A DATABASE FOR NUCLEAR RECEPTOR DNA BINDING SITES

Timothy E. Breen Ph.D.

Submitted to the faculty of the School of Informatics
in partial fulfillment of the requirements
for the degree
Master of Science
in Bioinformatics,
Indiana University

August 2007

Accepted by the Faculty of Indiana University,
in partial fulfillment of the requirements for the degree of Master of Science
in Bioinformatics.

Master's Thesis
Committee

_____
Vladimir Uversky, Ph.D.

_____
Lang Li, Ph.D.

_____
Yunlong Liu, Ph.D.

_____
Narayanan Perumal, Ph.D.

For my wife Nancy, daughter Erin and son Patrick, their support and patience have made it possible to continue and their presence has made it worthwhile.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACKNOWLEDGEMENTS

I want to express my gratitude to the members of my thesis committee for their patience and support.

I want to thank Erin E. Breen for her editorial review of this thesis.

I also would like to thank Beverly Musick for her support throughout this Masters Program.

Abstract
Timothy E. Breen Ph.D.

A DATABASE FOR NUCLEAR RECEPTOR DNA BINDING SITES

A database has been developed for nuclear receptor DNA binding sites and the corresponding nuclear receptor DNA binding domain (DBD) amino acid sequences. This database contains three main tables nuclear receptor DNA footprints, nuclear receptor DNA binding sites and the corresponding nuclear receptor DBD. The ultimate goal of this database is to provide a repository for all experimentally derived nuclear receptor DNA binding sites. Finite DNA binding sites and footprints are obtained from original peer reviewed research that used selected experimental methods to identify nuclear receptor DNA binding sites. Acceptable methods for the identification of DNA binding sites and/or footprints include reporter genes and DNA deletion analysis, DNAase I footprinting, exonuclease III footprinting, methylation protection, methylation interference, nuclear receptor protein mutagenesis and the electrophoretic mobility shift assay. With respect to this database, a nuclear receptor footprint is a DNA sequence less than 50 nucleotides in length identified by one of the methods listed above. A binding site is a DNA sequence identified within one of the footprints that is less than 15 nucleotides in length and corresponds to a previously identified half site, response element or consensus sequence. The five prime and three prime location of each footprint with respect to the transcription initiation site is noted in the database. The corresponding nuclear receptor DBD amino acid sequence is obtained from the Swiss-Prot database. The nuclear receptor data is annotated

with the literature reference, DNA source, DNA binding site identification method, the nuclear receptor Swiss-Prot primary accession number, and inter halfsite spacing if specified. The tables are related by the literature reference, DNA source, experimental method and nuclear receptor ID. Sixty one publications have been identified from 1983 through 1994 that contain nuclear receptor binding sequences identified with one of the methods listed above. These publications were identified by selected review articles, cited references, citation index and PubMed searches. Currently, the database contains 21 DNA binding sites and 30 DNA footprints from 14 publications between 1983 and 1987. As the database expands, the literature citations will be used to design new search strategies for the nuclear receptor literature. After the current literature has been surveyed and the database is up to date, submissions will be accepted from past and future research. This database will eventually become publicly accessible. In order to interpret the experimentally derived DNA footprints, the specific DNA base - NR protein interactions are summarized from six X ray crystallographic studies.

CHAPTER ONE: INTRODUCTION

Introduction

The concept of the nuclear receptor (NR) developed from investigations of the mechanism of action of steroid hormones (Jensen & DeSombre 1973, Tata 2002). The effect of the steroid hormone estrogen on tumor growth has been known for more than one hundred years (reviewed in Green & Chambone 1991, Jensen & Jordan 2003), however, it was not until 1966 that the receptor protein that specifically binds estrogen was isolated from rat uterus (Toft & Gorski 1966). The estrogen receptor was the first nuclear receptor isolated. In 1968, two laboratories independently proposed the existence of a cytosolic receptor that is activated by binding estrogen and that the activated estrogen-receptor complex relocates to the nucleus (Jensen et al. 1968, Gorski et al. 1968). These were the first formal descriptions of the nuclear receptor. During the same time period, evidence was accumulating that steroid hormones increase the expression of specific genes and the synthesis of the encoded proteins (Peterkofsky & Tomkins 1968). By the early seventies, the concept of the nuclear receptor was developing as a protein that binds with steroid hormones in the cytosol, the steroid-receptor complex then migrates to the nucleus and enhances the expression of specific genes.

The second NR to be isolated, in 1979, was the glucocorticoid receptor (Wrange et al. 1979, Westphal & Beato 1980). In addition, in 1971 the glucocorticoid, hydrocortisone was shown to induce the replication and release of

mammary tumor virus from mouse mammary adenocarcinoma cells (McGrath 1971). These findings and the availability of purified glucocorticoid NR provided a model to investigate the binding of the activated glucocorticoid NR to viral DNA (Ringgold 1985). One of the first demonstrations of specific binding of the activated glucocorticoid NR to viral DNA was provided by Yamamoto's laboratory in 1981 (Payvar et al. 1981). In this study, purified glucocorticoid receptor was shown to specifically bind to a distinct fragment of murine mammary tumor virus (MMTV) DNA as compared to other MMTV DNA fragments or E. coli plasmid DNA restriction fragments. The MMTV DNA sequence that specifically binds the glucocorticoid receptor was not detailed in this study. This finding was further investigated in a number of studies in 1982, however, the specific MMTV DNA sequence that binds the glucocorticoid receptor was not identified in these studies either. (Fasel et al. 1982, Geisse et al. 1982, Govindan et al. 1982, Pfahl 1982).

Prior to 1983, several genes including the human growth hormone gene (Robins et al. 1982) and the rat tryptophan oxygenase gene (Schmid et al.1982) were shown to be responsive to glucocorticoid hormones. Once again, the specific DNA sequences that mediate glucocorticoid responsiveness were not experimentally identified in these studies. In 1983, the first finite specific DNA binding sequence for a nuclear receptor was experimentally identified using the DNAase footprinting technique (Scheidereit et al. 1983). In this study, purified rat liver glucocorticoid receptor protected four regions of a 438 bp DNA fragment of the MMTV long terminal repeat region. The protected regions were 13, 14, 17 and 25 bp in length. This was the first application of the DNAase footprinting

technique in the identification of the DNA binding sites of a NR. The availability of purified NRs and the use of techniques that reveal the actual DNA binding sites have provided the capability to study NR-DNA interactions.

