DESIGN AND SYNTHESIS OF DNA MINOR GROOVE METHYLATING COMPOUNDS THAT TARGET ESTROGEN RECEPTOR POSITIVE CELLS

Tera L. Lynch

A Thesis Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
Of the Requirements for the Degree of
Master of Science

Department of Chemistry

University of North Carolina Wilmington

2006

Approved by

Advisory Committee

________________________ ________________________

____
Chair

Accepted By

____
Dean, Graduate School

This thesis has been prepared in the style and format

Consistent with the journal

Journal of Organic Chemistry

TABLE OF CONTENTS

ABSTRACT....................iv

ACKNOWLEDGEMENTS....................v

DEDICATION....................vi

LIST OF TABLES....................vii

LIST OF FIGURES....................viii

INTRODUCTION....................1

BACKGROUND....................3

SYNTHETIC STRATEGY....................13

EXPERIMENTAL....................24

RESULTS AND DISCUSSION....................36

CONCLUSION....................48

REFERENCES....................49

APPENDIX A....................51

APPENDIX B....................53

ABSTRACT

The goal of this project is to design and synthesize compounds that can potentially target breast cancer cells selectively, and can cause specific DNA-damage in these tumor cells. The molecules that will be synthesized will consist of 3 components: (a) a component known to selectively methylate the N3-position of adenine in A/T-rich regions of DNA, (b) a component known to target breast cancer cells due to its specific binding to the estrogen receptors (ER) that are over expressed in early stage breast cancer cells, and, (c) a variable linker component, that connects the above two functional components in such a way that the desired properties (DNA-methylation and cell-targeting) are optimized. The specific kind of DNA-damage, the N3-methyladenine adduct, is one that has been shown to cause only cell-death, and not cause mutations. The cell-targeting component chosen is estradiol, which has been often used to deliver agents to breast cancer cells. This thesis discusses the design of these compounds, and the synthetic strategy that has been developed for the preparation of these compounds. The progress towards the synthesis of one of these compounds is also described. Understanding how to target breast cancer cells preferentially, and how to cause only the kind of DNA-damage that results in cell-death, will aid in the design of new drugs that can destroy breast tumors while minimizing unwanted side-effects, and eliminating the incidence of secondary cancers.

ACKNOWLEDGEMENTS

I would like to thank Dr. Sridhar Varadarajan for being interested in this research that could some day be life changing. Also for his guidance and teaching me everything I know about organic chemistry. My thanks also go out to Dr. Tyrell and Dr. Halkides for their help and for being on my committee. I would also like to send a thank you to Dr. Sayed for his help.

Last but not least, I would like to give a great big thank you to my parents who have been there for me over the years. Without their love and support I wouldn't be where I'm at today.

DEDICATION

I would like to dedicate my thesis to my grandmother, Theresa. I know she is looking down on me with the biggest smile on her face. When I was younger she wasn't sure if I would even go to college and now here I am with a Degree in Master of Science. Thank you for being there for me when I was growing up and the influence you have had on my life. I would also like to dedicate my thesis to my family members who have passed on as a result of cancer my Grandpa Croyle, Grandpa Lynch, Uncle Jim, Uncle Roy, Uncle Al, Aunt Franny, Aunt Evelyn, and Aunt Mona.

LIST OF TABLES AND SCHEMES

Table Page

1 Percent yields for the compounds in the synthesis of the DNA recognition component. 40

Scheme Page

1 Outline for the synthetic strategy of the DNA recognition segment. 16

2 Outline for the synthesis of the cell-targeting segment 19

3 Strategy for assembling the components. 22

4 Introduction of the reactive methylsulfonate ester into the molecule..........................23

5 Synthesis of the DNA recognition component. 38

6 Synthesis of the cell-targeting component. 42

7 Synthetic strategy for coupling the DNA recognition and cell targeting components. 46

8 The strategy for conversion of the nitro compound into the methylsulfonate ester 46

LIST OF FIGURES

Figure Page

1 Structure of DNA showing the major and minor grooves, and base-pairing..............4

2 Structure and pairing of the bases: adenine with thymine and guanine with cytosine. 6

3 Structure of Me-Lex, the compound that produces N3-methyladenine adducts exclusively 8

4 (a) Binding of Me-Lex in the minor groove of DNA at A/T rich regions, and (b) transfer of the reactive methyl group to N3-adenine..8

5 Structure of estradiol, which selectively binds to the estrogen receptor (ER), which is over expressed in breast cancer cells 10

6 Structure of a large alkylating compound (nitrogen mustard) attached to the 7α position of estradiol..10

7 Overall design of compounds for this project includes three segments: a DNA alkylating segment, a cell-targeting segment, and a linker..12

8 Synthetic strategy used to prepare compounds that can target breast cancer cells and methylate N3-adenine in those cells...14

9 Structure of the cell-targeting segment that will be synthesized. 19

10 Examples of potential linkers with variations in length and composition................22

11 Structure of the DNA recognition with the linker attached and cell targeting components 38

12 Structure of the DNA recognition compound that was made...................................38

INTRODUCTION

Some cancer chemotherapy drugs such as mechlorethamine, busulfan, and dacarbazine are DNA-alkylating compounds[1]. Such alkylating compounds usually damage DNA extensively at multiple sites leading to the death of the cells. Cancer chemotherapy drugs attack all cells in the body, both normal and cancerous, and damage DNA. Most normal cells survive the DNA damage that occurs as a result of the cancer chemotherapy drugs because repair enzymes have time to mend the damage on DNA. On the other hand, in cancerous cells, the repair enzymes are either not functioning well, or they are unable to fix the damage before DNA replication is initiated in these rapidly dividing cells. This DNA damage that persists during cell-division in these cells triggers signals that result in the death of these cells. Normal cells that are rapidly dividing, for example, hair cells and cells that line the stomach, are also killed by this process, since in these cells too, the DNA-damage is not repaired before cell-division is initiated. This is why people undergoing cancer chemotherapy treatment with alkylating drugs have side effects such as hair loss and gastric irritation[1].

Damage at various sites on DNA can result in different biological outcomes depending upon the site of alkylation. Alkylation at some sites cause the death of the cell, which is the desired outcome in a cancer drug. Alkylation at other sites can result in mutations, which are believed to be a cause of secondary cancers. Thus, these secondary cancers are caused by the DNA-alkylating drug used to treat the primary cancer, and the drugs used to treat the primary cancer cannot be used to treat the new cancer. DNA alkylation at certain sites can cause both cell-death and mutations, while damage at some sites appear to have no biological consequence. Alkylation at sites that do not cause cell-toxicity are a waste of the chemotherapy drug and will

necessitate the use of higher doses for the treatment to be effective, which also leads to an increase in unwanted side effects.

An ideal cancer drug would be one that targets only cancerous cells, and one that causes damage only at those sites on the DNA in these cells that results in cell-death. Such ideal drugs would exhibit improved efficiency (therefore requiring lower dosage), minimize side-effects, and eliminate the incidence of secondary cancers.

This thesis describes our efforts to design and make compounds that address some of the issues discussed above. We are attempting to make compounds that could potentially target breast cancer cells, and cause only the specific kind of DNA-damage in those cells that have been shown, in literature, to cause only cell toxicity (without causing mutations). The results from this study can aid in the design of improved cancer chemotherapy drugs.

BACKGROUND AND SIGNIFICANCE

Background:

The goal of this research is to make a compound that will target a specific kind of cell (breast cancer) and cause specific damage to DNA in that cell that will result only in cell toxicity and not in mutations. The design of compounds that can target and react with specific sites on DNA requires a good understanding of DNA structure and accessibility.

DNA (deoxyribose nucleic acid) is a double helix that is made up of a sugar-phosphate backbone on the outside and the interior is a helix of nitrogenous base pairs, which are holding the two strands together through hydrogen bonding (Figure 1). There are four bases adenine (A), thymine (T), guanine (G), and cytosine (C), each of which can pair only to a specific base, adenine pairs with thymine and guanine pairs with cytosine. The base-pairing of the two strands of DNA results in the formation of two grooves of DNA, the major groove and the minor groove (see Figure 1). The major groove is wide and easy to access, and many proteins and enzymes interact with the major groove. The minor groove is deep, narrow and difficult to access.

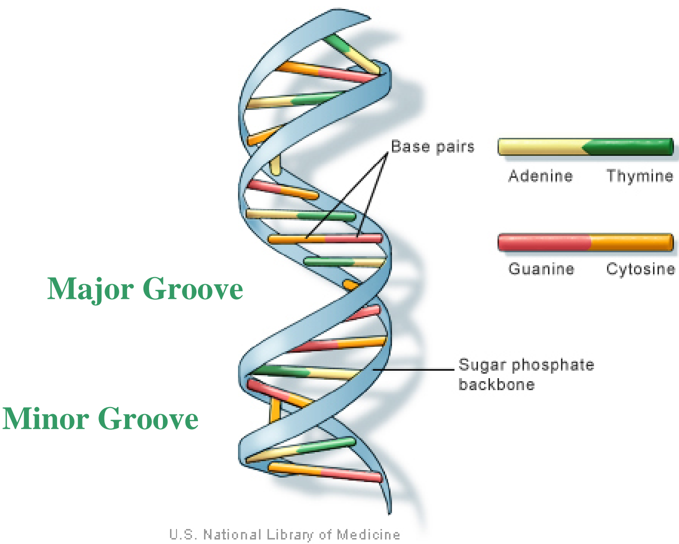


Figure 1. Structure of DNA showing the major and minor grooves, and base-pairing.

Figure 2 shows the top view of an AT and GC base pair, and the orientation of various sites that results from the base-pairing. In the AT base pair, the sites that lie in the major groove are N7 and N^6 of adenine and O^4 thymine . The N3 of adenine and O^2 of thymine lie in the minor groove. In the GC base pair, the sites that lie in the major groove are N7 and O^6 of guanine and N^4 of cytosine. The sites that lie in the minor groove for the GC base pair are the N3 and N^2 of guanine and O^2 of cytosine.

