide as an RHA surrogate. To investigate the ability of YK-4-279 to inhibit binding between EWS-FLI1 and the E9R RHA mimic, we first used SPR. Recombinant EWS-FLI1 was bound to the surface of a CM5 sensor chip. Initial kinetic binding studies were done to determine binding of the E9R peptide to EWS-FLI1. Next, an SPR displacement assay was developed using running buffer containing E9R compared to running buffer containing E9R plus 10 μM YK-4-279. The YK-4-279 is able to displace the E9R peptide from binding to EWS-FLI1, reducing the binding of 64 μM E9R from 17 RU to 7 RU (Figure 9).
**Figure 9: E9R peptide directly binds to EWS-FLI1**

Recombinant EWS-FLI1 was bound to a CM5 sensor chip. E9R peptide or E9R peptide with the addition of YK-4-279 was passed over the surface of the chip. SPR displacement assay shows 10 uM YK-4-279 reduces the binding of 64 uM E9R from 17 R.U. (red) to 7 R.U. (light red). Experiment was repeated twice in duplicate.

To further corroborate the ability of YK-4-279 to displace the E9R peptide from YK-4-279, we used fluorescence polarization (Figure 10A). We added increasing concentrations of FITC-E9R to 4.8 µM of EWS-FLI1 in an appropriate buffer to determine a saturated binding curve (Figure 10B) and determined that 3.2 µM E9R resulted in the largest dynamic range between free and bound E9R (Figure 10C). At 10 µM, YK-4-279 is able to completely displace 3.2 µM E9R (as determined from the saturated binding curve) from EWS-FLI1 binding, and we were able to titrate this back up to baseline (Figure 10D). This indicates that YK-4-279 functions similarly to E9R, which was previously shown to prevent EWS-FLI1 from binding to RHA[77].
**Figure 10: YK-4-279 is able to displace E9R peptide from EWS-FLI1**

(A) A fluorescence polarization assay was used to measure RHA peptide (E9R) displacement from EWS-FLI1. (B) A saturated binding curve first determined the amount of FITC-E9R necessary to bind 4.8 µM of EWS-FLI1. (C) 3.2 µM of FITC-E9R provided the largest dynamic range between unbound and bound peptide. (D) 3.2 µM of FITC-E9R is competitively inhibited by increasing concentrations of YK-4-279. The assay was performed in 20 mM Tris, 500 mM NaCl, and 0.67 M imidazole, pH 7.4, and was analyzed in a QuantaMaster fluorometer equipped with polymer sheet polarizers at an excitation wavelength of 495 nm and an emission wavelength of 517 nm. The data shown represents the mean ± standard deviation from one experiment with 3 replicates for each treatment. Experiment was repeated twice.

5. **(S)-YK-4-279 disrupts binding of RHA with EWS-FLI1**

The structure of YK-4-279 contains a chiral center. In order to evaluate the enantiospecific effects of YK-4-279, we tested each enantiomer in comparison to the racemate. YK-4-279 was resolved into its individual enantiomers using preparative high-pressure liquid chromatography (HPLC) on a chiral column[144]. The HPLC performed by AMRI indicates chemical and chiral purity of both fractions, with chemical purity and enantiomeric excess values of >99%.

To determine if enantiospecific blocking of the EWS-FLI1/RHA interaction occurs, binding between EWS-FLI1 and RHA was measured using ELISA. Wells coated with EWS-FLI1 were treated with a range from 1 µM to 30 µM YK-4-279, (S)-YK-4-279, (R)-YK-4-279, or vehicle. Both the racemate and (S)-YK-4-279 are able to inhibit the protein-protein interaction with as little as 1 µM, while (R)-YK-4-279 does not inhibit the binding at 30 µM (Figure 11A). Multiple ELISA runs suggest that (S)-YK-4-279 is only slightly more potent as disrupting the complex based upon increased dissociation at 1 µM.
Next, ESFT cells were treated with YK-4-279, (S)-YK-4-279, and (R)-YK-4-279, followed by immunoprecipitation of EWS-FLI1. We treated TC32 cells with 10 µM of the racemic or enantiomeric small molecule for 15 hours, consistent with the K_D value of YK-4-279[77]. Immunoblotting showed coimmunoprecipitation of EWS-FLI1 with RHA in the presence of vehicle or (R)-YK-4-279, but a significant reduction in complexed RHA in the lysates from cells treated with either YK-4-279 or (S)-YK-4-279 (Figure 11B). The control IgG lanes do not indicate the pulldown of either EWS-FLI1 or RHA.

To determine the relative potency of YK-4-279 and (S)-YK-4-279 to disrupt the binding between EWS-FLI1 and RHA, we next titrated down the amount of small molecule added to the TC32 cells to determine if (S)-YK-4-279 was able to inhibit binding at half the dose of the racemic YK-4-279. Cells were treated with 0.3, 1, 3, or 10 µM of small molecule for 15 hours before immunoprecipitation (Figure 11C). Cells treated with vehicle or (R)-YK-4-279 were again able to pull down RHA, as were cells treated with 0.3, 1, or 3 µM of YK-4-279 and 0.3 or 1 µM (S)-YK-4-279. However, by 10 µM YK-4-279 or 3 µM (S)-YK-4-279, the binding between EWS-FLI1 and RHA is interrupted. Treatment did not affect the EWS-FLI1 or RHA levels (Figure 11D). Densitometry of RHA relative to FLI1 was calculated (Figure 11E) and shows an IC_50 of 4.9 µM for the racemic and 1.8 µM for the (S)-YK-4-279 enantiomer, with a significant difference between vehicle and cells treated with 0.3 µM (S)-YK-4-279, but not for cells treated with racemic molecule. Experiment was repeated three times and one representative blot is shown.
Figure 11: (S)-YK-4-279 is able to disrupt the binding between EWS-FLI1 and RHA.

(A) An ELISA measured binding between EWS-FLI1 and RHA. 1, 3, 10, or 30 µM small molecule was added to the ELISA plate to disrupt the protein-protein interaction (**, p <0.05 compared to vehicle control, using a two-tailed Student’s t-test). (B) TC32 cells were treated with 10 µM YK-4-279, (S)-YK-4-279, or (R)-YK-4-279 for 15 hours. Immunoprecipitation was performed using 525 µg of nuclear lysate with 1 µg of FLI1 pAb or rabbit IgG as a control. EWS-FLI1 and RHA were then detected via western blot. (C) TC32 cells were treated with 0.3, 1, 3, and 10 µM of small molecule for 15 hours. Immunoprecipitation was performed using 250 µg of protein. (D) 10% input for each IP condition. (E) Densitometry was calculated for each band and the ratio of RHA/FLI1 was graphed.
6. EWS-FLI1 functional activity is reduced by only one enantiomer

Since (S)-YK-4-279 is able to block the binding between EWS-FLI1 and RHA, we next investigated the ability of the enantiomers to reduce transcriptional activity. We transfected COS7 cells with EWS-FLI1 and the NR0B1-luciferase reporter plasmid, which contains 25 EWS-FLI1 DNA-binding sites. YK-4-279 and (S)-YK-4-279 were able to inhibit EWS-FLI1 transcriptional activity in a dose-dependent manner compared to vehicle-treated cells (Figure 12A), with an IC$_{50}$ of 0.96 µM for racemic and 0.75 µM for the (S)-enantiomer. The (R)-enantiomer did not show significant reduction from control, with an IC$_{50}$ > 3 µM. (Figure 12B). At a dose of 0.3 µM, luciferase activity was significantly reduced in cells treated with (S)-YK-4-279 compared to vehicle (p < 0.05), while luciferase activity after treatment with racemic compound compared to control was not significant. The (R)-enantiomer did not show significant reduction from control. Data is averaged from three experiments, each performed in triplicate. Expression of EWS-FLI1 into COS7 was similar in each experiment and a representative Western blot is shown (Figure 12C).
Figure 12: (S)-YK-4-279 blocks transcriptional activity of EWS-FLI1

(A) COS7 cells were transfected with EWS-FLI1 and the EWS-FLI1-responsive promoter NR0B1. 2 hours after transfection, cells were treated for 18 hours with YK-4-279, (S)-YK-4-279, and (R)-YK-4-279 at 0.3, 1, and 3 μM. Relative luciferase activity was normalized per mg of protein by nanodrop. Transfection assays were performed in triplicate (**, p < 0.05 compared to vehicle control, using a two-tailed Student’s t-test). (B) IC_{50} values of luciferase inhibition of treated cells compared to vehicle treated cells were calculated using a non-linear sigmoidal dose-response curve fit. (C) Expression of EWS-FLI1 in transfected cells was detected by Western blot.