Nuclear receptors are a superfamily of ligand activated DNA binding proteins that regulate gene expression by binding to specific DNA regulatory elements (Mangelsdorf et al. 1995). This superfamily of proteins has a well conserved general structure consisting of four to five functional domains (Aranda & Pascaul 2001). From the amino terminus (N) to the carboxyl tenminus (C), the domains are labeled A/B, C, D, E and F (Figure 1). The A/B domain is the most variable region of NRs due to alternate splicing and the use of alternate promoters. This region defines multiple isoforms of a NR and is a target for phosphoylation. The D domain is the hinge region between the DNA binding domain and the ligand binding domain. This region is not well conserved and functions in nuclear localization and interaction with corepressors. The ligand binding domain (E) consists of two conserved regions the signature motif or Ti and AF2 (activation function 2) motif. In addition to ligand binding, several other functions are incorporated in this domain, including homo-hetero dimerization, interaction with heat shock proteins and ligand dependent transcription activation. A fifth region (F) is present in some NRs. The function of this region is unknown. Two sub-regions have also been described AF1 and AF2 (activation function 1 and 2). The AF1 sub-region is locatied in the A/B domain and is associated with constitutive ligand independent transcriptional activation. In contrast, the AF2 sub-region is a part of the ligand binding domain (E). AF2 is conserved among members of the NR superfamily and is associated with ligand dependent transcriptional activation.

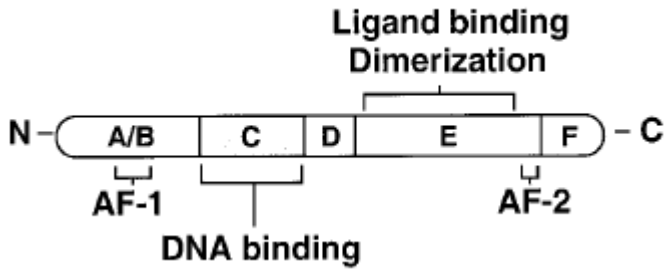

Figure 1. A schematic representation of the functional domains of nuclear receptors (Figure 2, Aranda & Pascual 2001).

The C – domain or DNA binding domain (DBD) is one of the most conserved regions of NRs and spans 60 to 70 amino acids. The DBD is stabilized by two zinc fingers (Figure 2). Each zinc finger is coordinated by four cysteine side chains. The secondary structure of the DBD is an alpha helix which has been designated the recognition helix. There is a second alpha helix in the carboxyl terminus of the second zinc finger. Two sub-regions of the DBD have been designated the P box (proximal) and the D box (distal or dimerization). The P box is located between the third and fourth cysteines of the first zinc finger (G-S in Figure 2). The D box is labeled the dimerization loop in figure 2 and functions in the protein-protien interactions between adjacent NRs bound to DNA. Another sub-region of the DBD is the carboxyl-terminal extension (CTE) which contains a T box and an A box. A third alpha helix is found in the CTE of some NRs.

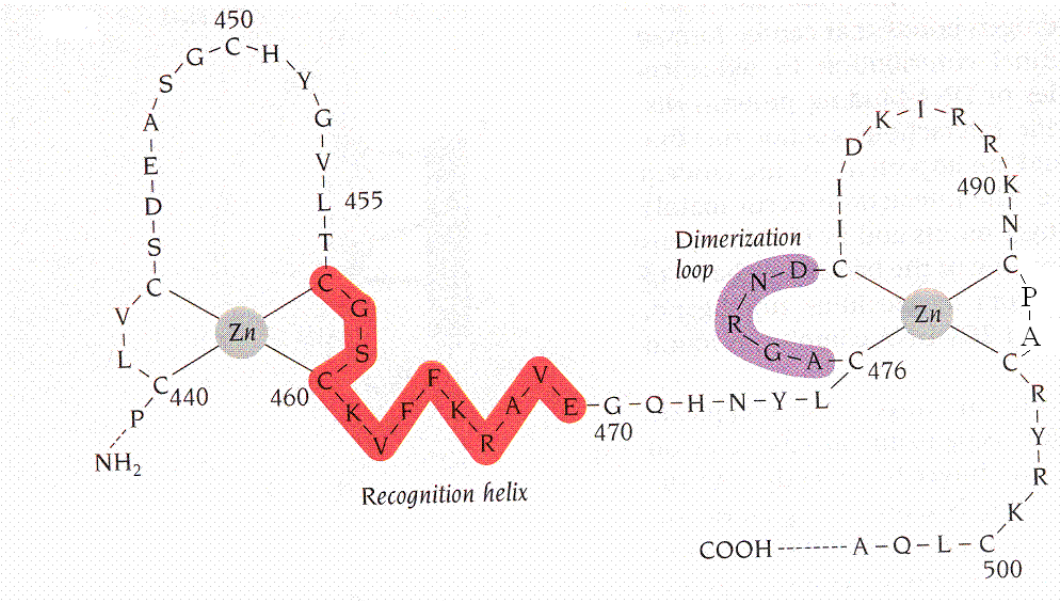

Figure 2. The DNA binding domain (DBD) of the glucocorticoid receptor. The amino acid sequence of the amino and carboxy zinc fingers are illustrated. The recognition helix is highlighted in red and the dimerization loop in purple (modified from Figure 10.8, Branden & Tooze 1999).

An estimate of the number of nuclear receptors in the human genome was generated by using a series of Bioinformatics methods and manual examination of all alignments with E-value $\leq 10^{-4}$ (Robinson-Rechavi 2001). Forty nine nuclear receptor genes were identified in the human genome along with an additional three nuclear receptor-related pseudogenes. A conservative estimate of 75 nuclear receptor protiens, based on alternate splicing and a review of the literature, was also made in the same study. With the availability of a draft of the rat genome, a comparison of human, mouse and rat nuclear receptor genes was published in 2004 (Zhang et al. 2004). This comparison revealed 47 rat, 49 mouse and 48 human NR genes. Likewise, 3 rat and 3 human pseudogenes were identified while 4 mouse pseudogenes were found. Eleven syntenic blocks containing NR genes

were found to be common to all three species. The DBD amino acid sequence showed little variation among the human, rat and mouse genomes. In contrast to mammalian species, only 20 NR genes were found in the fly, Drosophila melanogaster, genome, while at least 270 NR genes were predicted in a search of the genome of the nematode Caenorhabditis elegans (Adams et al. 2000, Sluder & Maina 2001).