The arrows shown in Figure 2 point to sites that can get damaged by DNA-alkylating agents. Damage at different sites can lead to different biological consequences. Damage at certain sites can lead to cell-death, while damage to other sites can lead to mutations. N7 guanine is easily accessible and the most reactive site on the DNA bases. The N7-methylguanine adduct has been shown to be non-cytotoxic and non-mutagenic[2]. The O^6-methylguanine adduct is both cytotoxic and mutagenic. Other sites that are mutagenic[3] and are involved in base pair hydrogen bonding are O sites and N2 guanine. When DNA-methylation occurs at N3 adenine, a cytotoxic nonmutagenic N3-methyladenine adduct is formed. Therefore, compounds that can form N3-methyl adenine adducts would be a good choice of a DNA – damaging agent that can be used to cause only cell-death, and not lead to mutations which can cause secondary cancer.

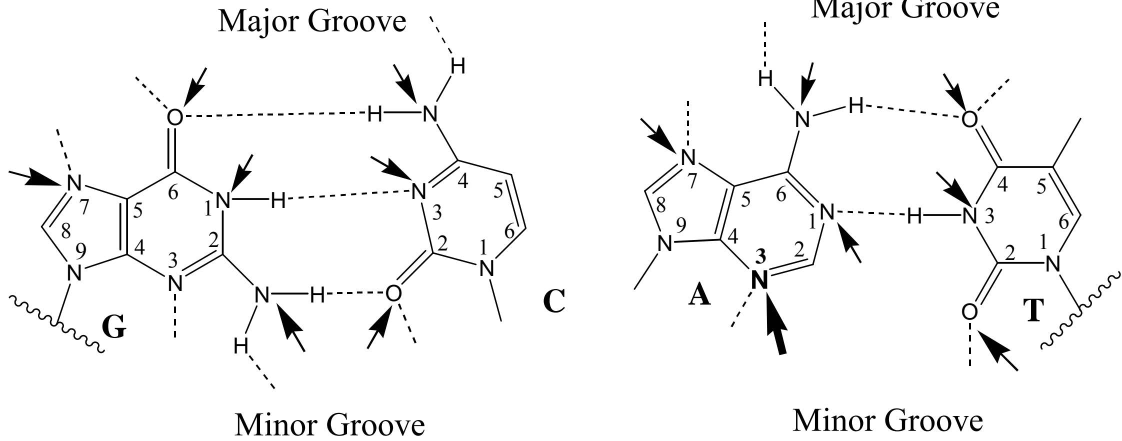


Figure 2. Structure and pairing of the bases: adenine with thymine and guanine with cytosine. The arrows indicate sites that can be alkylated by DNA-alkylating agents. The N3-adenine site, which when methylated, results in cytotoxicity without leading to mutations, is highlighted in bold.

Targeting N3 adenine:

A compound that selectively produces only N3-methyladenine adducts, a methyl sulfonate ester tethered to *N*-methylpyrrolecarboxamide dipeptide (Me-Lex) has been described in literature, and is shown in Figure 3. This compound has been shown to methylate adenines in the minor groove N3-site only in A/T rich regions[2,4,5]. Over 95% of the DNA-adducts formed by Me-Lex is N3-methyladenine. Me-Lex thus has the desired ability to produce exclusively cytotoxic, non-mutagenic adducts.

The binding of Me-Lex in the minor groove of DNA at A/T rich regions is shown in Figure 4a and involves van der Waals forces, hydrogen bonding and electrostatics[2]. The reason that this molecule binds selectively at A/T rich regions is because it is excluded from the minor groove at regions containing guanines due to the N^2 of guanine that protrudes into the minor groove and sterically hinders such binding[4]. Once Me-Lex binds in the minor groove at A/T rich regions, the reactive methyl group is transferred to the most nucleophilic site at A/T these regions, which is the N3 of adenine (Figure 4b). After the methyl group is transferred the sulfonate ester is converted into a negatively charged sulfonate anion, which is repelled from the negatively charged DNA-backbone[2]. N3-methyl adenine adducts formed by Me-Lex have been shown to be highly cytotoxic and nonmutagenic[4]. Thus Me-Lex is a compound which, if targeted to breast cancer cells, can be used to destroy these cells without causing mutations.

Figure 3. Structure of Me-Lex, the compound that produces N3-methyladenine adducts exclusively.

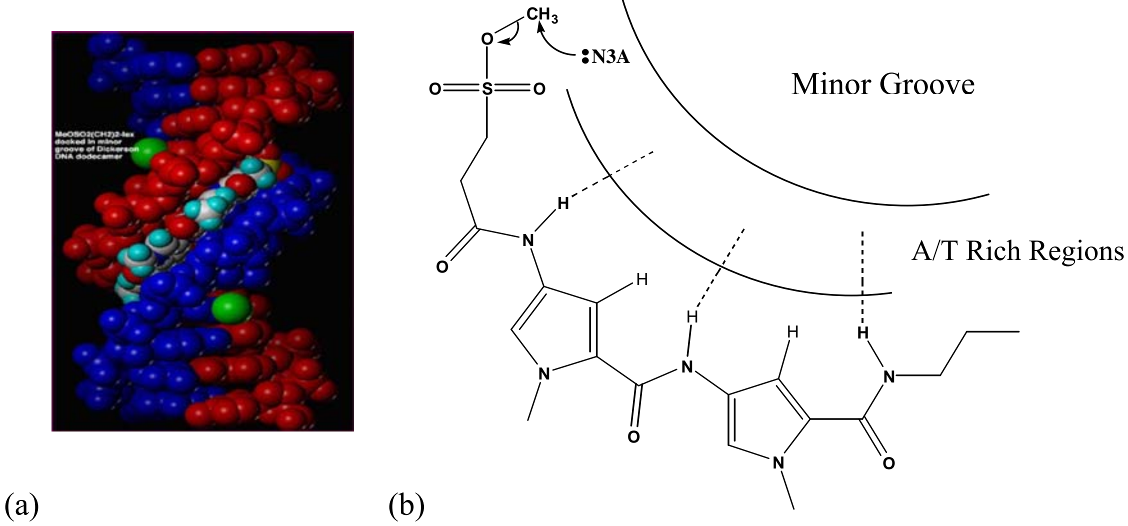


Figure 4. (a) Binding of Me-Lex in the minor groove of DNA at A/T rich regions, and (b) transfer of the reactive methyl group to N3-adenine.

Targeting breast cells:

Me-Lex can be targeted to breast cancer cells if it can be attached it to ligands, which are specific to breast cancer cells. Estrogen receptors (ER) are over-expressed on breast cancer cells and expressed at low levels in normal breast cells[6], and so ligands that target the ER can be used to target Me-Lex to these cells. Thus, by using a compound like estradiol (Figure 5), which selectively binds to the ER, to deliver Me-Lex to breast cancer cells, one could preferentially target the alkylating agent to these cells.

The ER has a selective ability to detect estradiol and a high affinity to bind to estradiol[7]. The ER's ability to recognize estradiol is due to specific hydrogen bonding interaction and estradiol's non-polar character[8]. The binding of estradiol to the ER causes a conformational change, which results in translocation of the estradiol-ER complex into the nucleus of the cell[7,9]. Thus estradiol is an ideal candidate to deliver Me-Lex to the DNA in ER-positive breast cancer cells.

It has been shown that modifications at the 7α, 11α, 16α, and 17β sites (Figure 5) do not radically affect the ability of estradiol to bind to the ER[7,10,11]. Radiolabels and DNA alkylators have been attached to these sites and have been taken into the nucleus by the ER. For example, a nitrogen mustard connected to estradiol at the 7α-position (Figure 6) has been shown to alkylate DNA in breast cancer cells at N7 of guanines[10]. Therefore, this 7α-site is a good site to attach Me-Lex, a DNA-methylating agent, to estradiol, in order to deliver it to the DNA in breast cancer cells.

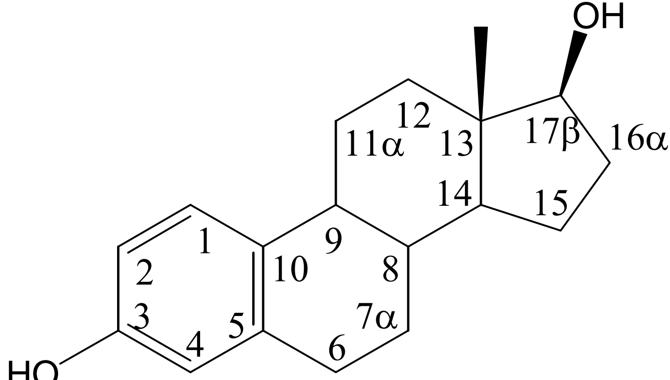


Figure 5. Structure of estradiol, which selectively binds to the estrogen receptor (ER), which is over expressed in breast cancer cells. The sites that can be modified on estradiol, without interfering with its binding to the ER, are 7α, 11α, 16α, and 17β.

Figure 6. Structure of a large alkylating compound (nitrogen mustard) attached to the 7α position of estradiol that is know to alkylate DNA in ER-positive breast cancer cells.

Significance:

The goal of this research is to design and synthesize compounds that will methylate N3-adenine at A/T rich regions in the minor groove of DNA in breast cancer cells to form cytotoxic nonmutagenic adducts. This will be achieved by connecting two compounds: Me-Lex, that produces cytotoxic N3 methyladenine lesions and estradiol, which targets breast cancer cells. Thus the overall design of the compounds is shown in Figure 7 and contains three segments: a DNA alkylating segment (Me-Lex), a cell-targeting segment (estradiol), and a linker segment. The linker is very important to the design of the molecules and the length and composition of the linker will be varied in order to achieve optimum cell-targeting and DNA-methylating properties. Also, the linker can be modified to vary the solubility of the compound in aqueous medium.

The purpose of making these compounds is to gain an understanding of the factors that govern the ability to deliver a selective DNA-alkylating agent to a specific kind of cell. Elucidation of these will pave the way for the design of new and improved cancer chemotherapy drugs that target only cancerous cells and minimize side effects.