EWS-FLI1 has been shown to increase Cyclin D1 levels by altering the D1b/D1a mRNA level in ESFT cells through effects on transcript elongation[143]. Since racemic YK-4-279 significantly decreases Cyclin D1 levels in TC32 cells[77], we evaluated whether this effect was enantiospecific. We treated A673i cells[145] with 1 μM of YK-4-279 or enantiomer for 8 hours and then analyzed mRNA for levels of Cyclin D1a and Cyclin D1b using quantitative RT-PCR. As a positive control, we reduced EWS-FLI1 protein levels in A673i with an induced shRNA. With 1 μM racemic YK-4-279 treatment, there is a reduction in the D1b/D1a levels (Figure 13A). The (S)-YK-4-279 is approximately 35% more potent at the D1b/D1a reduction than racemic and reduced this ratio to a level equivalent to the EWS-FLI1 reduction. Consistent with other experiments, the (R)-YK-4-279 cells maintained a D1b/D1a ratio equivalent to the vehicle treated cells. There was no decrease in EWS-FLI1 protein expression in cells treated with small molecule (Figure 13B). Therefore, the enantiospecific effect is due to changes in mRNA levels and not protein.
Figure 13: (S)-YK-4-279 reduces Cyclin D1b/D1a levels in A673i cells.
(A) A673i cells were treated with vehicle or 1 µM small molecule for 8 hours. Control cells were treated with tetracycline for 72 hours to reduce EWS-FLI1. q-RT-PCR was used to quantify changes of CyclinD1a and CyclinD1b mRNA levels. Expression was normalized to 18S. Data is averaged from four experiments performed in duplicate (**, p < 0.05 compared to vehicle, using a two-tailed Student's t-test). (B) Immunoblot of total cell lysates from one of the four experiments averaged in (A).

7. YK-4-279 demonstrates enantiospecific cellular effects

After determining that (S)-YK-4-279 is able to block the interaction between EWS-FLI1 and RHA while (R)-YK-4-279 does not, we tested the small molecules for cytotoxicity in a panel of ESFT cell lines compared to cell lines that lack ets rearrangements. TC32,
along with six other cell lines expressing EWS-FLI1, were treated with either a vehicle or dose of small molecule ranging from 0.1 to 30 µM of compound for three days (Figure 14A). Each of these cell lines demonstrated significant cytotoxicity to (S)-YK-4-279 while the (R)-YK-4-279 enantiomer demonstrated no specific toxicity. Experiments were repeated three times in triplicate and mean IC₅₀ values ranged from 0.33 µM to 1.83 µM for racemic YK-4-279, 0.16 µM to 0.87 µM for (S)-YK-4-279, and 11.69 µM to 25.98 µM for (R)-YK-4-279 (Figure 14B, Table 2), indicating that (S)-YK-4-279 is the active enantiomer in cytotoxicity studies. The effects of the enantiomers were next evaluated in a panel of carcinoma cell lines lacking ets rearrangements, including PC3, MCF7, MDA-MB-231, PANC1, and ASPC1 (Figure 14C, Table 2). Average IC₅₀ values for the five non-ESFT cell lines were 8.88 µM for YK-4-279, 6.86 µM for (S)-YK-4-279, and >30 µM for (R)-YK-4-279. YK-4-279 and the active enantiomer are therefore relatively specific for ESFT cells when compared to cancer cell lines lacking EWS-FLI1.

A4573 ESFT cells were treated with increasing concentrations of YK-4-279 and (S)-YK-4-279 for 18 hours and caspase-3 enzymatic activity was measured as an indication of apoptotic activity (Figure 14D). A panel of Ewing and non-Ewing cell lines were then treated with 10 µM small molecule for 18 hours and assayed for caspase activation. Caspase-3 activity increased from 3-fold to 18-fold upon treatment with YK-4-279 and 5-fold to 20-fold with (S)-YK-4-279 in ESFT cells (TC32, RDES, SKES, A4573, MMH-ES-1, STAE7) (Figure 14E), but showed no more than a 2-fold increase in apoptosis upon treatment with racemic or either enantiomer in non-ESFT cells (MCF7, MDA-MB-231, PANC1) (Figure 14F), showing the specificity to those cell lines containing EWS-FLI1.
Figure 14: (S)-YK-4-279 is the active enantiomer in biological assays

(A) A panel of ESFT and non-ESFT cells were treated with a dose range of YK-4-279, (S)-YK-4-279, and (R)-YK-4-279. Cell reduction was measured by WST after 72 hours of treatment. One representative graph from a cytotoxicity assay is shown. Graphs show IC\textsubscript{50} values for (B) ESFT and (C) non-ESFT cells (**, p < 0.05, using a two-tailed Student’s t-test). (D) A4573 cells were assayed for caspase-3 activation with increasing concentrations of
YK-4-279 and (S)-YK-4-279, along with the (R)-YK-4-279 concentration corresponding to the highest treatment dose, for 18 hours. Graph shows fold caspase activation of treated cell lysates compared to control cell lysates. ESFT (E) and non-ESFT (F) cells were treated with 10 µM small molecule for 18 hours. Graph shows fold caspase-3 activity of treated cells compared to untreated cells.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Histology</th>
<th>µM IC₅₀ at 3 days (+/- SEM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YK-4-279</td>
<td>(S)-YK-4-279</td>
</tr>
<tr>
<td>TC32</td>
<td>ESFT (Type 1)</td>
<td>1.02 (0.89)</td>
</tr>
<tr>
<td>TC71</td>
<td>ESFT (Type 1)</td>
<td>1.83 (0.41)</td>
</tr>
<tr>
<td>RDES</td>
<td>ESFT (Type 2)</td>
<td>1.03 (0.19)</td>
</tr>
<tr>
<td>SKES</td>
<td>ESFT (Type 2)</td>
<td>0.33 (0.03)</td>
</tr>
<tr>
<td>MMH-ES-1</td>
<td>ESFT (Type 2)</td>
<td>0.94 (0.13)</td>
</tr>
<tr>
<td>STA-ET 7.2</td>
<td>ESFT (Type 2)</td>
<td>0.60 (0.04)</td>
</tr>
<tr>
<td>A4573</td>
<td>ESFT (Type 3)</td>
<td>1.46 (0.31)</td>
</tr>
<tr>
<td>PC3</td>
<td>prostate</td>
<td>4.95 (3.62)</td>
</tr>
<tr>
<td>MCF7</td>
<td>breast</td>
<td>22.82 (7.19)</td>
</tr>
<tr>
<td>MDA-MB-231</td>
<td>breast</td>
<td>0.82 (0.02)</td>
</tr>
<tr>
<td>PANC1</td>
<td>pancreatic</td>
<td>1.51 (0.65)</td>
</tr>
<tr>
<td>ASPC1</td>
<td>pancreatic</td>
<td>14.28 (3.50)</td>
</tr>
</tbody>
</table>

Table 2: IC₅₀ values in a panel of cells.
Cells were plated in a 96-well plate and treated with YK-4-279, (S)-YK-4-279, or (R)-YK-4-279 at concentrations between 0.1 µM and 30 µM for 72 hours. Plates were treated with WST to determine cell viability. IC₅₀ values were calculated using a non-linear sigmoidal dose-response curve fit.

8. Racemic YK-4-279 inhibits in vivo growth of ESFT

To compare the racemic with the enantiomers in vivo, we established orthotopic ESFT tumors in severe combined immunodeficient-beige mice with TC71, A4573, and SKES cell lines. Preliminary studies compared vehicle and YK-4-279 to determine the best cell line for future studies. Initial treatment with YK-4-279 showed delayed tumor growth in
comparison to vehicle (Figure 15A). Based on the preliminary study, A4573 cells were used for a comparison of YK-4-279, (S)-YK-4-279, and (R)-YK-4-279. Two million A4573 cells were injected into the gastrocnemius of mice and palpable tumors developed within 2 to 3 weeks. Once masses were identified, animals were randomly assigned to either control or treatment groups. Mice received intraperitoneal injections of vehicle (DMSO) or small molecule at dose of 200 mg/kg in 20 µL every day. Tumors were measured daily and mice were euthanized when tumor volume reached 1.6 cm³. Despite the rapid growth rate of tumors, mice treated with the racemic small molecule exhibited reduced tumor growth rate (p = 0.0167) and survived longer (p = 0.028) when compared to vehicle. Following 8 days of treatment, mice dosed with the racemic small molecule (p = 0.0167) and the (S)-enantiomer (p = 0.0368) exhibited reduced tumor growth compared to the vehicle (Figure 15B). While there was no statistical significance between either the (S)- or (R)-enantiomers, mice treated with the (S)-enantiomer did have increased survival when compared to vehicle (p = 0.182), in contrast to the (R)-enantiomer (p = 0.602) (Figure 15C).
Figure 15: YK-4-279 inhibits ESFT cell proliferation in vivo in a mouse xenograft model.