Significance

Nuclear receptors have been implicated in a number of diseases (Tenbaum & Baniahmad 1997). Inactivating mutations in the androgen receptor result in androgen insensitivity syndrome. Similar mutations of the vitamin D receptor produce vitamin D resistance. Another mutation of the androgen receptor increases transcriptional activation and results in spinal and bulbar muscular atrophy known as Kennedy's disease (Achermann & Jameson 2003, Tenbaum & Baniahmad 1997). Perhaps the most well known association of a NR with disease is the involvement of the estrogen receptor with breast cancer (Deroo & Korach 2006). Today, estrogen receptor expression in some breast cancers is a biomarker to determine the use of estrogen receptor antagonists in the treatment of breast cancer.

Background

Identification of functional DNA regulatory elements has proven to be a challenging problem (Wasserman & Sandelin 2004). While much progress has been made in the characterization of the genomic coding regions for a number of organisms, the regulatory regions remain less well understood. Identification of

functional DNA regulatory elements requires delineation of the interaction of DNA regulatory proteins and DNA regulatory sequences in the context of DNA expression experiments.

Identification of nuclear receptor DNA regulatory elements is complicated by the diversity of NR-DNA interactions. Nuclear receptors can bind to functional DNA regulatory sequences as monomers, homodimers or heterodimers. The DNA regulatory elements binding NR dimers are called response elements and are composed of two half sites. Each half site binds an individual ligand activated NR. These half sites can be arranged as direct repeats, inverted repeats, everted repeats or unique half sites (Table 1). Nuclear Receptor binding to the DNA half site is mediated by the recognition helix, located at the carboxyl end of the first zinc finger (Figure 2). The NR protein-DNA interactions have been examined with x-ray crystallography for at least six different NR's: the glucocorticoid receptor (GR), the estrogen receptor (ER), RevErbA-a, the nerve growth factor-inducible-B (NGFI-B), the 9-cis retinoic acid receptor (RXR) and the all trans retinoic acid receptor (Luisi et al. 1991, Schwabe et al. 1993, Zhao et al. 1998, Meinke & Sigler 1999, Rastinejad et al. 2000, Zhao et al. 2000). All these studies demonstrate specific contacts between NR amino acid side chains and the DNA bases of the half sites. In the GR study, an 86 amino acid fragment of the GR DBD was bound to a response element composed of palindromic hexameric half sites separated by four base pairs (5'-CC**AGAACA**TCGA**TGTTCT**G-3') (Luisi et al., 1991). Three specific base contacts were observed between GR amino acid side chains of the recognition helix and the three prime half site. Labeling the base pairs of the three prime hexameric half site from five prime (1) to three prime (6),

the amino acid side chain of arginine 466 hydrogen bonds with the G2 and valine 462 makes van der Waals contact with T3. Lysine 461 forms a direct hydrogen bond with the guanine complementary to C5 and water mediated hydrogen bonds to the same guanine and T4. In addition to the specific base contacts, there are a number of amino acid side chain interactions with the DNA phosphate backbone. While this was the first x-ray crystallographic study of a NR DBD complexed with a DNA response element, this study was limited in that the inter-half site spacing allowed only one of the NR receptors of the dimer to bind in the major groove of one of the half sites. These crystallographic studies indicate that the specificity of NR-DNA binding is mediated by specific protein-base pair interactions. A comprehensive review of NR-DBD/DNA response element x-ray crystallographic findings will be presented in the results section.

Table 1.

An example of response elements for the thyroid alpha-1 nuclear receptor (TRE).

(a.) IAP-TRE1G 5’-acttTTGAACTCAgccTGAGGTTACcaaact-3’
(b.) core DR8.4.7 5’-aaAGGTCAtaaaAGGTCAc-3’
(c.) core ER7.4.7 5’-aTGACCTtaaaAGGTCAc-3’
(d.) core IR6.4.9 5’-AGGTCAtaaaTGACCTcgc-3’

a. the wild type footprint for the TRE of the human intestinal alkaline phosphatase gene (hIAP) with a three bp inter half site space, b. a direct repeat that functions as a TRE, c. an everted repeat that functions as a TRE, d. an inverted repeat that functions as a TRE. Response elements b-d have a four bp inter half site spacing. Upper case sequences are TRE half sites. (Malo et al., 2005).

A model of NR-DNA binding has been proposed based not only on NR-DNA interactions, but also on the protein-protein interactions between the individual NR monomers comprising the bound dimer (Freedman & Luisi, 1993). In addition to the protein-DNA interactions, x-ray crystallographic studies also

show extensive protein-protein interactions between the monomers bound to a response element. These protein-protein interactions are mediated mainly through protein residues between the first and second cysteines in the second zinc finger (Figure 2). Binding of a NR DBD monomer to the first half site increases the affinity at the second half site for the other monomer by two orders of magnitude (Freedman & Luisi, 1993). As evidenced by the crystallography study mentioned above, the response element inter half site spacing must accommodate the protein-protein interactions between NR monomers to optimize DNA Binding. Another model of NR-DNA binding emphasizes the interaction of five prime and three prime nucleotides at the boundaries of the response element (Malo et al., 2005). The "Swinging Cradle" model proposes that the response element acts as a platform for the bound NR dimer, while the nucleotides just beyond the boundaries of the response element stabilized the protein-DNA interactions. The human intestinal alkaline phosphatase gene (hIAP), thyroid alpha 1 NR and 9-cis retinoic acid NR were the basis for this study. However, electrophoretic mobility shift binding assays and luciferase reporter assays were carried out using consensus direct repeats, everted repeats and inverted repeats that are not found in the hIAP TRE (Table 1).