DNA Alkylating Segment

Linker

Cell Targeting Segment

Figure 7. Overall design of compounds for this project includes three segments: a DNA alkylating segment, a cell-targeting segment, and a linker, which connects the two functional segments of the molecule.

SYNTHETIC STRATEGY

The design of the molecules for this project involve two main components: one is the DNA alkylating unit, which will alkylate DNA in a sequence specific minor at A/T rich regions, and the other is the cell-targeting unit that will target the DNA alkylating unit to breast cancer cells. These two functional components have to be connected in the best possible way so that, after the connection, the DNA alkylating segment will still achieve its specific DNA alkylation, and the cell-targeting segment will also maintain its ability to target breast cancer cells. Many different linker components, which will link the above two components, will have to be tested, in order to determine which of the linkers will result in the optimum properties. Therefore a good synthetic strategy would be one that will make it easy to introduce many different linkers efficiently.

Different linkers can be incorporated efficiently into the molecules if the DNA-alkylating segment and the cell-targeting segment are first synthesized separately as appropriate intermediates, and then each of these components can be assembled together with the appropriately functionalized linker units at a late stage in the overall synthesis. Figure 8 outlines this synthetic strategy that can be used to achieve these goals.

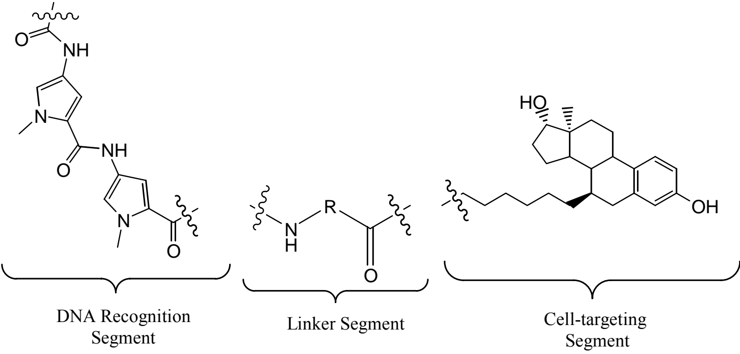


Figure 8. Synthetic strategy used to prepare compounds that can target breast cancer cells and methylate N3-adenine in those cells.

Strategy for the synthesis of the DNA recognition component:

The component that is designed to achieve the selective DNA methylation has two functional units, one that recognizes the DNA minor groove at A/T rich regions, and the other is the methylating functionality, the methyl sulfonate. Due to the reactive nature of the methyl sulfonate, it will have to be introduced at the very end of the overall synthesis, after all of the different units have been assembled together.

The over all synthetic strategy for the DNA recognition segment is outlined in Scheme 1. The commercially available N-methyl pyrrole can be converted into compound **1** by acylation with trichloroacetyl chloride. This trichloroacetyl group is stable to aqueous hydrolysis, which eliminates the need for dry solvents in subsequent reactions. Compound **1** can be nitrated at the 4-position efficiently by using nitric acid and acetic anhydride to give compound **2**. The linker can be added to compound **2** by using ethyl β-alanine hydrochloride and triethylamine to give compound **3**. Then this nitro compound **3** can be coupled to compound **2** in order to give the dipeptide ester **4**. The ester group in **4** can be hydrolyzed to obtain a carboxylic acid, which can then be coupled to the appropriate cell-targeting component functionalized as an amine.

Scheme 1. Outline for the synthetic strategy of the DNA recognition segment.

Strategy for the synthesis of the cell-targeting component:

The component that will enable the compounds to be targeted to breast cancer cells is estradiol because of its specific binding to the ER, which is over expressed in early stage breast cancer cells. Therefore, the site on estradiol, at which the rest of the molecule will be connected, should be such that this connection does not interfere with its binding to the ER. Several sites on estradiol have been used for attaching components without interfering with its ER-binding ability. Some of these sites are 7α, 11α, 16α, and 17β[7,10,11]. One particular site, the 7α-site, has been used to attach a large alkylating component to estradiol (see Figure 6) and this compound was shown to successfully alkylate the DNA in breast cancer cells[10]. Therefore, this site appeared to be a good choice for connecting the alkylating compound, Me-Lex, that is being used in this project.

It has been shown in literature that the tether which is used to connect different kinds of agents to estradiol at this 7α site has to be at least six-carbons in length for maintaining efficient binding to the ER[12]. Tethers of up to eighteen carbons in length have been successfully used and in fact longer tethers have resulted in better binding to the ER[13]. The end of the tether must have a suitable functional group in order to facilitate connections to other units. Functionalizing the end of the tether with an amine will make it easy to connect it to appropriate moieties ending with a carboxylic acid utilizing standard peptide coupling agents. Based on all these considerations the compound shown in Figure 9 was selected as the cell-targeting segment.

The synthesis of this compound shown in Figure 9, starting from commercially available β-estradiol, has been described in literature[14-18]. The overall strategy that was utilized

is outlined in synthetic scheme 2. In β-estradiol, the 7α-position to which the tether is to be connected is unreactive, and needs to be first converted into a reactive site. This can be accomplished by introducing a ketone at the benzylic C-6 by first converting it into an alcohol, and then oxidizing the alcohol to the ketone. In order to prevent interference form the 3-phenolic and 17-hydroxyl OH groups present on estradiol, they have to be first protected before this conversion can take place.

Figure 9. Structure of the cell-targeting segment that will be synthesized.

Scheme 2. Outline for the synthesis of the cell-targeting segment.

The presence of the ketone at the C-6 position renders the hydrogens at the 7-position acidic, and therefore the tether can now be introduced by using a base and the appropriate alkylating agent. Trapping the enolate formed with triethyl borate before the addition of 6-iodohexene as the alkylating agent results in the introduction of the alkyl grroup at the α-position. The terminal alkene can then be converted into an alcohol by the anti-markovnikov addition of water across the double bond, and the alcohol can then be converted into the desired amine via the Mitsunobu reaction. Thus the cell-targeting component, in the form of compound **14** is now ready for condensation with the DNA recognition component compound **5**.

Choice of Linker:

The linker unit can be varied to optimize the properties of the DNA recognition segment and cell-targeting segment. Some examples of linkers that can be used are shown in Figure 10. The length of the alkyl chain in the linker can be varied to optimize the binding properties of both the DNA alkylating and cell-targeting segments. If increase in water solubility is required, then functional groups capable of hydrogen binding can be introduced in the linker. A tertiary amine can be introduced into the linker if tighter binding to DNA is desired because the amine will get protonated at physiological pH, which will result in stronger binding to the negatively charged DNA backbone.

Assembly of the Final molecule:

Assembly of the components will be achieved as outlined in Scheme 3. Once the individual components are prepared the DNA recognition component, which terminates with carboxylate group will be first hydrolyzed into a carboxylic acid, and then combined with the amine of the cell-targeting segment to give compound **15**. The reactive methylsulfonate ester can

then be introduced as outlined in Scheme 4. The nitro group on compound **15** is reduced to an amine and then reacted with acryloyl chloride to give compound **16**. The alkene can be converted into the desired sulfonic acid **17** by the anti-Markovnikov addition of sodium bisulfite. The protecting groups on the estradiol will be removed before the methylation of the sulfonic acid to yield the final compound **18**, in order to minimize the number of manipulations required after the reactive methyl group is introduced. The free OH groups on the estradiol unit are not expected to interfere with the methylation of the sulfonic acid. However, if necessary, the protecting groups can be removed after the methylation reaction.

The synthetic procedure that has been outlined here will enable us to make the cell-targeting component and the DNA sequence-recognizing component separately. This strategy is also amenable to the introduction of various linker units efficiently just before the assembly of the different components.

Figure 10. Examples of potential linkers with variations in length and composition.

Scheme 3. Strategy for assembling the components.

15

1.) Pd/C, H_2, EtOH
2.) Acryloyl chloride, THF

16

1.) Sodium Bisulfite, H_2O_2 THF/H_2O
2.) HCl

17

1.) 3-methyl-p-tolyltriazine dioxane

18

P_2 =

Scheme 4. Introduction of the reactive methylsulfonate ester into the molecule.

EXPERIMENTAL

General

All solvents and reagents were purchased from VWR International (West Chester, Pennsylvania) or Sigma-Aldrich (Atlanta, Georgia) and were of the highest grade available unless otherwise noted. Flash chromatography was performed with silica gel (230/400 mesh, Life Force Inc.). All rotary evaporations were carried out using a Buchi R-3000 or a Buchi R-114 rotary evaporator equipped with a Brinkmann model B-16 vacuum aspirator. Hydrogenations were performed using a Parr Hydrogenation Apparatus in a 500 mL Parr jar. Melting points were determined using a Mel-Temp II.

All anhydrous reactions were carried out in a fume hood, under a positive pressure of argon. Glassware used for anhydrous reactions were dried overnight in an oven at 140°C, assembled while still hot, and cooled to room temperature under argon. Solvents and reagents (liquids) for anhydrous reactions were obtained in bottles with sure-seal caps and transferred by using oven-dried glass syringes.

All ^{1}H NMR and ^{13}C NMR spectra were recorded with a Bruker 400 MHz NMR spectrometer, using $CDCl_3$ or deuterated DMSO as the solvent. $CDCl_3$ was obtained from Alfa Aesar and the deuterated DMSO was obtained in sealed ampoules from ACROS Organics. The spectra are reported in ppm and referenced to $CDCl_3$ (7.24 ppm, 77 ppm) when acquired in $CDCl_3$ and to DMSO (2.49 ppm, 39.5 ppm) when acquired with deuterated DMSO. The samples were contained in 5 mm pyrex glass tubes obtained from Wilmad- LabGlass Buena, New Jersey.

All TLC were performed on glass plates coated with silica gel (Whatman International, Maidstone, England, 150 A) that had a fluorescence indicator and were detected by UV visualization.