Xenografts were established with orthotopic injections of 2 million A4573 cells (ESFT) into SCID mice. After formation of a palpable tumor, mice were treated daily with intraperitoneal injections of vehicle (n = 6) or 200 mg/kg (n = 6) of YK-4-279 until tumors reached 1.6 cm$^3$. (A) Initial treatment with YK-4-279 showed delayed tumor growth in comparison to vehicle (p > 0.05, curve regression and F-test). (B) Following 8 days of treatment, mice treated with racemic small molecule and the (S)-enantiomer exhibited significant reduction in tumor growth compared to vehicle (p > 0.05). Mice treated with the racemate survived longer when compared to vehicle (p > 0.05, Kaplan-Meyer analysis). While there was no statistical significance between either the (S)- or the (R)-enantiomers regarding survival, mice treated with (S)-enantiomer did have a longer median survival when compared to vehicle (p = 0.182) in comparison to (R)-enantiomer (p = 0.602).

B. Characterization of interaction between EWS-FLI1 and RHA

Given the challenges of drug design targeted to EWS-FLI1, we hypothesize that a better understanding of the binding between EWS-FLI1 and RHA will allow us to better alter the lead compound to block this protein-protein interaction. Initially, three fragments of EWS-FLI1 were used to pull down RHA to identify the site of the interaction. However, multiple repeats did not indicate a successful co-precipitation of the RHA by any of the fragments. Full length EWS-FLI1 was then crosslinked with RHA following improvements to protein purification, and the resulting crosslinked region was analyzed using mass spectrometry. Due to the scrambled order of individual amino acids in the resulting mass spectrometry identified peptide fragments, we have yet to identify which region of the proteins is actually crosslinked. We next experimentally determined the regions of
disorder in EWS-FLI1, and attempted to determine a binding site for YK-4-279 based on the sensitivity of disordered regions to protease digestion.

1. **Full-length EWS-FLI1 is required to pull down RHA**

   In order to better understand the interaction between EWS-FLI1 and RHA, we attempted to identify the region of the EWS-FLI1 protein to which the RHA molecule binds. Given the challenges of drug design targeted to EWS-FLI1, we hypothesized that a better understanding of the binding between these two proteins would allow us to alter the lead compound to block this protein-protein interaction. As no information is available regarding the structure of full-length EWS-FLI1 or RHA, we generated fragments of EWS-FLI1 for use in a pull-down assay with RHA. Three fragments of EWS-FLI1, consisting of the EWS domain (amino acids 1-246), the FLI1 domain (amino acids 281-497), and a small domain including the breakpoint region that encompasses part of the EWS and FLI1 fragments (amino acids 245-281), were cloned into a pET-42(a) GST-tagged vector and purified for use in the experiment (Figure 16A). GST-CBP was cloned into a pGEX vector for use as a positive control. To locate the site of the interaction between EWS-FLI1 and RHA, the three fragments were mixed with RHA obtained from HEK293 lysate in a pull-down experiment. The resulting complex was run on a Western and immunoblotted for both RHA (Figure 16B) and GST (Figure 16C). The experiment was repeated 6 times and 3 representations of the RHA blot are shown.

   While the various EWS-FLI1 fragments were present on the blot at the expected size, multiple repeats of the experiment did not clearly indicate a successful co-precipitation of
the RHA by any of the fragments. However, full-length GST-tagged EWS-FLI1 was able to pull down RHA from HEK293 cells (Figure 16D). The unsuccessful pull-down by any of the fragments may be due to a folding of EWS-FLI1 in such a way that RHA binds to residues in both the EWS and the FLI1 portions of the protein.

Figure 16: Full length GST-EWS-FLI1 pulls down endogenous RHA, but fragments do not.
(A) EWS-FLI1 fragments were cloned into pET-42(a) vectors containing a GST tag. Fragments were tumbled with endogenous RHA from HEK293 cells and immunoblotted for RHA (B) and GST (C) to ensure all constructs were present. Experiment was repeated six times and three representative RHA blots are shown. (D) Full-length GST-EWS-FLI1 is able to pull down the RHA.
2. RHA 630-1020 fragment can be purified for use in binding experiments

In order to pursue binding studies between EWS-FLI1 and RHA, purified recombinant RHA was necessary. The 630-1020 fragment of RHA, which binds to EWS-FLI1, was initially cloned into the GST-tagged pGEX vector. However, the protein did not properly bind to or elute from glutathione beads or a GST spin column, hindering the purification of the protein. Increasing amounts of SDS were added to the beads to wash away impurities, but this did not help the purification. A switch in buffer to a GST sonication buffer (120 mM KOAc, 20 mM Tris-OAC pH 7.9, 1 mM EDTA, 10% glycerol) followed by sonication, addition of N-laryl sarcosyl (1.5%), a second sonication, and neutralization with 1% Triton-X100 also did not help the GST-RHA to stick to the glutathione beads. We hypothesized that the RHA fragment was folding around the GST tag, thereby preventing the tag from binding to the glutathione beads.

The 630-1020 fragment of RHA was then cloned into a pET-28(b) His-tagged vector and protein was optimized for expression in BL21 cells. Purification attempts included chelating columns charged with NiSO₄, CoCl₂, CuSO₄, and ZnSO₄. Salt concentrations were varied from 150 mM NaCl up to 1 M NaCl. Native purification conditions using a lysis buffer containing 50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0, resulted in losing all the RHA in the pass. Ion exchange was attempted using various columns for different exchange, including SP-FF (strong cation exchanger), Q-FF (strong anion exchanger), and DEAE-FF (weak anion exchanger). Buffer with ethanolamine, Glychine-NaOH, and piperazine were all used in addition to lysozyme and sonication. Ammonium
sulfate precipitation was attempted to precipitate out RHA and then resuspend in an appropriate buffer. Figure 17 outlines many of the purification attempts and the outcome.
Figure 17: RHA 630-1020 fragment purification challenges.
Attempts to purify the His-tagged RHA 630-1020 fragment included FPLC elution with imidazole; low salt, physiological salt, or high salt buffers; charging a chelating column with NiSO₄, CoCl₂, CuSO₄, or ZnSO₄, purification in recommended native or denaturing conditions for His-tagged proteins, and ion exchange.

Successfully purified RHA resulted from denaturing the protein in 6 M guanadinium and refolding on a nickel-charged chelating column, similar to the purification process for EWS-FLI1. Following loading the protein on the column, the column buffer was slowly changed to 20 mM Tris, 500 mM NaCl, pH 7.4, to refold the protein while still on the chelating column. Protein was then eluted with an imidazole gradient. Purified RHA was obtained at a concentration of approximately 10.5 µM (Figure 18A) but is prone to precipitation if too concentrated. To overcome this aggregation at higher concentration levels, 1 mL RHA fractions are eluted directly into 1 mL of the elution buffer, immediately diluting the protein to a soluble concentration. To ensure that the RHA fragment was properly refolded, we confirmed that it was still able to bind to EWS-FLI1 in a pull-down assay. EWS-FLI1 was indeed able to pull-down RHA, indicating that RHA was properly refolded within the binding site for EWS-FLI1 (Figure 18B).
Figure 18: EWS-FLI1 pulls down purified recombinant RHA 630-1020.

(A) An inclusion body pellet of RHA was treated with BugBuster to extract protein from the insoluble fraction. The resulting pellet was resuspended in 35 mL of 20 mM Tris, 500 mM NaCl and 6 M guanadinium. The suspension was filtered and loaded on a 1 mL prepacked metal chelating column on an AKTA Explorer chromatography system. Column buffer was slowly changed to a refolding buffer containing 20 mM Tris and 500 mM NaCl. Elution buffer contained 2 M imidazole. Collected fractions were centrifuged to remove precipitated protein. (B) Recombinant EWS-FLI1 and RHA were used in an immunoprecipitation assay to determine if EWS-FLI1 bound to the refolded RHA fragment.
3. EWS-FLI1 purification improvements result in increased purity.

Although initial purification of EWS-FLI1 was done in a two-step process, first binding to a chelating column and eluting with a pH gradient, and then refolding on a second chelating column and elution with EDTA[117], the purification process evolved to a one-step purification with elution by imidazole. Unfortunately, there were many non-specific bands contaminating EWS-FLI1 after the imidazole purification. A sizing column was initially used to separate out the protein from the contaminants. However, the sizing column diluted the EWS-FLI1 to unusable levels and, as a result of intrinsic disorder, EWS-FLI1 cannot be concentrated without precipitation.

Binding and wash buffers for His-tagged proteins typically contain low concentrations of imidazole, in the 5 – 20 mM range, to help prevent nonspecific binding of proteins that contain histidine clusters. The low concentration is necessary to prevent elution of the protein, which generally occurs with 150 – 300 mM imidazole. However, EWS-FLI1, possibly due to its disordered nature, does not elute off the column until the elution gradient reaches a concentration of 1.2 M imidazole. As a result, we increased the imidazole concentration in the binding and wash buffers from 5 mM to 50 mM to further eliminate non-specific binding. The increase in imidazole significantly improved the purity of the recombinant EWS-FLI1 (Figure 19A). To ensure that the increase in imidazole did not affect protein refolding, we tested protein purified in each buffer using SPR. DNA oligonucleotides that bind to EWS-FLI1 were used for quality control and indicate an improvement in binding to the protein (Figure 19B: 5 mM buffers, 14.012 RU protein, $K_D = 9.531 \times 10^{-9}$ M; Figure 18C: 50 mM buffers, 13.725 RU protein, $K_D = 1.234 \times 10^{-8}$ M). The
improved EWS-FLI1 purification buffers were used for all experiments that follow; however, the initial SPR displacement study and fluorescent polarization assay used proteins purified with the 5 mM imidazole buffers.
Figure 19: Purity of recombinant EWS-FLI1 is improved with increase in imidazole.