Research Hypothesis

The complexity of NR-DNA interactions and diversity of response elements creates a challenge with respect to the prediction of functional NR response elements. Naturally occurring response elements rarely contain a consensus direct repeat, inverted repeat or everted repeat. For example, the first study to employ DNA footprinting techniques identified multiple copies of the consensus half site

TGTTCT for the GR, but the corresponding consensus half site was not located in any of the response elements (Scheidereit et al., 1983). Therefore, the prediction of functional NR response elements requires a more complete knowledge of experimentally determined NR regulatory elements. While several NR databases exist that contain consensus half sites, none of these databases catalogue the actual experimentally derived NR regulatory sequences (Martinez et al., 1998, Horn et al., 2001, Duarte et al., 2002). A database with experimentally derived NR regulatory elements and the corresponding NR DBDs could provide the information necessary to predict functional NR regulatory elements.

Hypothesis:

Experimentally derived nuclear receptor response elements obtained from a review of the literature will not contain theoretical consensus direct repeats, inverted repeats or everted repeats.

Therefore, a database containing experimentally identified nuclear receptor response elements will provide the information necessary to investigate and predict NR-DNA interactions. The creation of a comprehensive database of nuclear receptor DNA regulatory elements cannot be completed in the course of a Masters Thesis project or in any finite time period. The purpose of this project is to demonstrate the necessity and feasibility a database for nuclear receptor DNA regulatory elements. In addition, this project will establish a nuclear receptor DNA binding site database resource.

CHAPTER TWO: METHODOLOGY

Compilation of DNA Binding Sites

Nuclear receptor response elements were obtained from a review of the literature beginning with four seminal review articles. (Yamamoto 1985, Evans 1988, Glass 1994, Aranda & Pascul 2001). These four review articles were chosen because they emphasize NR-DNA interactions and provide a rich review of the NR response element literature. Once the first publication was identified, that detailed an acceptable response element, citation by later articles led to other literature sources of acceptable response elements. The literature references for the database were both manually and electronically procured. Manual procurement took the form of examining cited references from original research papers and identification of additional nuclear receptor literature reviews. Electronic techniques included Citation index, PubMed and other electronic literature search tools. It is envisioned that as this reference list of publications with experimentally identified responses elements grows, the reference list itself may be used to examine search algorithms for the biological literature.

Response Elements and Experimental Techniques

Two types of nuclear receptor DNA regulatory elements have been identified in the current project, footprints and binding sites. A nuclear receptor DNA footprint is defined as a DNA sequence less than 50 nucleotides in length and identified with one of a selected set of experimental techniques defined below. A nuclear receptor binding site is a DNA sequence identified within one of

the footprint sequences that is less than 15 nucleotides in length and corresponds to a previously identified half site, response element or consensus sequence. The DNA footprints were the primary DNA regulatory elements studied in this project. The actual DNA sequence entered into the database will be the sequence complementary to the DNA template strand for mRNA synthesis. Determination of the correct DNA sequence is problematic due to ambiguity in the description of each DNA strand over time. The definition of coding strand in the **Dictionary of Microbiology and Molecular Biology** best describes the problem of identifying the correct DNA strand (Singleton & Sainsbury 2006):

"**coding strand** A term which (like non-coding strand) is frequently used in the opposite meanings by different authors. Originally, the term was generally used to refer to that strand of a gene which acts as the template on which mRNA is synthesized during transcription – the strand whose sequence is *complementary* to that of mRNA. Currently, there appears to be a consensus for the opposite meaning, i.e. that strand of a gene which is *not* transcribed (i.e. not used as a template) and whose sequence is homologus to that of mRNA. (Even so, given the confusion present in the literature, it may be wise to define the term, in relation to mRNA, whenever it is used.) The coding strand is also called the sense strand or the plus (+) strand.

The non-coding strand of a gene (which is complementary to the coding strand) is thus the template strand; it is also called the antisense strand or the minus strand."

The coding or sense strand will be determined for each DNA regulatory element entered into the database. Table 2 lists other terms for each DNA strand encountered in the literature review during this project.

Table 2.
Other Terms for each DNA Strand.

DNA Strand Homologus to mRNA	DNA Strand Complementary to mRNA
Sense strand	Antisense strand
Plus (+) strand	Minus (-) strand
Anti-template strand	Template strand
Codogenic strand	Complementary strand
Non-transcribing strand	Transcribing strand

Only a limited number of experimental techniques reveal nuclear receptor DNA footprints as defined above. Currently, seven experimental techniques provide acceptable DNA footprints (Table 3).

Table 3.
Acceptable Experimental Techniques.

Experimental Technique
1. Reporter Genes and deletion analysis
2. DNAase I footprinting
3. Exonuclease III footprinting
4. Methylation protection assay
5. Methylation interference assay
6. Protein Mutagensis
7. Electrophoretic mobility shift assay

Each experimental technique is best paired with reporter gene experiments for the DNA footprint sequence under consideration, however, not all publications confirm an identified footprint with reporter gene experiments.

Database Design

The database used to manage the NR-DNA binding site data consists of five tables (Appendix 1 – table structures and Figure 3). One of the main tables (NR-DNA BINDSITE) contains the DNA binding site sequence, a text code for the DNA source information, inter-half site spacing, a text code for the experimental method of DNA binding site identification and a text identifier for the NR protein. This table is also annotated with the literature reference and publication authors. DNA source information consists of an identifier that links to another table with the specific source of the DNA used in experimental design. The method of identification also consists of an identifier which links to a table listing the techniques used to identify the DNA sequence involved in the DNA-protein interaction. The NR protein identifier is a combination of the NR nomenclature and the species source of the NR (Nuclear Receptors Nomenclature Committee 1999). For example, there are entries for the chicken and rabbit progesterone NR. The NR nomenclature for the progesterone NR is NR3C3. The NR protein identifiers for the chicken and rabbit progesterone receptors are CHXNR3C3 and RABNR3C3 respectively. For protein mutagenesis experiments the NR protein identifier is derived from the nomenclature used in the publication.

The other main table (NR-DNA FOOTPRINT) in the database contains information about the DNA footprint. Specifically, this table contains the five prime and three prime boundaries of the DNA footprint relative to the gene start site, the DNA footprint, a text code for the DNA source and the NR protein identifier. The literature reference is repeated in this table for convenience.