Synthesis of DNA-methylating Component:

(4-nitro-2-trichloroacetyl-1-methyl pyrrole) (2). To a cooled solution (-50 ºC) of **1** (10.6 g, 46.7 mmol) in acetic anhydride (70 mL) was added fuming HNO_3 (4 mL, 95.3 mmol) drop wise over 10 minutes. The solution was allowed to stir at room temperature for 5 h, after which an ice bath was placed under the reaction flask. Then distilled water (70 mL) was added and the reaction stirred overnight. The precipitate that was formed was filtered to obtain the product **2** as a yellow-orange powder in a yield of 8.79 g (79%). mp = 122 - 128 °C, R_f = 0.87 (ethyl acetate), ^{1}H NMR (DMSO): δ 4.00 (s, 3H), 7.78 (d, 1H, J = 2 Hz), 8.56 (s, 1H). ^{13}C NMR (DMSO): δ 95.02, 116.82, 121.08, 133.09, 134.73, 173.29.

(Ethyl β-(1-methyl-4-nitropyrrole-2-carboxamido) alanine) (3). To a solution of ethyl β-alanine hydrochloride (4.9 g, 31.89 mmol) in ethyl acetate (150 mL) was added TEA (8.8 mL, 63.14 mmol) and the mixture was stirred for 5 h at room temperature. At this time a white precipitate will have formed. Then **2** (10.86 g, 39.7 mmol) was added and the reaction was allowed to stir overnight. The white precipitate was formed was removed by filtration. The yellow crystals obtained upon solvent removal were recrystallized from 95% ethanol and water to give 3.7 g (43%) of **3** as a yellow crystal. mp = 125 - 129°C, R_f = 0.60 (ethyl acetate), ^{1}H NMR (DMSO): δ 1.16 (t, 3H, J = 7 Hz), 2.54 (t, 2H, J = 7 Hz), 3.39 (q, 2H, J = 7 Hz), 4.05 (q, 2H, J = 7 Hz), 7.39 (d, 1H, J = 2 Hz), 8.11 (d, 1H, J = 2 Hz), 8.48 (t, 1H, J = 5 Hz). ^{13}C NMR (DMSO): δ 14.53, 34.09, 35.40, 37.82, 60.43, 107.87, 125.98, 126.68, 134.20, 160.32, 171.65.

(Ethyl β-[1-methyl-4-(1-methyl-4-nitroimidazole-2-carboxamido) pyrrole-2-carboxamido] alaninate) (4). To a solution of **3** (1.02 g, 3.8 mmol) in ethanol (25 mL) in a Parr jar was added Pd/C (0.10 g). The reaction jar was then connected to the hydrogenator, flushed with H_2 and evacuated twice, then filled with H_2 to a pressure of 70 psi, and shaken, until TLC (ethyl acetate) indicated the complete disappearance of the starting material **4**. The reaction mixture was then filtered through celite to remove Pd/C and the Pd/C was washed thoroughly with ethanol. The filtrate and washings were combined and the solvent was removed by rotary evaporation. The residue was kept under vacuum for 30 minutes and then ethyl acetate (30 mL) was added to the residue, followed by the addition of **2** (1.02 g, 3.7 mmol). This mixture was allowed to stir overnight upon which a yellow precipitate was formed. The precipitate was then filtered to give 0.62 g (42%) of **4** as a yellow powder. The product was pure and no further purification was necessary. mp = 188 - 192°C, R_f = 0.53 (ethyl acetate), ^{1}H NMR (DMSO): δ 1.16 (t, 3H, J = 7 Hz), 2.53 (t, 2H, J = 7 Hz), 3.37 (q, 2H, J = 7 Hz), 3.78 (s, 3H), 3.94 (s, 3H), 4.05 (q, 2H, J = 7 Hz), 6.82 (d, 1H, J= 2 Hz), 7.20 (d, 1H, J = 2Hz), 7.57 (d, 1H, J = 2 Hz), 8.10 (t, 1H, J = 5 Hz), 8.16 (d, 1H, J = 2 Hz), 10.2 (s, 1H). ^{13}C NMR (DMSO): δ 14.55, 34.46, 35.28, 36.50, 37.94, 60.35, 104.57, 108.03, 118.56, 121.83, 123.42, 126.74, 128.68, 134.23, 157.30, 161.63, 171.83.

(Ethyl β-[1-methyl-4-(1-methyl-4-nitroimidazole-2-carboxamido) pyrrole-2-carboxamido] carboxylic acid) (5). To a solution of **4** (0.5 g, 1.3 mmol) in methanol was added a solution of KOH (0.46 g, 8.20 mmol) dissolved in 0.5 mL of deionized water. The reaction was stirred overnight and then 10% HCl (1 mL) was added. The reaction was allowed to stir overnight when a yellow precipitate fell out. The precipitate was filtered out to give 0.41 g (88%) of **5** as a yellow powder. mp = 235 - 237°C, R_f = 0.14 (ethyl acetate), ^{1}H NMR

(DMSO): δ 2.44 – 2.48 (m, 2H), 3.34 – 3.38 (m, 2H), 3.79 (s, 3H), 3.93 (s, 3H), 6.82 (d, 1H, J = 2 Hz), 7.19 (d, 1H, J = 1 Hz), 7.55 (d, 1H, J = 2 Hz), 8.07 (t, 1H, J = 6 Hz), 8.17 (d, 1H, J = 1 Hz), 10.22 (s, 1H), 12.17 (s, 1H). ^{13}C NMR (DMSO): δ 34.47, 35.33, 36.52, 37.94, 96.19, 98.82, 116.50, 121.81, 123.47, 126.13, 128.69, 133.31, 155.78, 161.61, 170.40.

Synthesis of Estradiol Component:

(3,17β-Bis(2-tetrahydropyranyloxy)estra-1,3,5(10)triene) (6). To a solution of β-estradiol (10 g, 36.7 mmol) in CH_2Cl_2 (125 mL) was added dihydropyran (33 mL, 295.5 mmol) and then pyridinium ρ-toluenesulfonate (8 mg). The reaction mixture was then equipped with a reflux condenser and was allowed to reflux for 3 hours while following by TLC. The solvent was removed by rotary evaporation to give a yellow oil. This yellow oil was purified by flash chromatography (5:1, hexane/EtOAc) to give a clear oil. Ethanol was then added to the clear oil, the flask was then spun in a warm water bath until the oil dissolved and the ethanol turned white. Once this happened the flask was allowed to cool to room temperature upon which a white solid will start to fall out, then the flask was placed into the freezer overnight upon which a white solid precipitated. The precipitate was then filtered off to give 13.05 g (80%) of **6** as a white solid. mp = 75 - 83°C, R_f = 0.42 (5:1, henaxe/EtOAc), ^{1}H NMR ($CDCl_3$): δ 2.82 (d, 2H, J = 4 Hz), 3.72 – 3.94 (m, 1H), 4.64 – 4.68 (m, 1H), 5.38 (s, 1H), 6.77 (d, 1H, J = 2 Hz), 6.83 (d, 1H, J = 8 Hz), 7.17 – 7.21 (m, 1H). ^{13}C NMR ($CDCl_3$): δ 19.20, 19.90, 19.92, 19.96, 23.07, 25.46, 25.56, 25.67, 26.35, 27.25, 28.82, 29.81, 30.44, 31.14, 37.22, 37.80, 38.56, 42.85, 43.36, 44.02, 50.04, 61.95, 62.79, 84.16, 86.65, 96.37, 99.44, 113.88, 116.45, 126.25, 133.71, 137.97, 154.84.

(3,17β-Bis(2-tetrahydropyranyloxy)estra-1,3,5(10)-triene-6-ol) (**7**). To a cooled (-78 °C) solution of **6** (7.0 g, 16 mmol) in THF (40 mL) was added 1.0 M potassium tert-butoxide (KOt-Bu) in THF (33.2 mL, 33.2 mmol), diisopropylamine (0.58 mL 4.10 mmol), and 1.6 M n-

BuLi in hexanes (19 mL, 30.4 mmol). Upon addition of n-BuLi the reaction mixture turned dark red. After 10 minutes trimethyl borate (4.4 mL, 38.75 mmol) was added and the reaction mixture turned yellow/orange. After 5 minutes 35% H_2O_2 (20 mL) was added to the reaction, and a white solid forms. Then after stirring the solid dissolves to give a clear light yellow liquid. The reaction mixture was allowed to stir at room temperature for 1 hour. The organic phase was partitioned between toluene diluted with 10% HCl. The organic layer was dried over $MgSO_4$ and filtered. The solvent was removed by rotary evaporation and the product was purified by flash chromatography (2:1, hexane/EtOAc) to give 10.32 g (75%) of **7** as a white solid. mp = 73 - 77°C, R_f =0.34 (2:1, hexane/EtOAc), ^{1}H NMR ($CDCl_3$): δ 0.81 (d, 1H J = 7 Hz), 3.25 (q, 1H, J = 8 Hz), 4.64 – 4.68 (m, 1H), 4.79 – 4.85 (m, 1H), 5.44 (dt, 1H, J = 3 Hz), 6.95 (d, 1H, J = 6 Hz), 7.18 – 7.22 (m, 1H), 7.29 (1H). ^{13}C NMR ($CDCl_3$): δ 11.72, 18.76, 18.80, 19.35, 19.89, 23.02, 23.14, 25.26, 25.54, 25.64, 26.26, 27.19, 28.75, 30.37, 30.39, 31.07, 31.12, 37.08, 37.63, 37.88, 37.93, 38.10, 38.21, 42.78, 43.28, 44.36, 44.49, 49.39, 49.51, 61.83, 61.99, 62.05, 62.73, 69.93, 69.96, 70.00, 70.03, 84.03, 86.50, 96.20, 96.48, 96.61, 99.42, 114.78, 114.98, 115.70, 115.77, 116.53, 126.27, 126.32, 133.43, 133.49, 133.55, 140.79, 140.84, 155.43, 155.52.