(A) An increase of imidazole in binding, wash, and refolding buffers from 5 mM to 50 mM significantly improved the purity of recombinant EWS-FLI1. Increasing quantities of EWS-FLI1 from equivalent fractions (Fraction 11) were run on a 10% SDS-PAGE gel and coomassie stained for the identification of protein. EWS-FLI1 from the (B) 5 mM buffer preparation was immobilized on a chip at 14.021 RU or (C) 50 mM preparation was immobilized on a chip at 13.725 RU and a kinetic experiment using DNA oligonucleotides determined binding to the protein. The binding affinity constant, $K_D$, is $9.531 \times 10^{-9}$ M for the 5mM buffer preparation and $1.234 \times 10^{-8}$ M for the 50mM buffer preparation.

4. Recombinant EWS-FLI1 is dependent on imidazole in the buffer.

Imidazole is commonly used as an elution agent in the purification of His-tagged proteins. His-tagged proteins bind to metal ions attached to the beads in a chelating column. The excess of imidazole in the elution buffer displaces the His-tag from the nickel, thereby eluting off the protein. However, the presence of imidazole in a buffer is problematic in several assays, including crosslinking with BS3 (due to the presence of a primary amine) and circular dichroism (imidazole absorbs below 220 nm).

Significant time and effort was invested in attempts to dialyze out the imidazole from purified EWS-FLI1. 2, 3, 4, 6, and 8 step dialysis was tried to lower the imidazole from 1.2 M down to 0 M, but protein aggregated each time during the dialysis process. To stabilize the protein, a DNA oligonucleotide that binds to EWS-FLI1 was added to the dialysis cuvette. The addition of glycerol or 0.1% Triton did not prevent protein aggregation. In the event the protein aggregation was due to high concentration levels, we diluted the protein in the elution buffer prior to multi-step dialysis. We also tried replacing the 500 mM NaCl with the next salt in the Hoffmeister series to help with protein solubility,
attempting both (NH₄)₂SO₄ and KCl in one and multi-step dialyses. Attempts included both simultaneously exchanging salts and removal imidazole and the exchange of salts first, followed by the removal of imidazole. Unfortunately, all these attempts at protein dialysis resulted in precipitation of EWS-FLI1 within the dialysis cuvette.

We attempted to dilute down the concentration of imidazole in the buffer for experiments (circular dichroism, protein crosslinking, AlphaScreen) by doing a 1:1 dilution of a fraction of EWS-FLI1 in a buffer containing no imidazole. While the protein did not appear to precipitate with the dilution of imidazole, protein melting studies indicated a significant change in melting following the dilution with 0 M imidazole buffer compared to an equivalent dilution with 1.2 M imidazole buffer. As the proteins were diluted the same amount, and all components in the buffer were identical except imidazole concentration (1.2 M versus 0.6 M imidazole), the change is thought to be due to a change in protein stability in the reduced imidazole buffer (Figure 20). The negative first derivative drawing of the melting curve should show a single negative peak, as seen in EWS-FLI1 with 1.2 M imidazole. There is a positive peak in the imidazole-diluted sample, indicative of a loss of protein stability (personal communication, Steve Martin, Roche field specialist).
Figure 20: Dilution of EWS-FLI1 changes the profile of the protein melting curve.

EWS-FLI1 was diluted 1:1 in either elution buffer (20 mM Tris-HCl, 500 mM NaCl, 1.2 M imidazole, pH 7.4) or an imidazole-free buffer (20 mM Tris-HCl, 500 mM NaCl, pH 7.4). Resultant protein was either in 1.2 M (plotted in blue) or 0.6 M imidazole (plotted in red). SYPRO Orange dye, added to a final concentration of 6.66X, was incubated with 5 µM protein and underwent a melting curve analysis from 20 °C – 85 °C over 20 minutes. Experiments were performed with 10 replicates each in a 96 well plate. Data was collected and analyzed on a Roche LightCycler 480 Real-Time PCR system.

Next, at the suggestion of Vladimir Uversky (Associate Professor, College of Medicine, University of South Florida), we tried dialysis with the addition of 1 M urea, enough to stabilize the protein without unfolding any structure. The most dialysis success was seen with this addition of 1 M urea. A 6-step dialysis brought the imidazole content down to 1 M, 0.8 M, 0.6 M, 0.4 M, 0.2 M, and finally 0 M imidazole, while keeping a constant 20 mM Tris-Hcl, 500 mM NaCl, and 1 M urea throughout all the buffers. 250 µL EWS-FLI1 in a 0.1 – 0.5 mL capacity cuvette progressed through the series of buffers over the course
of 3 hours. Following the final dialysis step, EWS-FLI1 was centrifuged to remove any aggregated protein and run on a gel for coomassie staining (Figure 21). While a substantial amount of protein was lost during dialysis, this is the only dialysis attempt that resulted in any EWS-FLI1 remaining after dialysis for use in further experiments.

![Pre-Dialysis Post-Dialysis BSA Standards](image)

**Figure 21: Addition of urea improves EWS-FLI1 solubility without imidazole.**

EWS-FLI1 was purified using 50 mM imidazole buffers. Following purification, 250 mL underwent a 6-step dialysis in buffers containing 20 mM Tris-HCl, 500 mM NaCl, 1 M urea, pH 7.4, and decreasing concentrations of imidazole from 1 M to 0 M. The dialysis cuvette remained in each buffer for 30-60 minutes. Following completion of dialysis, EWS-FLI1 was centrifuged to remove any precipitated protein and run on an SDS-PAGE gel, along with un-dialyzed aliquots.

5. **Imidazole prevents structural analysis by circular dichroism.**

While circular dichroism (CD) cannot determine if a particular region of a protein is structured, it can characterize the secondary or tertiary structure of the overall protein. CD can determine whether protein-protein interactions, such as the interaction between EWS-FLI1 and RHA, alter the overall conformation of a protein, thus indicating a disorder-to-order binding transition. We hoped to use CD to look
at structural changes upon binding to the 630-1020 fragment of RHA, the E9R peptide, or YK-4-279 to determine if a disorder-to-order transition happens upon binding.

Unfortunately, buffers that absorb in the region of interest (250 – 190 nm) are not compatible with circular dichroism, including imidazole, Tris, and β-mercaptoethanol. EWS-FLI1 was purified in buffers containing 20 mM sodium phosphate in place of Tris, but imidazole content proved to be problematic. We first tried diluting the protein with imidazole-free buffer, but the imidazole still overwhelmed the spectra in the far-UV range. Next, we tried using protein dialyzed into urea, but the concentration was too low to generate an interpretable CD signal, even with an increase in cell size. To ensure proper use of the spectropolarimeter, we first used two ordered proteins, myoglobin and concanavalin A. Myoglobin has an 80% helical structure, which is seen in the double minimum at 222 nm and 208 nm (Figure 22A), while concanavalin A has 40% beta-sheet structure, exemplified by the single minimum between 210 nm and 255 nm and a single positive between 190 and 200 nm (Figure 22B). EWS-FLI1, however, does not exhibit structure characteristic of an α-helix, β-sheet, or random coil, as the protein is not concentrated enough for accurate circular dichroism measurements (Figure 22C).
**Figure 22:** Circular dichroism of EWS-FLI1 is hindered due to buffer constraints.

(A) Myoglobin and (B) concanavalin A give the expected spectra at a 10mg/mL concentration. An alpha-helical shape is seen in myoglobin, with two negative peaks at 222 nm and 208 nm. The beta-sheet is visible for concanavalin A, with a negative peak between 210 nm and 255 nm and a single positive peak between 190 nm and 200 nm. (C) No useful data is detected from the EWS-FLI1 spectrum due to low concentration following dialysis into the urea buffer.
6. EWS-FLI1 and RHA can be crosslinked.

Since the GST-tagged fragment pull-down did not provide information regarding the specific binding location of RHA on EWS-FLI1 and we were unable to study binding via circular dichroism, our next attempt was to crosslink the two proteins. Recombinant EWS-FLI1 and the 630-1020 fragment of RHA underwent rapid step-wise dialysis to remove imidazole from the elution buffer into a phosphate buffer containing 1 M urea. Crosslinking was completed using 0.25 mM and 1 mM bis[sulfosuccinimidyl] suberate. Following coomassie staining, bands are visible at 46 kDa (RHA), 68 kDa (EWS-FLI1), and 100 kDa (crosslinked band of RHA and EWS-FLI1 complex) (Figure 23).