There are three additional tables in the database that contain annotation

information. The first of these tables (NR-DNA PROTEIN) includes the NR protein identifier, a text description of the NR protein, the species source of the NR protein, the Swiss-Prot primary accession number and the amino acid sequence of the NR recognition helix. The second table (DNA Source) lists a text description of the DNA source such as "Rat Tryptophan Oxygenase Gene". The third table (DNA Methods) contains a text description of the experimental method used to reveal the DNA footprint such as "methylation protection".

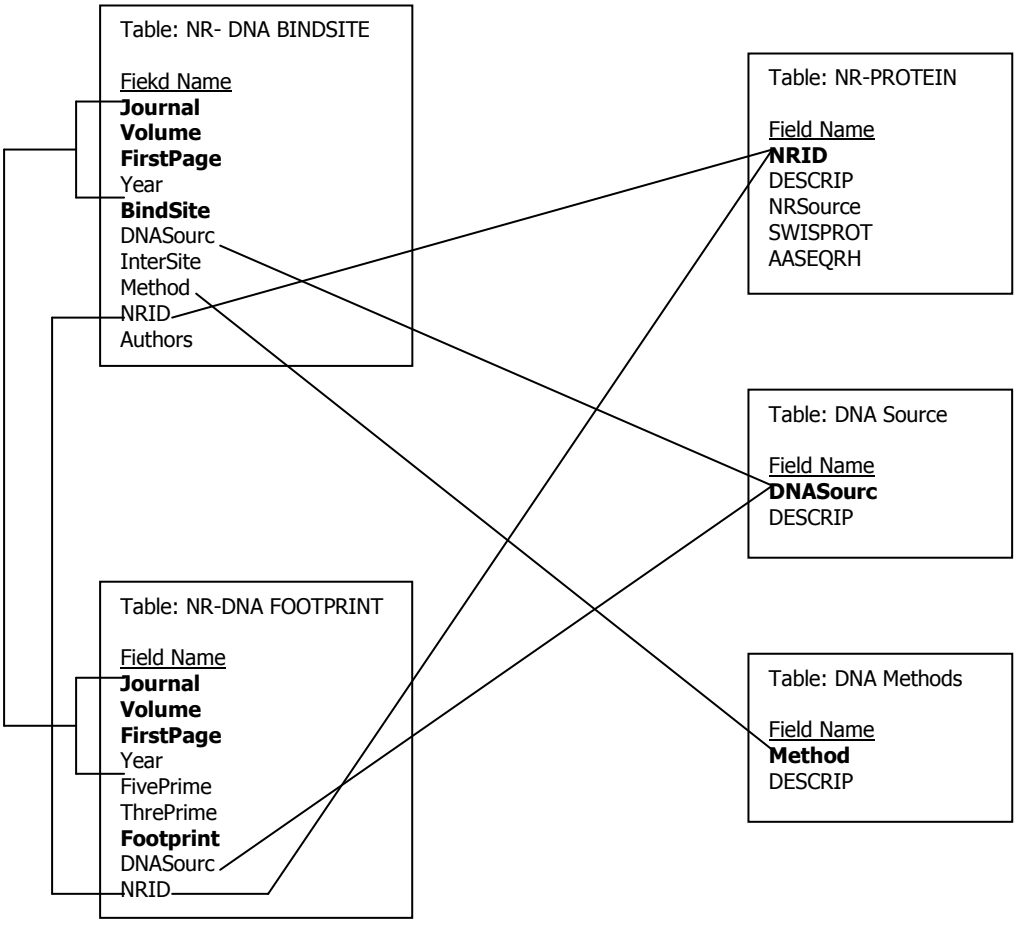

Figure 3. Entity Relationship Diagram of the Nuclear Receptor DNA Binding Site Database. The primary key fields for each table are in bold.

CHAPTER THREE: RESULTS

The results consist of examples of six acceptable experimental techniques that identified NR DNA binding sites and footprints, a review of X ray crystallographic analysis of six NR-DNA interactions, a description of experimentally identified GR footprints and analysis of the five prime location of the GR footprints. The five selected experimental techniques represent the acceptable techniques from Table 3. A review of X ray crystallographic findings provides the necessary basic understanding of NR-DNA interactions to interpret the experimentally identified NR footprints. The analysis of the five prime location of experimentally derived NR footprints will provide an indication of the importance of relative location with respect to the transcription initiation start site. Finally, the examination of experimentally derived NR footprints will test the hypothesis that natural GR response elements are not consensus direct repeats, inverted repeats or everted repeats.

Examples of Acceptable Experimental Techniques

Reporter genes and DNA deletion analysis is the most fundamental method of NR DNA binding site identification. The DNA regulatory region of a gene known to be regulated by a NR is deleted from the five prime direction and the three prime direction until gene expression is inhibited. A classic example of this technique was the identification of estrogen response elements by Seiler-Tuyns et al. in 1986. The regulatory region of the Xenopus Laevis Vitellogenin B1 gene was hybridized with the reporter gene, chloramphenicol acetyltransferase (CAT). Five prime and three prime deletion mutants were transfected into estrogen

responsive MCF-7 cells. Seventeen five prime deletion mutants and 10 three prime deletion mutants from -596 to -42 base pairs from the Xenopus Laevis Vitellogenin B1 gene transcription initiation start site were evaluated for CAT activity. A 48 base pair sequence from -334 to -287 was identified that displayed estrogen responsiveness (5'-AGTCACTGTGACCCAACCCAAGTTATCATGACCTCTTAGTTGGCTATG-3'). The authors speculate that a 13 base pair sequence in the five prime region of the footprint is critical for estrogen responsiveness (5'-AGTCACTGTGACC-3'). The two sequences demonstrate the difference between a DNA footprint and a DNA binding site as described above.

The DNAase I protection assay was first utilized to identify the DNA binding site for rat liver glucocorticoid receptor (GR) (Scheidereit et al. 1983). A 438 base pair fragment of MMTV DNA was incubated with GR isolated from rat liver. Four DNA footprints were protected following DNAase I digestion of the receptor-DNA complex (-187 to -163, -122 to -105, -102 to -89, -83 to -71). The footprints ranged in length from 13 base pairs to 25 base pairs. A hexameric sequence, 5'-TGTTCT-3', was common to all four footprints. Functional GR binding to MMTV DNA was confirmed with a thymidine kinase reporter gene assay and monoclonal antibodies to the GR. Once again the difference between footprints (four protected sequences) and DNA binding sites (5'-TGTTCT-3') is illustrated.