(3,17β-Bis(2-tetrahydropyranyloxy)estra-1,3,5(10)-triene-6-one) (**8**). To a cooled (0 °C) solution of **7** (9.14 g, 20.0 mmol) in CH_2Cl_2 was added 2,2,6,6-Tetramethylpiperidine 1-oxy (0.168 g, 1.1 mmol) and KBr (0.252 g, 2.12 mmol in 8.4 mL of DI H_2O). 14% Sodium hypochlorite (pH was adjusted to 8.5 by adding saturated $NaHCO_3$) was added until a less polar compound was formed. The organic phase was then washed with 10% sodium thiosulfate and deionized H_2O. The organic layer was dried over $MgSO_4$ and gravity filtered. The solvent was removed by rotary evaporation and the product was then purified by flash chromatography (3:1, hexane/EtOAc) to give 7.48 g (82%) of **8** as a white solid. mp = 48 - 60°C, R_f = 0.71 (3:1,

hexane/EtOAc), ^1H NMR ($CDCl_3$): δ 0.81 (d, 3H, J = 6Hz), 2.73 (dd, 1H, J = 17 & 4 Hz), 3.69 – 3.76 (m, 1H), 4.65 (dt, 1H, J = 17 Hz), 5.47 (q, 1H, J = 3 Hz), 7.21 (ddd, 1H, J = 1 Hz, 3 Hz, 8 Hz), 7.33 – 7.36 (m, 1H), 7.71 (t, 1H, J = 3 Hz). ^{13}C NMR ($CDCl_3$) δ 11.56, 11.61, 18.78, 19.41, 19.79, 22.81, 22.92, 25.16, 25.53, 25.63, 27.12, 28.86, 30.23, 31.03, 31.11, 36.71, 37.25, 39.82, 39.85, 39.95, 42.65, 42.97, 43.02, 43.13, 44.12, 49.90, 50.04, 61.98, 62.15, 62.65, 83.87, 86.29, 96.35, 96.39, 96.75, 99.40, 114.06, 122.53, 122.63, 126.48, 126.54, 133.46, 140.40, 140.51, 197.90, 197.97.

(3,17β-Bis(2-tetrahydropyranyloxy)-7α-(5-hexen-1-yl)-estra-1,3,5(10)-triene-6-one) (9). To a solution of **8** (1.0 g, 2.20 mmol) in ethyl ether (10.5 mL) was added 1.0 M KOt-Bu in THF (2.5 mL, 2.5 mmol). After 10 minutes 1.0 M triethylborane (2.8 mL, 2.8 mmol) was added and the reaction mixture was stirred for 1 hour. 6-iodohexene (0.6g, 2.86 mmol) in ethyl ether was canulated into the reaction flask mixture. After 30 minutes of stirring an equivalent amount of 1.0 M KOt-Bu in THF (2.5 mL, 2.5 mmol) was added and the reaction mixture was stirred overnight. The reaction was then quenched with DI water and extracted with CH_2Cl_2. The organic layer was dried over $MgSO_4$ and gravity filtered. The solvent was removed by rotary evaporation and the product was purified by flash chromatography (5.5:0.5, hexane/EtOAc) to give 0.5 g (42%) of **9** as a white solid. mp = 37 - 48°C, R_f = 0.17 (5.5:0.5, hexane/EtOAc), ^1H NMR ($CDCl_3$): δ 0.80 (d, 3H, J = 6 Hz), 2.31 – 2.40 (m, 1H), 2.46 (d, 1H, J = 11 Hz), 2.70 (t, 1H, J = 11 Hz), 3.75 (q, 1H, J = 8 Hz), 4.65 (dt, 1H, J = 15 Hz), 4.89 – 4.97 (m, 2H), 5.47 (d, 1H, J = 3 Hz), 5.71 – 5.82 (m, 1H), 7.20 (d, 1H, J = 8 Hz), 7.30 – 7.33 (m, 1H), 7.69 (t, 1H, J = 2 Hz). ^{13}C NMR ($CDCl_3$): δ 26.62, 26.81, 27.03, 28.62, 28.96, 29.72, 30.26, 30.31, 33.63, 36.97, 37.41, 37.51, 42.25, 42.29, 42.37, 42.80, 43.28, 45.27, 45.36, 48.61, 48.69, 61.95, 62.19, 62.32, 62.62, 83.90, 86.31, 96.33, 96.50, 96.70, 99.35, 114.35, 114.57, 114.66,

122.30, 122.36, 126.99, 127.03, 127.08, 132.27, 132.32, 138.84, 139.41, 139.54, 155.50, 155.56, 200.84, 200.90.

(6-iodo-1-hexene). To a solution of 5-hexen-1-ol (3.34g, 33.4 mmol) in a mixture of diethyl ether/acetonitrile (160/40 ml) was added triphenylphosphine (18g, 68.6 mmol) and imidazole (5g, 73.4 mmol). Iodine (17 g, 66.9 mmol) was then added slowly to the reaction and the reaction was allowed to stir at room temperature until only one spot was seen by TLC. Diethyl ether (100ml) was then added and the reaction mixture was then filtered through celite. The filtrate was then washed with saturated sodium bicarbonate and brine, followed by a water wash. The organic layer was then dried over Mg SO_4 and gravity filtered. The solvent was then removed by rotary evaporation and the product was purified by flash chromatography (hexanes) to give 6.10g (97%) of 6-iodo-1-hexene as a clear liquid. ^{1}H NMR ($CDCl_3$): δ 1.51 (m, 2H), 1.84 (m, 2H), 2.08 (m, 2H), 3.19, (t, 2H, J = 7 Hz), 4.99 (m, 2H), 5.78 (m, 1H).

(7α-(5-hexen-1-yl)-estra-1,3,5(10)-triene-3,17β-diol) (10). To a cooled (0 °C) solution of **9** (4.16 g, 7.8 mmol) in MeOH (186 mL) was added acetyl chloride (18.6 mL, 261.6 mmol) in 1 mL aliquots. The ice water bath was removed and the reaction was allowed to stir at room temperature for 30 minutes. The solvent was removed by rotary evaporation to give a yellow oil. To a cooled (0 °C) solution of the crude oil in CH_2Cl_2 (93 mL) was added triethylsilane (94 mL, 588.5 mmol) and Boron trifluoride diethyl etherate (96 mL, 777.8 mmol). The ice water bath was removed and the reaction solution was allowed to stir overnight. 20% K_2CO_3 (930 mL) was added resulting in a biphasic mixture. The biphasic mixture was then filtered through silica and to the filtrate was added CH_2Cl_2. The organic layer was then dried over Mg SO_4 and gravity filtered. The solvent was then removed by rotary evaporation and the product was purified by flash chromatography (2:1, hexane/EtOAc) to give 2.36 g (85%) of **10**

as a white solid. mp = 109 - 112°C, R_f = 0.24 (2:1,, ^{1}H NMR ($CDCl_3$): δ 0.78 (s, 3H), 2.30 (t, 2H, J = 10 Hz), 2.72 (d, 1H, J = 16 Hz), 2.87 (dd, 1H, J = 17 & 5 Hz), 3.75 (t, 1H, J = 8 Hz), 4.91 – 4.99 (m, 2H), 5.75 – 5.82 (m, 1H), 6.55 (d, 1H, J = 3 Hz), 6.62 (dd, 1H, J = 8 & 2 Hz), 7.15 (d, 1H, J = 8 Hz). ^{13}C NMR ($CDCl_3$): δ 11.10, 14.21, 21.09, 22.67, 25.45, 27.28, 27.65, 29.18, 30.56, 33.17, 33.87, 34.57, 36.88, 38.05, 41.95, 43.39, 46.47, 60.47, 82.06, 112.83, 114.27, 116.13, 127.09, 131.92, 139.04, 153.38.

(3,17β-Bis(τ-butyldimethylsilanyloxy)-7α-(5-hexen-1-yl)-estra-1,3,5(10)-triene) (11). To a cooled (0 °C) solution of imidazole (5 g) in DMF (76 mL) was added tertbutyl dimethylsilyl chloride (5 g, 33.2 mmol). The reaction was then stirred for 30 minutes and then a solution of **10** (2.43 g, 6.85 mmol) in DMF (11 mL) was added all at once. The reaction was then stirred overnight and then was hydrolyzed with 0.1% K_2CO_3. The reaction mixture was then extracted with CH_2Cl_2 and washed with deionized H_2O. The organic layer was dried over $MgSO_4$ and gravity filtered. The product was then purified by flash chromatography (2:1, hexane/EtOAc) to give 3.32 g (82%) of **11** as an oil. ^{1}H NMR ($CDCl_3$): δ 0.05 (d, 6H, J = 8 Hz), 0.2 (s, 6H), 0.8 (s, 3H), 0.93 (s, 9H), 1.0 (s, 9H), 2.3 (t, 2H, J = 10 Hz), 2.7 (d, 1H, J = 17 Hz), 2.85 (dd, 1H, J = 17 & 5 Hz), 3.67 (t, 1H, J = 8 Hz), 4.92 – 5.00 (m, 2H), 5.75 – 5.85 (m, 1H), 6.55 (d, 1H, J = 2 Hz), 6.63 (dd, 1H, J = 8 & 2 Hz), 7.14 (d, 1H, J = 8 Hz). ^{13}C NMR ($CDCl_3$): δ -4.77, -4.45, -4.37, 1.03, 11.40, 14.14, 14.21, 18.13, 18.16, 22.71, 22.81, 25.39, 25.71, 27.31, 27.65, 29.20, 29.38, 30.96, 31.94, 33.27, 33.86, 34.63, 37.39, 38.28, 41.96, 43.73, 46.13, 81.88, 114.20, 117.17, 120.83, 126.68, 132.56, 136.82, 139.11, 153.23.