The bands of interest (EWS-FLI1, RHA, and crosslinked) were isolated from the gel and analyzed using LC MS/MS via QSTAR. The 68 kDa band contained peptides identified as Friend leukemia integration 1 transcription factor (FLI1), while the 46 kDa band digestion resulted in a peptide from ATP-dependent RNA helicase A. The crosslinked band contained peptides from both these proteins, showing successful crosslinking of EWS-FLI1 and RHA, along with several unidentified proteins. A BLAST search was performed with unidentified peptides (confidence ≥ 90) from the crosslinked lane, but no more than 3 consecutive amino acids from either EWS-FLI1 or RHA are contained in the unidentified peptides (Table 3). Due to the scrambled order of individual amino acids in the identified peptide fragments, we were unable to identify which region of the proteins is actually crosslinked.
Figure 23: EWS-FLI1 and RHA crosslink in the presence of BS3.
Recombinant EWS-FLI1 and the 630-1020 fragment of RHA crosslink with 1mM bis[sulfosuccinimidyl] suberate. RHA is present at 46 kDa, EWS-FLI1 is present at 68 kDa, and the crosslinked band is present at 100 kDa. Circled bands were excised from the gel and sent for mass spectrometry analysis.

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>BLAST results</th>
<th>EWS-FLI1</th>
<th>RHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLEQQNQVLQTK</td>
<td>keratin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTEQRQPDPYQILGPTSSR</td>
<td>FLI1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELDALDANDELTPLGR</td>
<td>RHA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLDLDSIIAEVK</td>
<td>keratin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSDHVALLSVFQAWDDAR</td>
<td>RHA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDESKEGSIQGLEMQVER</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GASSSTNDASDPOTTKEFETLIK</td>
<td>SSS</td>
<td>STN</td>
<td></td>
</tr>
<tr>
<td>SPPLGGAQTISK</td>
<td>FLI1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDDFQLKGIVEEKFVK</td>
<td>zinc-finger CCHC domain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLDLDSIIAEVK</td>
<td>keratin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Mass spectrometry analysis of crosslinked bands
The bands of interest (EWS-FLI1, RHA, and crosslinked) were isolated from the gel, digested with trypsin, and analyzed using LC MS/MS via QSTAR. The 68 kDa band contained peptides identified as Friend leukemia integration 1 transcription factor (FLI1), while the 46 kDa band digestion resulted in a peptide from ATP-dependent RNA helicase A (RHA). The crosslinked band contained peptides from both these proteins, along with several unidentified proteins. A BLAST search was performed with the unidentified peptides from the crosslinked lane.

7. Native state proteolysis shows EWS-FLI1 regions of order and disorder.

After successfully crosslinking EWS-FLI1 and RHA but lacking the ability to determine a binding site with the mass spectrometry data, we focused on assays to determine the structure of EWS-FLI1, with the ultimate goal of identifying structural changes in a region following binding to a small molecule or peptide. EWS-FLI1 is predicted to be heavily disordered using a variety of different disorder predictors (Figure 4). To test this experimentally, we looked at the melting point of EWS-FLI1 in comparison to two ordered proteins, myoglobin and concanavalin A. The melting temperature, Tm, is defined as the temperature at which 50% of the protein is unfolded[146]. Since a disordered protein naturally contains regions of unfolding, the Tm may be lower than for a structured protein. SYPRO orange was added to the proteins of interest (5 µM) and a melting curve analysis from 20 °C to 85 °C was run over the course of 20 minutes (Figure 24). The melting temperature of myoglobin and concanavalin A were 74 °C and 79 °C, respectively, while the melting temperature of EWS-FLI1 was 42 °C, which may indicate disorder in EWS-FLI1 compared to either myoglobin or concanavalin A.
Figure 24: EWS-FLI1 has a lower melting point than two ordered proteins.
Myoglobin, concanavalin A, and EWS-FLI1 underwent a melting temperature analysis as detected by SYPRO orange dye. 5 µM protein was added to a 10X solution of SYPRO orange in a total volume of 20 µL. The melting temperature was taken from 20 °C to 85 °C over 20 minutes.

Next, we determined the location of the disordered regions of EWS-FLI1. Using a native state proteolysis experiment, EWS-FLI1 was subjected to proteases to determine the regions of disorder and order. Those areas of the protein that are disordered will be digested first, while the tightly ordered regions are protected. EWS-FLI1 was digested with proteinase K (191 possible cleavage sites), trypsin (21 possible cleavage sites), and chymotrypsin (58 possible cleavage sites). BSA, an ordered protein, was used as a control.

Protein was incubated with proteases for 1, 5, 15, 30, 60, or 120 minutes, followed by addition of 10% trichloroacidic acid to precipitate the protein and stop the reaction. Protein from each time point was run on a gel and coomassie stained (Figure 25A). Circled bands were cut from the gel and sent for analysis by mass spectrometry. A peptide
coverage map for EWS-FLI1 shows almost complete coverage at the initial zero time point and almost complete digestion by 30 minutes (Figure 25B). Sequence coverage is first lost from the EWS region of the protein, which is predicted to have significant disorder, followed by the C-terminus domain of FLI1. The peptides found in Band 4 correspond to the structured helix-turn-helix region in the ets DNA binding domain of FLI1 and match with the predicted regions of order (Figure 25C). No degradation is seen in BSA at the 120 minute time point, consistent with the known structural order of BSA.
Figure 25: Native state proteolysis identifies regions of EWS-FLI1 disorder.

(A) Recombinant EWS-FLI1 and BSA were subject to digestion by proteinase K, trypsin, and chymotrypsin for time points ranging from 1 to 120 minutes. Following the incubation period, 10% TCA was added to the reaction to precipitate the protein, therefore stopping the digestion. The remaining pellet was resuspended in sample buffer and run on an SDS-PAGE gel, followed by coomassie staining. The circled bands were excised and sent for mass spectrometry analysis. (B) Peptide coverage of EWS-FLI1, corresponding to each of the circled bands. (C) Comparison of the peptide sequences found in the band from the 30 minute time point to the predicted order of EWS-FLI1.

Native state proteolysis allows for a comparison of free and ligand-bound proteins. If a binding partner impacts the stability and structure of the protein, a shift will occur in the bands. The digestion of a certain region may take longer, as the ligand is protecting the region of the protein from rapid digestion. We then wanted to see if the addition of YK-4-279 changed the digestion profile. Trypsin was used for the digestion, as there are fewer cleavage sites, and EWS-FLI1 was incubated with either vehicle (DMSO) or 100 µM YK-4-279 prior to proteolysis. Two shifts in bands were identified in an initial experiment, and these bands were excised and sent for mass spectrometry analysis (Figure 26). However, a protein sequence coverage analysis did not identify peptides from EWS-FLI1. This may be due to technical error and this experiment will be pursued again in the future.
Figure 26: Addition of YK-4-279 changes digestion profile of EWS-FLI1. EWS-FLI1 was incubated with either (A) vehicle or (B) 100 µM YK-4-279 prior to digestion with trypsin. Two bands were identified that are present or more prominent in the YK-4-279 treated sample in comparison to the equivalent time point in the DMSO treated sample. Bands were excised and sent for mass spectrometry analysis, but no peptide fragments of EWS-FLI1 were identified.

C. Assay development for small molecule probes of structure and function

While we initially identified the small molecule NSC635437 through an SPR screen, not all the derivative small molecules bind well on the Biacore format. As more analogues have been developed, we wanted to generate an assay allowing for high-throughput testing of the binding of these compounds to EWS-FLI1. Given the challenges of drug design targeted to EWS-FLI1, we hypothesized that the optimization of a binding assay using small
molecules can predict and direct the design of more potent analogues. Initially, fluorescence polarization was used to measure RHA peptide displacement from EWS-FLI1, but the reaction is both time consuming and reagent limiting, so we evaluated new strategies using a 96-well plate format with the goal of high-throughput data collection. Strategies evaluated include fluorescence polarization via a plate reader, protein melting curves, and AlphaScreen.

1. Fluorescence Polarization

Initial results with fluorescence polarization resulted in binding between EWS-FLI1 and the E9R peptide that can be titrated down with an increase in concentration of small molecule. Although the assay would allow for comparison of various small molecules, the experiment required significant amounts of EWS-FLI1 and was very time consuming. The first approach to convert the assay to a higher-throughput system used a TECAN ULTRA 384 for 96-well plate fluorescence polarization assay with an excitation wavelength of 485 nm and an emission wavelength of 535 nm (Figure 27A). However, the readings resulted in high background noise and low signal. Less than a two-fold range was established between free and bound E9R, which does not provide a large enough dynamic range to titrate down the fluorescent signal (Figure 27B).
Figure 27: High-throughput FP does not yield a large dynamic range.

(A) FP was attempted in a black 96-well plate using a TECAN ULTRA 384 with an excitation wavelength of 485 nm and an emission wavelength of 535 nm. A saturated binding curve was used to determine the amount of FITC-E9R necessary for binding to EWS-FLI1. (B) 1.6 µM FITC-E9R provided the largest dynamic range between unbound and bound peptide, but low signal resulted in less than a two-fold dynamic range between free and bound E9R. Response units are plotted as milli-polarization.