The exonuclease III protection assay identifies discrete DNA footprints in a manner similar to the DNAase I protection assay. Exonuclease III digests double stranded DNA from the three prime end, while DNAase I digests single and

double stranded DNA starting from the middle of the molecule. In the first report to describe the use of the exonuclease III protection assay to identify NR footprints, von der Ahe et al. identified progesterone receptor DNA binding sites (von der Ahe et al. 1985). Chicken lysozyme gene regulatory DNA was bound with rabbit uterus progesterone receptor and exposed to exonuclease III digestion. Two DNA footprints were protected from exonuclease digestion at -200 to -160 and -80 to -54 base pairs with respect to the transcription initiation start site. Two binding sites were also described (5'-TGTTCT-3', 5'-ATTCCTCTGT-3'). The exonuclease footprints of the GR and progesterone receptors were compared on chicken lysozyme gene regulatory DNA and both NRs were found to bind to the same sites.

Methylation protection and methylation interference are two addition assays that can be used to identify DNA-protein binding sites (Guille & Kneale 1997, Shaw & Stewart 2001). These are complementary procedures. In the methylation protection assay, protein is bound to DNA first then the DNA is methylated. Conversely, DNA is partially methylated first and then protein bound in the methylation interference assay. Both techniques were used by Scheidereit and Beato to detect rat liver GR binding sites on MMTV DNA (Scheidereit and Beato 1984). Methylation interference identified one guanine residue (-174) on the sense strand and two guanine residues (-171 and -180) on the antisense strand that interfere with GR binding. The methylation protection assay identified seven sense strand guanine residues (-82, -97, -118, -145, -153, -174, -184) and five anti-sense guanine residues (-79, -94, -115, -171, -180) that were protected by GR binding. On the basis of a binding assay and both methylation assays the

consensus sequence for the GR binding site was -180 to -164 (5'-CAAACTGTTCTTAAAAC-3'). A binding site was described (5'-TGTTCT-3') that was located in sequences identified by both methylation protection and methylation interference assays.

Information concerning NR-DNA interactions can also be obtained by modifying the NR amino acid sequence and evaluating binding to known DNA sequences. Umesono and Evans used NR protein mutagenesis to investigate the effect of various amino acid substitutions on functional binding to GR or thyroid receptor response elements (Umesonon & Evans 1989). Chimeric NRs composed of the GR amino terminal, the thyroid receptor DNA binding domain and the GR carboxyl terminal (GTG) were tested with a GR response element and thyroid receptor response element linked to a luciferase reporter gene. The GR (GTG1) with a glycine to glutamic acid substitution in the P-Box displayed luciferase activity similar to the full human GR with the GR response element (5'-AGAACAnnnTGTTCT-3') but little activity with the thyroid response element (5'-AGGTCATGACCT-3'). The GR (GTG8B) with three amino acid substitutions in the P-Box (GS—V to EG—G) and five amino acid substitutions in the D-Box (AGRND to KYEGK) displayed luciferase activity similar to the full GTG chimeria with the thyroid response element but no activity with the GR response element. The additional five amino acid substitution in the D-Box changed the specificity of the GR to that of a thyroid receptor. NR protein mutagenesis information can be stored in the database by recording the novel protein sequences with the known DNA binding sites.

The database entries for some of the preceding five examples are displayed

in tables 4-8.

Table 4.
NR-DNA Binding Site Table.

Journal	Volume	FirstPage	Year	BindSite	DNASourc	InterSite	Method	NRID	Authors
NATURE	304	749	1983	TGTTCT	MMTV		DNASEI	RATNR3C1	SCHEIDEREIT,GEISSE,WESTPHAL,BEATO
PROC NATL ACAD SCI USA	81	3029	1984	TGTTCT	MMTV		MI,MP	RATNR3C1	SCHEIDEREIT,BEATO
NATURE	313	706	1985	TGTTCT	MMTV		EXONUCIII	RABNR3C3	VON DER AHE, JANICH,SCHEIDEREIT,RENKAWITZ,SCHUTZ,BEATO
NATURE	313	706	1985	ATTCCTCTGT	CLYSOZYM		EXONUCIII	RABNR3C3	VON DER AHE, JANICH,SCHEIDEREIT,RENKAWITZ,SCHUTZ,BEATO
NATURE	313	706	1985	TGTTCT	CLYSOZYM		EXONUCIII	RATNR3C1	VON DER AHE, JANICH,SCHEIDEREIT,RENKAWITZ,SCHUTZ,BEATO
NATURE	313	706	1985	ATTCCTCTGT	CLYSOZYM		EXONUCIII	RATNR3C1	VON DER AHE, JANICH,SCHEIDEREIT,RENKAWITZ,SCHUTZ,BEATO
NATURE	313	706	1985	TCTTCT	CLYSOZYM		EXONUCIII	RABNR3C3	VON DER AHE, JANICH,SCHEIDEREIT,RENKAWITZ,SCHUTZ,BEATO
NUCLEIC ACIDS RES	14	8755	1986	AGTCACTGTGACC	VITELLB1		DNAMUTA	HUMNR3A1	SEILER-TUYNS,WALKER,MARTINEZ,MERILLAT,GIVEL,WAHLI
CELL	57	1139	1989	TGTTCT	RPTPLSMD	3	PROTMUTA	HGTG1	UMESONO,EVANS
CELL	57	1139	1989	AGAACA	RPTPLSMD	3	PROTMUTA	HGTG1	UMESONO,EVANS
CELL	57	1139	1989	TGACCT	RPTPLSMD	0	PROTMUTA	HGTG8B	UMESONO,EVANS
CELL	57	1139	1989	AGGTCA	RPTPLSMD	0	PROTMUTA	HGTG8B	UMESONO,EVANS

Table 5.
NR-DNA Footprint Table.