(3,17β-Bis(τ-butyldimethylsilanyloxy)-7α-(6-hydroxy-hexen-1-yl)-estra-1,3,5(10)-triene) (12) . To a solution of **11** (0.4663 g, 0.800 mmol) in THF (33 mL) was added 0.5 M 9-borabicyclo[3.3.1]nonane in THF (10 mL, 5 mmol). The reaction was then stirred overnight and

hydrolyzed with 3 M KOH (4.2 mL, 12.6 mmol). After 5 minutes 30% H_2O_2 (4.2 mL) was added and the reaction was stirred for 3 hours. Saturated $NaHCO_3$ (42 mL) was added and then extracted with CH_2Cl_2. The organic layer was dried over $MgSO_4$ and gravity filtered. The solvent was removed by rotary evaporation and the product was then purified by flash chromatography (20%, EtOAc/hexane) to give 0.3477 g (72%) of **12** as a white solid. mp = 55°C, R_f = 0.60 (20%, EtOAc/hexane), ^{1}H NMR ($CDCl_3$): δ 0.045 (d, 6H, J = 4 Hz), 0.199 (s, 6H), 0.752 (s, 3H), 0.903 (s, 9H), 0.986 (s, 9H), 2.29 (t, 2H), 2.69 (d, 1H, J = 17 Hz), 2.85 (dd, 1H, J = 17 & 6 Hz), 3.61 – 3.68 (m, 1H), 6.54 (d, 1H, J = 2 Hz), 6.64 (dd, 1H, J = 8 & 3 Hz), 7.12 (d, 1H, J = 8 Hz). ^{13}C NMR ($CDCl_3$): δ - 4.76, - 4.44, - 4.36, 11.40, 18.14, 18.16, 22.82, 25.71, 25.89, 29.80, 30.96, 32.81, 33.30, 34.63, 37.39, 38.29, 41.97, 43.73, 46.12, 63.05, 81.87, 117.18, 120.83, 126.69, 132.58, 136.82, 153.22.

(3,17β-Bis(τ-butyldimethylsilanyloxy)-7α-(6-phthalimido-hexen-1-yl)-estra-1,3,5(10)-triene) (13). To a solution of triphenylphosphine (0.654 g, 2.23 mmol) in THF (22 mL) was added diisopropyl azodicarboxylate (0.491 mL, 2.49 mmol) dropwise. The reaction was allowed to stir for 40 mins and a white precipitate formed. The phthalimide (0.366 g, 2.49 mmol) was then added followed by **12** (0.500 g, 0.83 mmol). The reaction was stirred at 0 °C for 1 hour and then at room temperature overnight. The solvent was removed by rotary evaporation and the product was then purified by flash chromatography (5% hexane/EtOAc) to give 0.5266 g (86%) of **13** as a yellow solid. mp = 58°C, R_f = 0.81 (5% hexane/EtOAc), ^{1}H NMR ($CDCl_3$): δ 0.05 (d, 6H, J = 4 Hz), 0.199 (s, 6H), 0.74 (s, 3H), 0.9 (s, 9H), 0.98 (s, 9H), 2.27 (t, 2H, J = 10 Hz), 2.68 (d, 1H, J = 17 Hz), 2.84 (dd, 1H, J = 17 and 5 Hz), 3.65 (t, 2H, J = 7 Hz), 5.23 – 5.28 (m, 1H), 6.53 (d, 1H, J = 2 Hz), 2.61 (dd, 1H, J = 8 & 3 Hz), 7.11 (d, 1H, J = 8 Hz), 7.72 (q, 2H, J = 3 Hz), 7.84 (q, 2H, J = 3 Hz). ^{13}C NMR ($CDCl_3$): δ - 4.76, - 4.44, - 4.36, 11.39, 18.13, 18.15,

21.62, 22.80, 25.71, 25.89, 28.06, 28.62, 29.60, 30.94, 33.25, 34.57, 37.38, 38.07, 38.26, 41.95, 43.72, 46.09, 81.86, 117.16, 120.82, 123.16, 126.67, 132.19, 132.55, 133.84, 136.80, 153.22, 168.47.

(3,17β-Bis(τ-butyldimethylsilanyloxy)-7α-(6-phthalimido-hexen-1-yl)-estra-1,3,5(10)-triene) (14). To a solution of **13** (0.070 g, 0.096 mmol) in ethyl ether and EtOH was added anhydrous hydrazine (470 μL, 13.37 mmol) at room temperature in one installment. The reaction mixture was then equipped with a reflux condenser and was refluxed for 2 hours. During this time a white solid formed on the walls and the solution turned a slight green/yellow. 5% NaOH was added after the reaction cooled to dissolve the white solid. After 30 minutes DI H_2O was added and then was extracted with CH_2Cl_2. The organic layer was dried over $MgSO_4$ and gravity filtered. The solvent was then removed to give 0.053 g (91%) of **14** as a white solid. ^{1}H NMR ($CDCl_3$): δ -0.013 (d, 6H, J = 4 Hz), 0.014 (s, 6H), 0.69 (s, 3H), 0.84 (s, 9H), 0.93 (s, 9H), 2.21 (t, 2H), 2.65 (d, 2H), 2.79 (dd, 1H, J = 16 Hz), 3.60 (t, 1H, J = 8 Hz), 6.49 (d, 1H, J = 2 Hz), 6.55 (dd, 1H, J = 8 & 2 Hz), 7.06 (d, 1H, J = 8 Hz). ^{13}C NMR ($CDCl_3$): δ - 4.75, - 4.43, - 4.35, - 4.34, - 3.47, 1.04, 11.41, 18.06, 18.13, 18.16, 22.00, 22.83, 25.73, 25.90, 29.82, 29.86, 30.97, 33.31, 33.69, 34.64, 37.40, 38.29, 41.99, 42.10, 43.73, 46.13, 81.88, 113.25, 116.51, 117.18, 120.84, 126.70, 126.87, 130.75, 132.57, 136.82, 136.86, 153.22, 154.98.

Synthesis of coupling DNA recognition and cell targeting segments:

2-{[5-({[5-({[3-({6-[(7β,8ξ,9ξ,13α,14ξ,17α)-3,17-bis{[*tert*-butyl(dimethyl)silyl]oxy}estra-1,3,5(10)-trien-7-yl]-1-hexyl} amino)-3-oxopropyl]amino}carbonyl)-1-methyl-1*H*-pyrrol-3-yl]amino}carbonyl)-1-methyl-1*H*-pyrrol-3-yl]amino}-2-nitro (15). To a dry flask under argon was added **23** (0.032g, 0.0882 mmol) and **17** (0.053g, 0.08804 mmol). Then 1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide

hydrochloride (0.023g, 0.1315 mmol), 4-Dimethylaminopyridine (0.027g, 0.221 mmol), 1-Hydroxybenzotriazole hydrate (0.041g, 0.2678 mmol), and $CuCl_2$ (0.0012g, 0.0089 mmol) was added to the flask. Once all the solids were added DMF (1 mL) was added and the reaction solution was stirred until TLC showed that the reaction was complete. The reaction mixture was then washed with CH_2Cl_2 then $NaHCO_3$, HCl, and finally NaCl. The organic layer was then dried over $MgSO_4$ and filtered. The solvent was removed by rotary evaporation and the product was then purified by flash chromatography (3:1.5 Ethyl Acetate/Hexane) to give 0.0288g (34.5%) of **18** as a yellow solid. ^{1}H NMR ($CDCl_3$): δ 0.023 (d, 6H, J = 4 Hz), 0.18 (s, 6H), 0.73 (s, 3H), 0.89 (s, 10H), 0.97 (s, 10H), 2.47 (t, 2H, J = 6 Hz), 2.65 (d, 1H, J = 16 Hz), 2.85 (dd, 1H, J = 17 & 5 Hz), 3.25 (q, 2H, J = 7 & 13 Hz), 3.65 (m, 3H), 3.86 (s, 3H), 4.03 (s, 3H), 6.38 (s, 1H), 6.4 (d, 1H, J = 2 Hz), 6.53 (d, 1H, J = 2 Hz), 6.6 (dd, 1H, J = 2 & 8 Hz), 2.72 (t, 1H, J = 6 Hz), 7.1 (d, 1H, J = 8 Hz), 7.3 (d, 1H, J = 2 Hz), 7.55 (d, 2H, J = 18 Hz), 9.00 (s, 1H). ^{13}C NMR ($CDCl_3$): δ -4.81, -4.49, -4.41, -4.40, 11.35, 18.07, 18.10, 22.77, 25.66, 25.83, 28.11, 29.46, 29.70, 30.91, 33.26, 34.55, 35.45, 35.68, 36.55, 37.32, 38.12, 38.23, 39.74, 41.90, 43.68, 46.06, 81.79, 103.34, 107.68, 117.17, 119.35, 120.79, 121.08, 123.32, 126.17, 126.71, 126.96, 132.56, 134.96, 136.74, 153.16, 157.67, 161.83, 171.81.

RESULTS AND DISCUSSION

The overall goal of this project was to make compounds that alkylate the N3 adenine of DNA in the minor groove at A/T rich regions in breast cancer cells. In order to do this our tactic was to combine a specific DNA methylating agent with a cell-targeting agent with appropriate linkers. The linkers have to be optimized in order to achieve the best properties of cell targeting and DNA methylation. Therefore it was realized early on that numerous molecules with various linkers will have to be made and then tested to see what effect the linker variations have on the DNA methylation and cell targeting abilities of these molecules. Thus a modular design was developed that would allow the DNA recognition and cell-targeting components to be synthesized separately with appropriate functional groups so that they can be later assembled very quickly with different linker units.

Based on consideration of raw materials available and the design constraints it was decided that compounds **5** and **14** shown in Figure 11 would be made. The DNA recognition component was synthesized as a nitro carboxylic acid without the reactive methylsulfonate and the cell-targeting component was synthesized with an amine at the end so that it can be connected to appropriate carboxylic acids.

Once the linker units to be used, containing an amine at one end and an ester at the other end, are identified, the amine will first be reacted with the carboxylic acid of the DNA recognition component. After the formation of this amide bond, the ester at the other end will be hydrolyzed into a carboxylic acid, which will then be coupled with the amine of the cell-targeting component. Once the units are assembled, the DNA recognition component can be converted into the methylating compound by introducing a methylsulfonate ester, which has the reactive methyl group that will be transferred to the DNA.