2. Protein Melting Curves

Protein melting can determine the stability of protein in different conditions, including the interaction of the protein with a compound. An increase in Tm indicates a stabilizing effect, proportional to the concentration and binding affinity of the ligand[147, 148]. While we previously showed a difference in protein melting of EWS-FLI1 with a change in buffer (Figure 20), and the difference in Tm based on the level of disorder of a protein (Figure 24), there was unfortunately no significant difference in Tm upon the addition of small molecules. We know from previous studies that YK-4-279 binds to EWS-FLI1 while other analogues do not, but this difference was not reflected in the change in melting temperature between protein and protein plus small molecule (Figure 28).
addition of small molecules therefore likely does induce an overall structural change to the entire protein substantial enough to increase the stability of EWS-FLI1.

Figure 28: Addition of small molecule does not shift melting temperature of EWS-FLI1.

EWS-FLI1 (5 µM) was incubated with small molecules (range from 3 – 100 µM, data from 100 µM experiment is shown) for 1 hour at 4 °C. After 1 hour, SYPRO orange dye was added and the protein was subjected to melting from 25 – 85 °C over 20 minutes.

3. AlphaScreen

In an AlphaScreen (PerkinElmer) assay, donor and acceptor beads are used to study the interaction between two proteins, molecules, or peptides. If binding occurs, the donor and acceptor beads are brought together in close proximity. Upon laser excitation at 680
nm, a photosensitizer inside the donor bead (phthalocyanine) converts the ambient oxygen to a more excited single state, which diffuses to produce a chemiluminescent reaction. Without an interaction between the two molecules, the single state oxygen molecules produced by the donor go undetected.

For our experiment, we used nickel chelate donor beads, which bind to EWS-FLI1, and anti-FITC acceptor beads, which bind to the FITC-E9R peptide (Figure 29A). As the experiment is executed in a 96-well plate, each experiment would allow for testing multiple small molecules at once. Each well contained 800 nm (3 µg) EWS-FLI1, allowing for multiple plates per protein purification. Although one initial experiment provided promising data regarding titration with increasing concentration of YK-4-279 (Figure 29B), many repeats did not indicate any consistency of results between multiple experiments (Figure 29C). Troubleshooting included changing incubation times after addition of donor beads, titrating amounts of DMSO, changing the order of addition of reagents, changing buffer from imidazole to urea, adding in increasing amounts of BSA to decrease the background, and degassing buffers. Experimenting with different buffers indicated that perhaps EWS-FLI1 itself was quenching the reaction (Figure 29D). 1 M urea buffer had more than a 10-fold higher fluorescence reading than the same 1 M urea buffer with EWS-FLI1. Meanwhile, the 1 M imidazole buffer had fluorescence levels equivalent to the 1 M urea buffer with EWS-FLI1. Furthermore, the buffer alone signal increased with decreasing levels of imidazole in the buffer, suggesting that perhaps the imidazole was also quenching the fluorescence signal. Peptide alone studies did not indicate the reduction in background seen in the protein plus peptide controls (data not shown).
Figure 29: AlphaScreen provides inconsistent results and may be quenched by protein.

(A) The AlphaScreen assay results in a chemiluminescent signal when acceptor and donor beads are brought together during binding. When donor
and acceptor beads are in close proximity due to biological partners, a fluorescent signal is produced. When there is no binding, singlet oxygen decays and no signal is produced. (B) Addition of increasing concentrations of YK-4-279 initially competitively inhibited the binding of EWS-FLI1 to RHA. EWS-FLI1, E9R, and small molecule were added to a 96-well plate and allowed to incubate for 30 minutes. Anti-FITC acceptor beads were added, followed by an hour-long incubation, and then addition of nickel chelate donor beads. The plate was read in an EnVision Multilabel Reader. (C) Results were no consistent, despite many attempts to troubleshoot. (D) Comparison of protein and peptide versus buffer alone in multiple buffers indicates that the recombinant EWS-FLI1 may be quenching the fluorescent signal.
Chapter IV: Discussion
A. Single enantiomer of YK-4-279 demonstrates specificity in targeting EWS-FLI1

EWS-FLI1 is tumor specific and is necessary for the survival of Ewing sarcoma cells, and is therefore an ideal molecular target. Previous studies have indicated that RHA enhances EWS-FLI1 driven oncogenesis, and interruption of this protein-protein complex by small molecular inhibitors may be a potential therapeutic target. As EWS-FLI1 is a disordered protein with unknown structure, standard target oriented synthesis was not an option for structure-based small molecule design. Instead, plates of small molecule compounds were screened using SPR, identifying several compounds that bound to EWS-FLI1 as potential disruptors of function. Following an initial single concentration binding screen of 300 compounds, 20 were analyzed for binding kinetics. From this, a lead compound, NSC635437, was selected based on cytotoxicity, drug-like properties, and binding affinity to EWS-FLI1, and fourteen derivative compounds were synthesized. The lead compound shows significant structural homology to a peptide that corresponds to amino acids 823 to 832 of RHA and is able to block EWS-FLI1 from binding to RHA. A derivative of NSC635437, YK-4-279, has shown significant toxicity against ESFT cells in cytotoxicity studies and blocks RHA from binding to EWS-FLI1.

We recognized that YK-4-279 has a chiral center and can be separated into its enantiomers. In binding assays and transcriptional assays, (S)-YK-4-279 is active and titrations indicate more potency when compared to the racemic, while (R)-YK-4-279 is inactive. The cellular response to (S)-YK-4-279 demonstrates that it is the active enantiomer in both apoptosis and growth assays in EWS-FLI1 containing cells. We show
that YK-4-279 and (S)-YK-4-279 are able to block the EWS-FLI1/RHA interaction, but (R)-YK-4-279 cannot. This work is significant for identifying a single enantiomer effect for the targeted disruption of small molecule modulators of protein-protein interactions with a disordered protein. It further supports the hypothesis that small molecule interactions with intrinsically disordered proteins can have significant specificity, despite relatively low binding affinity[90]. While the structure of the full-length EWS-FLI1 protein still eludes us due to its intrinsic disorder, the functional difference of the two enantiomers may help us learn more in future studies about the binding relationship between EWS-FLI1 and RHA, along with other protein-protein interactions.

In order to evaluate the potency of the enantiomers towards cells containing EWS-FLI1, we tested a panel of cells in cytotoxicity and caspase assays. Since a racemic small molecule consists of equal parts of the (S)- and (R)-enantiomers, one would expect a two-fold difference in potency between the active enantiomer and the racemate. Through a dose titration, we saw the expected two-fold difference in caspase activation between ESFT cells treated with YK-4-279 and (S)-YK-4-279 for 18 hours at low concentrations, while this difference saturates at the highest concentrations. In each of the ESFT cell lines studied for cytotoxicity, (S)-YK-4-279 had a lower IC₅₀ value than YK-4-279. The actual fold-differences between the racemic and the active enantiomer vary from 1.18-fold in RDES cells to more than a log-fold difference in TC71 cells. We tested cells containing all three major variant translocation types of Ewing sarcoma and this difference cannot be attributed to the specific translocation[21]. One explanation for this is the steep nature of the dose response curve, which causes large changes in IC₅₀ based on small shifts in
variability. Differences in the cellular uptake or metabolism of YK-4-279, independent of translocation type, may also account for the fold-difference between racemic and active enantiomer.

The cytotoxicity and caspase results indicate that (S)-YK-4-279 is the active enantiomer, and that it retains specificity to ESFT cells containing an ETS family transcription factor in comparison to non-Ewing cells. It is important to note that when testing the enantiomers, cell growth of both MDA-MB-231 and PANC1 cells, breast and pancreas respectively, were reduced by YK-4-279 and (S)-YK-4-279 (Figure 14C). However, previous testing indicated IC₅₀ values of >20 µM (Figure 7A). As the two experiments were performed several years apart, the cell lines were received from different sources and the latter clones exhibited increased sensitivity. Despite the increased sensitivity in cytotoxicity studies, the IC₅₀ values for MDA-MB-231 and PANC1 cells were lower for the racemic than the active enantiomer and there was no statistical significance between treatments with the racemic and (S)-YK-4-279. When tested for caspase-3 activation, neither of the cell lines exhibited more than 2-fold caspase activation, in comparison to ESFT cells, which averaged 8.8-fold for YK-4-279 and 10.3-fold for (S)-YK-4-279 treatment across the panel of six cell lines. Although YK-4-279 and the enantiomer may have increased toxicity in MDA-MB-231 and PANC1 cells, the mechanism of action may be different than in ESFT cell lines. Recent studies have identified caspase-independent cell death factors induced by anticancer drugs[149], including apoptosis-inducing factor (AIF)[150], endonuclease G[151], and HtrA2[152], which may explain the increase in cytotoxicity without caspase activation in these two non-ESFT cell lines.
Further studies focusing on caspase-independent cell death may explain why these cells exhibit cytotoxicity toward YK-4-279 but do not undergo apoptosis and may suggest other potential mechanisms of activity.