Journal	Volume	FirstPage	Year	FivePrime	ThrePrime	Footprint	DNASourc	NRID
NATURE	304	749	1983	-83	-71	TGTTCTTTTGGAA	MMTV	RATNR3C1
NATURE	304	749	1983	-102	-89	TTAGTGTTCTATTT	MMTV	RATNR3C1
NATURE	304	749	1983	-122	-105	AAATGTTCTGATCTGAGC	MMTV	RATNR3C1
NATURE	304	749	1983	-187	-163	ATGGTTACAAACTGTTCTTAAAACA	MMTV	RATNR3C1
PROC NATL ACAD SCI USA	81	3029	1984	-180	-164	CAAACTGTTCTTAAAAC	MMTV	RATNR3C1
NATURE	313	706	1985	-80	-54	GTTTTTGACAACTGTAGAACAGAGGAA	CLYSOZYM	RABNR3C3
NATURE	313	706	1985	-200	-160	ACAGACTATAAAATTCCTCTGTGGCTTAGCCAATGTGGTAC	CLYSOZYM	RABNR3C3
NUCLEIC ACIDS RES	14	8755	1986	-334	-287	AGTCACTGTGACCCAACCCAAGTTATCATGACCTCTTAGTTGGCTATG	VITELLB1	HUMNR3A1

Table 6.
NR Protein Table.

NRID	DESCRIP	NRSource	SWISPROT	AASEQRH
CHXNR3C3	CHICKEN PROGESTERONE	CHICKEN	P07812	CLICGDEASGCHYGVLTCGSCKVFFKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRKC
HE30	HUMAN ER CHIMERIA	HUMAN		CAVCNDYASGYHYGVWSCEGCKAFFKRSIQGVNDYMCPATNQCTIDKNRRKSCQACRLRKC
HGTG1	HUMAN GR CHIMERIA	HUMAN		CLVCSDEASGCHYGVLTCESCKVFFKRAVEGQHNYLCAGRNDCIIDKIRRKNCPACRYRKC
HGTG8B	HUMAN GR CHIMERIA	HUMAN		CLVCSDEASGCHYGVLTCEGCKGFFKRAVEGQHNYLCKYEGKCIIDKIRRKNCPACRYRKC
HUMNR3A1	HUMAN ESTROGEN	HUMAN	P03372	CAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKC
RABNR3C3	RABBIT PROGESTERONE	RABBIT	P06186	CLICGDEASGCHYGVLTCGSCKVFFKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRKC
RATNR3C1	RAT GLUCOCORTICOID	RAT	P06536	CLVCSDEASGCHYGVLTCGSCKVFFKRAVEGQHNYLCAGRNDCIIDKIRRKNCPACRYRKC

Table 7.
DNA Methods Table.

Method	DESCRIP
DNAMUTA	DNA Mutagensis
DNASEI	DNAase I Protection Assay
EXONUCIII	Exonuclease III
MI	Methylation Interference
MP	Methylation Protection
PROTMUTA	Protein Mutagenesis

Table 8.
DNA Source Table.

DNASourc	DESCRIP
CLYSOZYM	Chicken Lysozyme
HGHPE1I1	Human Growth Hormone (promoter, exon1,intron1)
HMT-IIA	Human Metallothionein IIA
MMTV	Mouse Mammary Tumor Virus
MOMSVLTR	Moloney Murine Sarcoma Virus LTR
RPTPLSMD	Reporter Plasmid
RTO	Rat Tryptophan Oxygenase Gene
RUG	Rabbit Uteroglobin Gene
VITELLB1	Xenopus vitellogenin B1
VITELLB2	Xenopus vitellogenin B2

X ray Crystallographic Literature Review

The interpretation of experimentally derived NR-DNA binding sites and footprints would be more meaningful if the molecular interactions between NRs and DNA were understood. One method which provides an approximation of such information is X ray crystallographic analysis of the NR DBD/DNA complex. Currently, at least seven studies have successfully analyzed the crystal structure of the NR DBD/DNA complex (Luisi et al. 1991, Schwabe et al. 1993, Rastinejad et al. 1995, Zhao et al. 1998, Meinke & Sigler 1999, Rastinejad et al. 2000, Zhao et al. 2000). Six of these seven studies are summarized herein, the seventh study did not detail most of the specific NR DBD/DNA base contacts and is not included in the summary (Rastinejad et al. 1995). The six X ray crystallographic studies summarized, examined the following NRs : glucocorticoid receptor - GR, estrogen receptor - ER, RevErbA-a, nerve growth factor-inducible-B – NGFI-B, 9-cis retinoic acid receptor – RXR and all trans retinoic acid receptor – RAR. The DNA major groove interactions with these selected NRs are summarized in figure 4, while minor groove NR-DNA interactions are summarized in figure 6. The NR amino acid residues which interact with DNA bases are summarized in figure 5.

Seven response elements binding five NR dimers (GR-GR, RevErbA-a-RevErbA-a, RXR-RAR, RXR-RXR) and one NR monomer (NGFI-B) are displayed in Figure 4. Double stranded DNA response elements are represented by the single letter codes for each base. Response element half sites are displayed in blue and red. NR amino acid residues are represented by the single letter designation followed by the DBD sequence position relative to the third cysteine residue in the first zinc finger (i.e. the third cysteine residue would be labeled C1).

NR amino acid residues are positioned under the response element DNA base to which they bind. The molecular interactions between amino acid residue side chains and DNA bases are indicated in parentheses following the amino acid designation (H – direct hydrogen bond, W – water mediated hydrogen bond and VW – van der Waals interaction). For example, the direct hydrogen bond between the fifth lysine residue (K5) of the GR DBD and the fourth guanine base of the response element is listed below position four of the response element as K5(H)-G in Figure 4, Panel A.. The GR-GR major groove interactions in Panel A. show limited binding in the 5' half site due to the four base pair inter-half site spacing (Luisi et al., 1991). Likewise, ER-ER major groove interactions were only reported for the 5' half site (Schwabe et al. 1993). The DNA construct used by Zhao et al. 2000, consisted of two direct repeats joined in one double stranded DNA sequence (Zhao et al. 2000). The two direct repeats are separated in Panel F. for comparison with the other response elements.