In this project the DNA recognition component was synthesized as the compound shown in Figure 12. Details of the synthesis of this component are given in the experimental section and an outline can be seen in Scheme 5 below. While the addition of the linker unit has been accomplished in the second step of this synthesis following methods described in literature[19], it is possible to attach this unit to the dipyrrole later on in the synthetic scheme.

The nitro pyrrole **2** was obtained by nitration of pyrrole **1** using fuming nitric acid in acetic anhydride at –40 ^{o}C. Upon completion of the reaction, an excess of water is added to convert all the acetic anhydride into acetic acid, in which the desired product has lower solubility than the starting material. Therefore, the product falls out of solution and can simply be isolated by simple filtration. A small amount of the undesired 5-nitro isomer is detected by NMR, but the 5-nitro can be separated from the 4-nitro by their difference in solubility since the 5-nitro is more soluble in acetic acid. On occasion, when the acetic acid is diluted with an excessive amount of water, both isomers precipitate. Therefore, careful dilution of acetic anhydride with water results in the precipitation of only the desired 4-nitro compound **2**.

Figure 11. Structure of the DNA recognition with the linker attached and cell targeting components.

Figure 12. Structure of the DNA recognition compound that was actually made.

Scheme 5. Synthesis of the DNA recognition component.

The trichloro group in compound **1** is stable to aqueous hydrolysis; however, it reacts very efficiently with amines. This enables the subsequent coupling reaction to be performed without the use of dry solvents, which would be required if acid chlorides were used to achieve the coupling. Also, upon reaction with an amine, the trichloro group is converted to chloroform, which can be easily removed by rotary evaporation.

The linker (scheme 5) is obtained as the hydrochloride salt of the amine, and the free amine is then generated, in situ, by using triethylamine as a scavenger of HCl. As the free amine is generated, it reacts with the trichloro compound to form the peptide linkage to give the nitro pyrrole ester **3**. Compound **3** can be further purified through recrystalization from ethanol.

The nitro compound **3** was hydrogenated under high pressure in order to obtain the amine. The amine is sensitive to air oxidation and has to be coupled to compound **2** immediately in order to obtain dipeptide ester compound **4**.

The ester of compound **4** can be easily hydrolyzed to the carboxylic acid of **5** and then can be used to couple with the cell-targeting segment, which will be in the form of an amine. The yields for each step in the synthesis of these compounds are summarized in Table 1 below. The overall yield of scheme 6 is 11%. Some improvements can be made particularly to the reaction that gives the dipeptide ester compound **4** by bubbling argon through the solution so that air oxidation could have been minimized. This was not attempted because starting materials for this part of the synthesis were inexpensive and large quantities were easily obtained. Now that the synthesis of the DNA recognition component is complete, it is now ready to be coupled to the cell-targeting component.

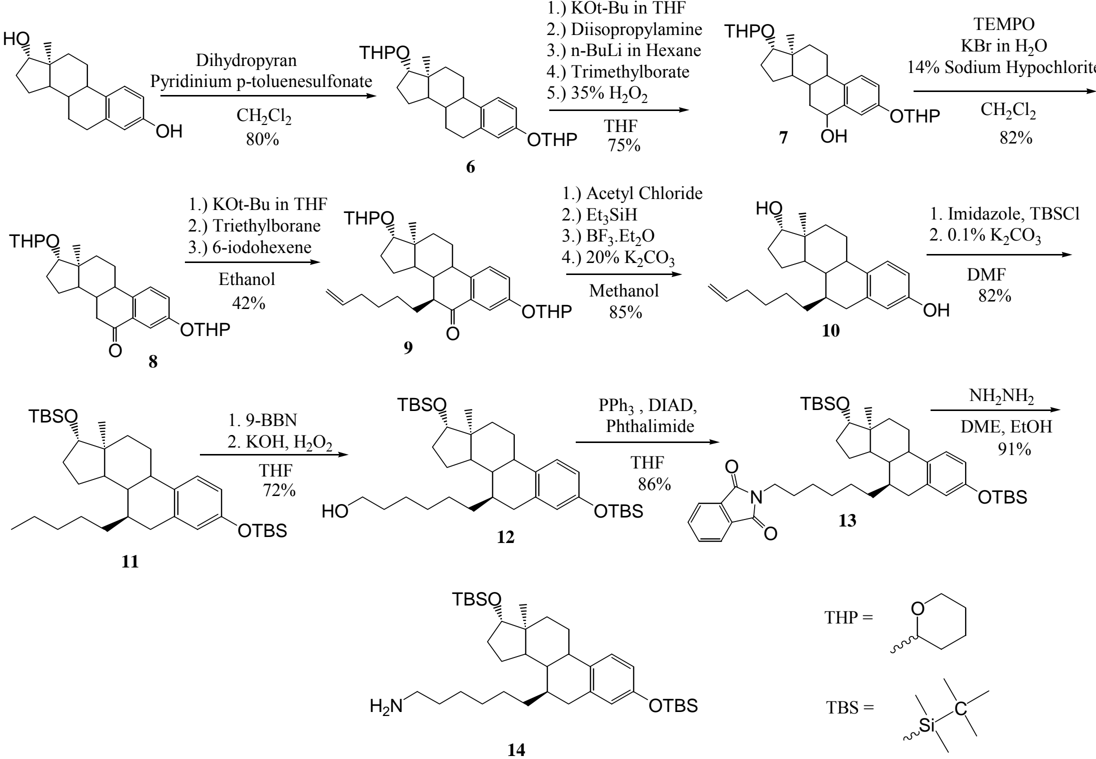
Structure	Precent	Compound #
O_2N N CCl_3 O	79%	2
O_2N N H N O O OEt	43%	3
NO_2 N NH O N H N O O OEt	42%	4
NO_2 N NH O N H N O O OH	80%	5

Table 1. Percent yields for the compounds in the synthesis of the DNA recognition component.

In this project the cell-targeting component **14** was synthesized as the compound shown below in scheme 6. The synthesis of the cell-targeting component has been described in

literature in different ways. The procedures that were used were adapted from the literature to achieve the synthesis that was used and these procedures have been described in the experimental section and are outlined in scheme 6.

As was discussed earlier, before functionalizing the benzylic position the –OH groups have to first be protected and the procedure described in literature used dihydropyran[14]. However the use of dihydropyran as the protecting group leads to problems in the characterization of compound **6** using 13C NMR because there are more carbons in the NMR than in the compound. This problem is due to the mixture of diastereoisomers as a result of an asymmetric center that is formed during the addition of the dihydropyran protecting groups[20]. The proton NMR of protected compound **6** was compared to the spectrum given in the literature of the synthesis that was followed. Also the proton NMR spectra of the subsequent steps had the expected changes showing that the protected compound **6** was formed.



Scheme 6. Synthesis of the cell-targeting component.

All descriptions of compound **6** in literature reveal it to be a viscous oil and we obtained an oil after the work up of **6.** The actual yield obtained was much larger than the theoretical yield and when placed under prolonged vacuum a decrease in weight was noticed. This analysis showed that a significant amount of this oil was solvent and despite extensive evacuation not all of the solvent was removed from this viscous oil. Since we could not quantify the proportion of the mixture that was solvent, calculations of the appropriate quantities for the subsequent steps were very difficult. Therefore we under took further efforts to better purify this compound **6** and for the very first time we succeeded in crystallizing this compound as a solid by using ethanol. Numerous attempts were taken to recrystallize protected compound **6** and it was found that recrystallization at low temperatures worked best.

Protected ketone compound **8** is obtained through a two-step procedure by converting protected estradiol **6** into protected alcohol compound **7**[15]. Then the alcohol of **7** is converted into the ketone of **8**. The alcohol in compound **7** is a mixture of two isomers but this does not matter because the alcohol is oxidized into ketone. Initially pyridinium chlorochromate (PCC) was used to convert the alcohol of **7** into the ketone of **8**[14] however high yields were not obtained because the compound stuck to the PCC. Therefore we followed a different procedure that used sodium hypochlorite and TEMPO, which gave much-improved yields[16].

6-iodo-1-hexene was used to introduce the 6-carbon tether to the 7α position of compound **8** but 6-iodo-1-hexene is not commercially available. Initially 6-iodo-1-hexene was obtained by reacting 6-bromo-1-hexene with NaI[17] but due to incomplete conversion of 6-bromo into 6-iodo there was a mixture of the two compounds. The two compounds are so similar in polarity that they cannot be separated by column chromatography. Furthermore 6-bromo-1-hexene was expensive; therefore, an alternate procedure was used in which 5-hexen-1-ol[18] is

converted into 6-Iodo-1-hexene by reacting imidazole, triphenylphosphine, and iodine. 5-hexen-1-ol is less expensive than the previous compound 6-bromo-1-hexene. Using this procedure we were able to obtain 6-iodo-1-hexene in better yields. Also due to the large difference in polarity between the product alkene and the starting alcohol, they can be easily separated by using column chromatography.

In order to attach 6-carbon tether to the 7α position of **8** the enolate is first generated by using potassium tertbutoxide. Then triethyl borane is used to trap the enolate to prevent O-alkylation and then water is added in order to quench the reaction[17]. The TLC showed that only one compound was formed which was confirmed by NMR. When comparing the NMR obtained and the NMR in literature all peaks matched up. This evidence shows that this reaction produces only compound **9**. The yield of **9** is only 42% however unreacted starting material from this reaction can be recovered and the adjusted yield is 60%.

The next step is to reduce the ketone of **9** to give compound **10**. It was shown in the literature that the protecting groups should be removed first to give better yields. Thus the procedure that was used removed the protecting group first by using acetyl chloride and methanol[16]. Then the ketone was reduced by using Et_3SiH and BF_3Et_2 to give compound **10**. The alcohol groups on **10** were then reprotected with tertbutyl dimethylsilyl (TBS) to give protected compound **11**[17].