Initial in vivo studies indicated that YK-4-279 inhibits the growth of ESFT xenograft. SCID mice were orthotopically injected with TC71 cells and animals treated with small molecule showed a significant overall tumor reduction compared to vehicle-treated mice[77]. Studying the effect of the enantiomers in A4573 xenografts indicated a statistical significance in tumor growth and survival between YK-4-279 and control, but not with either enantiomer (Figure 14). Given the improvement of specificity and potency of (S)-YK-4-279 in every other assay, we theorize that the underwhelming results are due to problems with formulation of YK-4-279.

A drug molecule must be water soluble to reach the cell membrane, but must also be hydrophobic to cross through the membrane[153]. YK-4-279 is a Biopharmaceuticals Classification System (BCS) class II drug, characterized by low solubility but reasonable membrane permeability[154]. Over 40% of new drug candidates are not water soluble, which can lead to ineffective absorption and poor pharmacokinetics[155]. Compounds such as YK-4-279, which have been designed and optimized based on receptor binding potency, are usually water insoluble[156]. However, this does not exclude successful development of a pharmaceutical, as more than one-third of drugs available in the United States are either poorly water-soluble or completely water-insoluble. Instead, the development of solubilization systems is necessary to help deliver these water-insoluble drugs to their targets. Current techniques for the delivery of insoluble drugs include
modifications to particle size, crystal structure, polymorphs, solubilization through surfactants or cyclodextrins, or drug dispersion in carriers[157]. Future experiments are necessary to determine the appropriate formulation and delivery of YK-4-279 to ensure that a maximal amount of the small molecule is reaching the tumor.

In further preparation for clinical trials, we evaluated the potential racemization of YK-4-279 in ESFT xenografts [144]. YK-4-279 and its enantiomers were administrated by intravenous (IV) injection or oral gavage to Sprague-Dawley rats. As expected, since the racemic contains a 1:1 mixture of each enantiomer, rats treated with each enantiomer individually had 1.8- to 2.3-fold higher concentrations of the specific enantiomer in plasma than rats treated with the racemic. We do see increased racemization in the urine, but not in the plasma, following gavage administration. This may be attributed to low oral bioavailability and slower elimination of YK-4-279 compared to IV administration. Importantly, no enantioconversion was seen in the tumors following implantation of xenograft tumors in SCID mice. These results suggest that the chirality of YK-4-279 does not change in the plasma or tumor tissues, but there may be a tissue-specific enzyme that isomerizes YK-4-279 in the urine. Additional experiments testing biospecificity are necessary to further study the pharmacokinetics of (S)-YK-4-279.

While further pharmacologic studies are necessary to advance YK-4-279 towards the clinic, it is also important to consider implications in other diseases. ETS rearrangements are present in 40-70% of all prostate cancers, including those that are the most clinically aggressive[158-160]. These aggressive prostate cancers fuse the promoter of TMPRSS2, an androgen responsive gene, to an ETS family transcription factor, such as
ETV1 or ERG[161], both of which have also been identified in joining to EWS in ESFT to form an oncogenic fusion protein[22-24]. Recent experiments from our group have indicated that YK-4-279 is also able to inhibit ERG and ETV1 fusion-positive prostate cancer cell lines[162], and the activity and function of the enantiomers should be further investigated in prostate cancer. Our data support the ability to disrupt a critical oncogenic protein-protein interaction with YK-4-279 and the importance of enantiospecificity in disrupting EWS-FLI1 from RHA. A small molecule targeted to the ETS family of transcription factors with minimal side effects would be an important development in treatment for patients with ESFT and other ETS-family oncogenic fusion proteins.

**B. EWS-FLI1 binds to RHA but region of interaction remains unknown.**

Given the challenges of drug design targeted to EWS-FLI1, we hope that a better understanding of the structure of EWS-FLI1 and its binding interaction with RHA will allow us to better design small molecules to block the protein-protein interaction. Initially, three fragments of EWS-FLI1 were immunoprecipitated with endogenous RHA to identify the site of the interaction. Multiple repetitions did not indicate a successful co-precipitation of RHA by any of the fragments. Successful purification of the RHA 630-1020 fragment and improvements in EWS-FLI1 purification permitted crosslinking studies between the two proteins. The resulting crosslinked region was analyzed using mass spectrometry. However, due to the scrambled order of individual amino acids in the resulting mass spectrometry identified peptide fragments, we were unable to identify which region of the proteins is actually crosslinked.
While we were successfully able to crosslink recombinant EWS-FLI1 and RHA, we were unable to interpret the resulting peptide data. Other options exist with published methods for determining protein sequences involved in crosslinking[163]. First, a photocrosslinking technique site-specifically incorporates p-benzoyl-L-phenylalaine (pBpa) into an E. coli-expressed protein[164-166]. Amber mutants of the gene of interest are generated, with the mutation at the sites for photo-crosslinker incorporation. An orthogonal aminoacyl tRNA synthetase-tRNA pair incorporates the pBpa at the position encoded by the amber codon. To photo-crosslink the protein to its binding partner, the proteins are exposed to UV light. However, generation of the mutant gene requires knowledge of where crosslinking occurs within the protein. While we know the region of RHA that binds to EWS-FLI1, we do not know the binding region on EWS-FLI1, thus requiring the generation of many mutants to determine the proper area.

Digestion with CNBr is a better option for proteins with an unknown region of binding[167]. Similar to the experiment demonstrated in Figure 22, proteins are crosslinked with BS3. However, instead of digesting the excised crosslinked proteins with trypsin, CNBr is used for digestion. As CNBr cuts after Met, larger fractions result. There are 9 Met in EWS-FLI1 and 13 in the 630-1020 fragment of RHA. The digested product can then be run on a gel to identify a band present in the sample that has been crosslinked, but not present in the uncrosslinked control samples. This band is then digested with typical proteases for mass spectrometry (chymotrypsin for EWS-FLI1, as trypsin does not cut in the EWS region), similar to the method we attempted. However, pre-digestion with CNBr
narrowst down the location of peptide fragments, as the size of the excised crosslinked band is the result of a fragment of EWS-FLI1 plus a fragment from the RHA.

There are four main techniques for studying the structure of a disordered protein: NMR spectroscopy, x-ray crystallography, circular dichroism, and protease sensitivity[168]. NMR is able to detect any molecular motion, along with any secondary or tertiary structure. Although NMR could provide a great deal of information about EWS-FLI1 and a possible conformational change following binding to EWS-FLI1, we are unable to concentrate our protein to levels required for NMR. At least 1 mM protein is necessary, and in a buffer containing less than 250 mM salt, both of which are currently not obtainable. X-ray crystallography led to the initial discovery of disordered proteins, as regions of disorder fail to scatter x-rays and lead to missing electron density. However, EWS-FLI1 is too fluid in solution for formation of crystals. Collaborative efforts with researchers at Los Alamos National Laboratory proved futile in the attempt to grow an EWS-FLI1 crystal, despite the addition of small molecule, peptides, and other proteins in an attempt to stabilize or co-crystallize the protein. In circular dichroism, a disordered protein is characterized by low intensity in the near-UV CD spectra. While EWS-FLI1 was previously studied by circular dichroism[117], we are currently unable to repeat the purification with EDTA elution, and have had very limited success with dialysis of the imidazole, preventing repeat CD studies.

The fourth main technique for studying the disorder of proteins is protease sensitivity, as ordered proteins are more resistant to protease digestion than disordered proteins. Through native state proteolysis experiments, we are able to experimentally show that EWS-FLI1 is a disordered protein and the EWS region and C-terminal domain of
FLI1 exhibit the most disorder, while the helix-turn-helix ets binding domain is highly ordered. Current experimental plans include further studying the protease sensitivity of EWS-FLI1. BSA, used as a positive control, is an extremely ordered protein. Comparing EWS-FLI1 to other proteins with varying degrees of disorder, such as apomyoglobin which contains a disordered F helix or the Lac repressor which has disorder in its DNA unbound state, will establish the degree of disorder relative to a known disordered protein [169].

We plan to repeat the experiment and send digested bands from the chymotrypsin digestion for mass spectrometry analysis. There are fewer cleavage sites with chymotrypsin compared to proteinase K, while trypsin is not an ideal protease for digestion as there are no cleavage sites in the EWS region. Furthermore, we hypothesize that we may be able to identify the binding site for YK-4-279 if the binding occurs at a cleavage site. The small molecule or peptide may protect the site from digestion, showing up as a missing band when compared to a digestion without small molecule. While we initially looked for a shift in bands or a band that was more prominent in the YK-4-279-treated samples (Figure 26), we now realize it would be more useful to look for a band that is present in the control lane, but not the treated lane, and send the control band for mass spectrometry analysis. If the small molecule is protecting this region of protein, that band may not appear until a later time point of digestion.

Similar to native state proteolysis, hydrogen/deuterium (H/D) exchange can monitor conformational changes in proteins after binding to a small molecule. H/D exchange replaces a covalently bonded hydrogen atom with a deuterium atom. If a structural change occurs, there will be a detectable change to the H/D exchange [170].
Individual peptides are analyzed for overall deutroneration of each fragment and can indicate regions of order versus disorder. H/D exchange could be used to confirm a suspected region of YK-4-279 binding discovered through protease sensitivity.

New purification methods have been developed in the Toretsky laboratory for recombinant EWS-FLI1, including elution with arginine instead of imidazole. While optimizations and dialysis experiments continue, the possibility of an imidazole-free, low salt preparation of EWS-FLI1 will hopefully allow us to utilize some previously unsuccessful assays. Circular dichroism may be an option, along with NMR if the arginine preparation can be concentrated. Both experiments would allow us to study structural changes to the protein upon binding, with full length protein, protein fragments, and wild-type EWS and FLI1, which may allow us to narrow down the region of binding for YK-4-279 and RHA.

C. Small molecule assay for structure-function relationship.

While attempts at high-throughput fluorescence polarization, protein melting, and AlphaScreen were not successful, other experiments presented in this dissertation offer a higher-throughput method to quantify activity of a small molecule outside of cellular cytotoxicity assays. An ELISA has been established which shows a clear titration of binding signal with the addition of increasing amounts of small molecule (Figure 11A). While requiring less protein than fluorescence polarization assays, the ELISA is a 3 day assay, and therefore not high-throughput. However, since we are able to purify recombinant wild-
type EWS and FLI1, the ELISA assay may provide useful data on small molecules that are able to bind to either partner protein, in addition to EWS-FLI1.

Initial luciferase assays tested small molecule at 3 µM, 10 µM, and 30 µM, and while we were able to see a clear enantiomeric difference between (S)-YK-4-279 and (R)-YK-4-279, there was no dose-dependent titration seen with the active enantiomer or racemic. However, as we continued to decrease the concentration of the small molecule to 0.3 µM, a dose-dependent effect was clearly visible (Figure 12), allowing us to calculate an IC$_{50}$ of the luciferase inhibition. This assay has been optimized for use in a 96-well plate, and allows for testing multiple compounds per plate to determine the effect on transcriptional activity of EWS-FLI1. While cells are treated with the small molecule for 16 hours, cells can be transfected the afternoon of Day 1 and results are obtained by the afternoon of Day 2, and multiple plates can be assayed at a time.

Results from the luciferase screen, combined with IC$_{50}$ cytotoxicity data, can be imported into a Free-Wilson QSAR model. The structures of current small molecule derivatives have already been converted into a numerical format, allowing for a mathematical comparison between structures used to predict the biological response of other chemical structures. This data may help to predict the binding of other analogues with similar chemical structure and direct the synthesis of new analogues.

**D. Proposed models of interaction between EWS-FLI1 and RHA**

Experimentally, we have been unable thus far to determine the region of EWS-FLI1 to which RHA is binding. We can, however, propose three models that may describe the
interaction between the two proteins. The simplest hypothesis is that the binding location may be on the ets domain of FLI1, with a structured protein binding to a region of structure in another protein. An alternative model suggests that RHA binds to the disordered EWS region of the fusion protein. As a third option, the protein may loosely fold in an overall globular structure, and RHA binding may require bringing together the C-terminal and N-terminal regions of the protein.

Since only 30-50% of proteins contain a region of disorder of at least 30 amino acids[86], half of all proteins are therefore estimated to be fully ordered. Many of these proteins bind to other ordered proteins, and RHA, an ordered protein, is already known to bind to p65, another ordered protein[171]. The first attempt at the GST-fragment pulldown with RHA indicated that RHA might be binding with the FLI1 region, visible in Figure 16B, top blot. We repeated the experiment several more times but were unable to achieve the same results. Because we were using crude, unpurified lysates for the experiment, it is possible that differences in contaminants of the FLI1 may account for an inability to achieve the same results. It is also important to note that the ordered region of FLI1 may not fold the same way when fragmented as it does in full-length EWS-FLI1, thereby preventing an interaction between the proteins. Our laboratory previously studied the interaction between RHA and full length FLI1. MOLT4 cells, which overexpress full length FLI1, were able to pull down RHA in a co-immunoprecipitation assay. Furthermore, YK-4-279 does target other ets binding domains, notably the ERG fusions in prostate cancer. Although the mechanism of action is not known, and RHA does not play a role, this suggests that YK-4-279 is binding to the ets DNA binding domain. If YK-4-279 serves to block the
binding site between EWS-FLI1 and RHA, the binding site likely involves residues from the ets domain. While not conclusive evidence, these data indicate that one method of interaction may involve binding in the FLI1 region of EWS-FLI1, possibly in the ordered ets domain.

Inversely, EWS-FLI1 may be interacting with RHA in the EWS region of disorder. As previously mentioned, many disordered proteins undergo a disorder-to-order transition upon binding with another protein. The flexibility of disordered proteins allows for binding to many partners in different confirmations, and EWS-FLI1 is already known to interact with many partner proteins, perhaps even serving as a “bottleneck” in the signaling pathway[172]. Binding within a disordered region provides many advantages, including very specific binding with a lower binding strength, increased speed of binding, and binding to multiple partners[173], all of which would be advantageous for EWS-FLI1’s binding to RHA. While structured proteins tend to bind via polar-polar surface interactions, disordered regions favor hydrophobic contacts. Although the 630-1020 fragment of RHA contains both hydrophilic and hydrophobic regions, the PPPLDAVIEA peptide, which binds to EWS-FLI1, is mostly hydrophobic, according to the Kyte-Doolittle Scale of hydrophobicity.

A third proposed mechanism of interaction between EWS-FLI1 involves a binding site that requires interaction with a region from both EWS and FLI1. While mostly disordered, EWS-FLI1 does contain regions of order, which may loosely bring together the C-terminal and N-terminal domains of the protein. This may mean that RHA is interacting with residues from each half of the fusion protein. This model does corroborate with the
data obtained from the GST-tagged fragment pulldown, as full-length EWS-FLI1 is able to pull down RHA, while the EWS, FLI1, or fusion fragments are not. Furthermore, a small molecule microarray, completed at the Broad Institute, tested 10,000 small molecules for their ability to bind to EWS, FLI1, or EWS-FLI1. Of these small molecules, 12 were identified which bound solely to EWS-FLI1, and not to either EWS or FLI1. These 12 molecules, unable to bind to either partner protein, suggest that EWS-FLI1 folds in such a way that presents binding sites unique to only the fusion protein, perhaps a result of folding that brings together both ends of the protein.

ANCHOR, a protein-binding predictor that looks at properties of disordered proteins, predicts binding sites based on the disorder predictor, IUPred. IUPred characterizes proteins based on amino acid composition, as disorder-inducing residues do not allow for well-defined structural characteristics[174, 175]. IUPred recognizes the regions that are most likely to undergo the disorder-to-order transition, as it estimates the energy of each residue based on both the individual amino acid and the composition of its neighborhood. ANCHOR identifies regions that cannot fold on their own, but are likely to gain stabilizing energy through the interaction with an ordered protein[176]. The ANCHOR binding prediction looks for residues that are unable to form favorable contacts within a region to induce folding, but can form interactions with globular proteins upon binding in the form of entropy gain. Interestingly, importing the sequence of EWS-FLI1 into ANCHOR predicts two areas that are most likely to interact with a structured protein: one in an area of disorder in EWS, and one in the helix-turn-helix domain of FLI1 (Figure 30).
Figure 30: ANCHOR predicts two sites of interaction on EWS-FLI1. The ANCHOR algorithm is a protein binding predictor for disordered proteins, looking for the most likely sites of interaction with a structured globular protein. For EWS-FLI1, the most likely sites of protein interaction are highlighted with an arrow: one in the disordered region of EWS, and one in the helix-turn-helix structured ets domain in FLI1.

Any of the three models appear to be feasible locations for binding between the two protein partners. Techniques already studied in the lab may narrow down the binding site. We are able to purify full-length EWS and FLI1 in the lab, and could clone the EWS-FLI1 fractions into His-tagged vectors, similar to the process taken to purify the 630-1020 fragment of RHA. Using ELISA, it may be possible to narrow down the binding location by experimentally showing that RHA binds to full length EWS, full length FLI1, or one of the fragments. Continued optimization of the AlphaScreen assay would allow for the same testing of the fragments and the wild-type partner proteins. If the third model were correct, neither of these options would provide positive results, as residues from both EWS and FLI1 are required to bind to RHA. Instead, site-directed mutagenesis may provide an answer about a binding location. While not practical for the entire EWS-FLI1 fraction,
mutating select amino acids on either the C-terminus or N-terminus end of the fusion protein may prevent binding to RHA.
Chapter VI: References


117