The terminal alkene of compound **11** was then converted into the terminal alcohol of compound **12** by the anti-Markovnikov addition of water using 9-BBN and H_2O_2[17]. The terminal alcohol of **12** can be converted into the phthalimide of **13** by using Mitsunobo conditions[17]. The conversion of the phthalimide of **13** into the amine of **14** is very efficient and can be achieved in yields greater than 90%[17]. Thus the amine compound **14** can be stored as the phthalimide **13** to

avoid any possible decomposition. The synthesis of the cell-targeting component is complete and it is now ready to be coupled to the linker.

The next step was to put together compound **5** and compound **14** to form compound **15**. Compound **15** has both the DNA recognition and the cell-targeting components connected together to form one compound that could now potentially alkylate DNA in breast cancer cells. The details of this coupling reaction are found in the experimental chapter and are outlined in scheme 7. Thus for the first time we have succeeded in assembling a DNA recognition segment with a cell-targeting segment with a certain linker.

Further efforts are on-going in the laboratory to convert the nitro into the methylsulfonate ester (scheme 8). This conversion involves the reaction of the nitro group into the amine immediately followed by the condensation of the amine with acryloyl chloride to make the olefin. The olefin will then be reacted with sodium bisulfite to give the sulfonic acid. The sulfonic acid will then be converted into the methylsulfonate ester upon reaction with 3-methyl-p-toyltriazine to give the final compound **18**.

NO_2 N NH O H N O OH O N **5**

TBSO H_2N OTBS **14**

NO_2 N NH O N H N O O NH TBSO OTBS **15**

Scheme 7. Synthetic strategy for coupling the DNA recognition and cell targeting components.

Scheme 8. The strategy for conversion of the nitro into the methylsulfonate ester.

In this project we have succeeded in assembling together a DNA recognition component and a cell-targeting component into a single molecule, and developed an efficient synthetic procedure that is amendable to the synthesis of many different compounds varying in composition of the linker. Studies in the laboratory have shown that various linkers with different lengths can be added to the carboxylic acid end of the DNA recognition component shown in figure 13. Therefore we have developed a procedure that can be used to easily assemble multiple components in order to make numerous compounds.

Once the methylating compounds are made, there are two different properties that have to be evaluated, one is the cell targeting ability and the other is the ability to form N3

methyladenine adducts. In future studies, experiments will be conducted to determine the DNA methylating ability by conducting HPLC experiments. The compound will be reacted with genomic DNA and the methylated bases will be isolated, characterized, and quantified by using HPLC methods. Chemical sequencing analysis and gel electrophoresis will be used to determine at what sequence the methylation of the bases occurs. These experiments will verify if the DNA recognition component still methylates DNA in the minor groove at A/T rich regions at the N3 adenine. These results from the methylation studies will be correlated to the structure and composition of the linker. Cells that do and do not express the estrogen receptor will be used to test the cell targeting ability of these compounds. The cells will be exposed to the N3-methyl adenine forming compounds and if targeting is achieved then the compounds will be more toxic to cells that express the ER.

CONCLUSION

The main goal of this project was to make a new compound by coupling a DNA recognition component and a cell-targeting component. The other goal was to develop a synthetic strategy that will allow easy variation in the mode of connection of the above two components. In this project, for the very first time, we have managed to make a compound that has a DNA recognition component and a cell-targeting component in a single molecule. This compound is three steps away from being converted into the desired methylating compound, which can be tested for its DNA-methylating and breast cancer cell-targeting properties. We have also succeed in developing a modular synthetic strategy in which the two different functional components – the DNA recognizing unit and the breast cancer cell-targeting unit – are synthesized separately, and can then be assembled together with varying linking units as desired.

REFERENCES

1. *The Anticancer Drugs* (1994) 2nd ed. (Pratt, W. B., Ruddon, R. W., Ensminger, W. D., and Maybaum, J., Eds) Oxford Press, New York.

2. Zhang, Y.; Chen, F.; Mehta, P.; Gold, B. "Groove- and Sequence-Selective Alkylation of DNA by Sulfonate Esters Tethered to Lexitropsins" Biochem. 1993, 32, 7954-7965.

3. *Chemical Carcinogens* (1984) 2nd ed., ACS Monograph (Searle, C. E., Ed.) Vols. I and II, American Chemical Spciety, Washington, D. C.

4. Shah, D.; Kelly, J.; Zhang,Y.; Dande, P.; Martinez, J.; Ortiz, G.; Fronza, G.; Tran, H.; Soto, A. M.; Marky, L.; Gold, B. "Evidence in Escherchia coli that N3-Methyladenine Lesions Induced by a Minor Groove Binding Methyl Sulfonate Ester Can Be Processed by both Base and Nucleotide Excision Repair" Biochem. 2001, 40, 1796-1803.

5. Varadarajan, S.; Shah, D.; Dande, P.; Settles, S.; Chen, F.; Fronza, G.; Gold, B. "DNA Damge and Cytotoxicity Induced by Minor Groove Binding Methyl Sulfonate Esters" Biochem. 2003, 42, 14318-14327.

6. Lemieux, P.; Fugua, S. "The Role of the Estrogen Receptor in Tumor Progression" J. Steriod Biochem. Molec. Biol. 1996, 56, 87-91.

7. Swamy, N.; James, D. A.; Mohr, S. C.; Hanson, R. N.; Ray, R. " An Estradiol-Porphyrin Conjugate Selectively Localizes Into Estrogen Recepotr-Positive Breast Cancer Cells" Bioorg. Med. Chem. 2002, 10, 3237-3243.

8. Pike, A. C. W.; Brzozowski, A. M.; Walton, J; Hubbard, R. E.; Thorsell, A.; Li, Y.; Gustafsson, j.; Cariquist, M. "Structural Insights into the Mode of Action of a Pure Antiestrogen" Structure. 2001, 9, 145-153.

9. Murphy, L. C.; Dotzlaw, H.; Leygue, E.; Coutts, A.; Watson, P. "The Pathophysiological Role of Estrogen Receptor Variants in Human Breast Cancer" J. Steriod Biochem. Molec. Biol. 1998, 65, 175-180.

10. Essigmann, J. M.; Rink, S. M.; Park, H.; Croy, R. G. "Design of DNA Damaging Agents That Hijack Transcription Factors And Block DNA Repair" Bio. Reac. Inter. VI. 2001, 301-313.

11. Kuduk, S. D.; Zheng, F. F.; Sepp-Lorenzino, L.; Rosen, N.; Danishefsky, S. J. "Synthesis and Evaluation of Geldanamycin-Estradiol Hybrids" Bioorg. Med. Chem. Let. 1999, 9, 1233-1238.

REFERENCES CONT.

12. Rink, S. M.; Yarema, K. J.; Solomon, M. S.; Paige, L. A.; Tadayoni-Rebek, B. M.; Essigmann, J. M.; Croy, R. G. "Synthesis and biological activity of DNA damaging agents that form decoy binding sites for the estrogen receptor" Biochem. 1996, 93, 15063-15068.

13. Anstead, G. M.; Carlson, K. E.; Katzenellenbogen, J. A. "The estradiol pharmacophore: Ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site" Steriods. 1997, 62, 268-303.

14. Mitra, K.; Marquis, J. C.; Hillier, S. M.; Rye, P. T.; Zayas, B.; Essigmann, J. M.; Croy, R. G. "A Rationally Designed Genotoxin that Selectively Destroys Estrogen Receptor-Positive Breast Cancer Cells" J. Am. Chem. Soc. 2002, 124, 1862-1863.

15. Spera, D.; Cabrera, G.; Fiaschi, R.; Carlson, K. E.; Katzenellenbogen, J. A.; Napolitano, E. "Estradiol derivatives bearing sulfur-containing substituents at the 11β or 7α positions: versatile reagents for the preparation of estrogen conjugates" Bioorg. Med. Chem. 2004, 12, 4393-4401.

16. Tedasco, R.; fiaschi, R.; Napolitano, E. "6-Oxoestradiol from Estradiols: Exploiting Stie Selective Metalation of Aralkyl Systems with Superbases" Synthesis. 1995, 1493-1495.

17. Skaddan, M. B.; Wust, F. R.; Katzenellenbogen, J. A. "Synthesis and Binding Affinities of Novel Re-Containing 7α-Substituted Estradiol Complexes: Models for Breast Cancer Imaging Agents" J. Org. Chem. 1999, 64, 8108-8121.

18. Spino, C.; Barriaylt, N. "Radical Cyclization of Polyenes Initiated by Attack of Trialkyltin or Germanium Radical on an Ynone" J. Org. Chem. 1999, 64, 5292-5298.

19. Xiao, J.; Yuan, G.; Huang, W. "A Covenient Method for the Synthesis of DNA-Recognizing Polyamides in Solution" J. Org. Chem. 2000, 65, 5506-5513.

20. Boucheau, V.; Renaud, M.; Ravel de Rolland, M.; Mappus, E.; Cuilleron, C. Y. "Proton and carbon-13 nuclear magnetic resonance spectroscopy of diastereoisomeric 3- and 17β-tetrahydropyranyl ether derivatives of estrone and estradiol" Steriods 1990, 55, 209-221.

APPENDIX A

Numbering of Compounds

....1

....2

....3

....4

....5

....6

....7

....8

APPENDIX A CONT.

……………………………………………………………9

……………………………………………………………10

……………………………………………………………11

……………………………………………………………12

……………………………………………………………13

……………………………………………………………14

…………………………………..15

APPENDIX B

Abbreviations

Deoxyribosenucleic acid..DNA

Adenine ....A

Thymine ....T

Guanine ....G

Cytosine C

Estrogen Receptor...ER

Deuterated Chloroform……………………………………………………….$CDCl_3$

Dimethyl Sulfoxide..DMSO

Melting Point…………………………………………………………………mp

Tetrahydrofuran………………………………………………………………THF

Tertbutyl dimethylsilyl..TBS

Tetrahydropyranyl...THP

Potassium tert-butoxide ....KOt-Bu

Retention Factor..R_f