The NR DBD – DNA major groove interactions are similar across the seven response elements displayed in Figure 4. The lysine residue in the fifth position (K5) binds to a half site guanine (G) in eleven of the twelve half sites either through a direct hydrogen bond (H) or water mediated hydrogen bond (W). This K5 – G interaction includes the singular NR monomer, NGFI-B, in Figure 4, Panel D. Arginine ten (R10) also binds to a half site guanine (G) in all response elements except for the NR monomer, NGFI-B, Panel D. The glutamic acid residue in position two (E2) is present in five of the six NRs examined and binds to the response element for each. E2 makes single cytosine base contacts on eight half sites and interacts with two adjacent cytosine bases on two half sites (Panel

E). In addition to the cytosine base contacts, E2 interacts with an adenine base on six half sites. Five of the E2 – adenine interactions are water mediated. Interestingly, the RXR interactions with the half site 5'-AGGTCA-3' show some variability (Panels E and F). The three prime half sites of Panels E and F display three different patterns of interaction with respect to E2. The ninth amino acid position in the DBD also contributes to NR DBD – DNA binding. In four of the NRs the ninth position is a lysine (GR, ER, NGFI-B, RXR), while two NRs contain an arginine at this position (RevErbA-a, RAR). The ninth position lysine (K9) in the GR and NGFI-B do not interact with half site bases (Panels A and D). Likewise, the RXR ninth position lysine does not interact with the five prime DR1 half site, however the three prime DR1 half site interacts with K9 through a single water mediated hydrogen bond to a guanine (Panel F, DR1). The ER and RXR ninth position lysine make multiple contacts through both direct and water mediated hydrogen bonds (Panels B and F). The ninth position arginine (R9) in the RevErbA-a receptor and RAR also interact through multiple direct and water mediated hydrogen bonds (Panels C and E). The GR provides an additional amino acid side chain – half site interaction through the valine in the sixth position (V6). This valine interacts with a thymine base by means of a van der Waals interaction (Panel A).

NR – half site interactions are not limited to DBD amino acid residues. The tyrosine residue five amino acids upstream of the DBD (Y-5) in the RAR, forms a water mediated hydrogen bond with a guanine half site base (Panel E). The RXR displays a water mediated hydrogen bond between arginine 57 (R57) in the second zinc finger with an adenine half site base and the thymine of the inter-half site

spacer (Panel E.). The DBD amino acid residues involved in specific DNA base interactions, including tyrosine -5, are summarized in Figure 5. Amino acid residues interacting with half site bases are shown in red. The amino acid residue positions interacting with DNA half sites, across all NRs examined, are quite similar.

NRs exhibit DNA base specific binding outside of the DNA major groove. DNA minor groove – NR interactions are summarized in Figure 6. The format of Figure 6 is similar to Figure 5 (see the description above). The NR RevErbA-a makes several base specific minor groove contacts with bases upstream of the three prime halfsite (Panel A). The four amino acid residues involved are part of the carboxyl-terminal extension (V55, G58, P61 and K62). These interactions are by means of direct and water mediated hydrogen bonds and a van der Waals interaction. The DNA major groove interactions of the monomer NGFI-B receptor consists of two hydrogen bonds, however, overall NR-DNA binding is enhanced by multiple DNA minor groove contacts (Panel B). The minor groove of the B-DNA configuration is widened with the interaction of three carboxyl-terminal extension amino acid residues of NGFI-B (R63, G64 and R65). All three amino acid residues interact with DNA bases through direct hydrogen bonds.

Glucocorticoid Receptor DNA Footprints

The glucocorticoid receptor was the first NR used in a DNAase I footprinting experiment (Scheidereit et al.1983). Following this initial NR footprinting experiment, the DNA footprint of the GR has been identified in many additional studies. The database for NR binding sites currently contains 29 GR footprints (Table 9). These footprints are used to explore the GR response elements for direct repeats, inverted repeats and everted repeats. The first three response elements listed in table 9 are the consensus half site 5'-TGTTCT-3' (red) configured as a direct repeat, palindromic inverted repeat and palindromic everted repeat, all with three base pair inter-half site spacing. The footprints below the consensus response elements are experimentally derived from ten studies using seven different DNA sources. The DNA sources include viral and eukaryotic DNA (Table 9, column 4 and abbreviations). The five prime and three prime positions relative to the transcription initiation start site are listed in columns one and two respectively (Table 9.). The methods used to identify the GR footprints are also listed in table 9. The GR DNA footprints were searched for the following half sites: 5'-TGTTCT-3', 5'-TCTTGT-3', 5'-AGAACA-3', 5'-ACAAGA-3' and 5'-TGTACC-3'. The matches for any of these half sites are listed in red on table 9. Three studies used MMTV DNA and two studies used Moloney murine sarcoma virus long terminal repeat DNA (MOMSVLTR).

A consensus direct repeat, inverted repeat or everted repeat was not found in any of the 29 experimentally derived footprints (Table 9.). Fifteen of the 29 footprints contained one of the consensus half sites, but not one of the footprints

contained a complete consensus response element. One footprint contained a sequence that was one nucleotide from a direct repeat (Table 9, Cato et al. 1984, -2644 to -2619). Fourteen footprints did not contain any consensus half sites. These footprints, without consensus half sites, were observed in viral DNA and in eukaryotic DNA (Table 9 - MMTV, MOMSVLTR, HMT-IIA, RUG, HGHPE1I1, CLYSOZYM and RTO). MMTV DNA was analyzed in two studies using DNAase I footprinting (Scheidereit et al. 1983, Payvar et al. 1983). Two footprints were found in common in these studies (Table 9 Scheidereit et al. 1983: -187 to -163, -102 to -89 and Payvar et al. 1983: -189 to -166, -127 to -84). These two pairs of MMTV common sequences show a single base difference (Table 9 bases highlighted in blue). A third study analyzing MMTV DNA using the methylation interference (MI) and methylation protection (MP), found only one footprint (Scheidereit & Beato 1984). This single footprint was also identified in the two other studies that examined MMTV DNA. The difference in the number of MMTV footprints between the DNAase I and MI/MP studies may result from a difference in sensitivity of the two techniques. Each of the two studies analyzing MOMSVLTR DNA using DNAase I footprinting, found three footprints corresponding to the same sequences (Miksicek et al. 1986, Defranco et al. 1986). The consensus half sites identified in MOMSVLTR DNA were identical in both studies. Interestingly, the same MOMSVLTR DNA sequence, without a consensus half site, was identified in both studies.

The GR footprints, displayed in Table 9, range in length from 13 to 45 base pairs. The mean GR footprint length was 25 base pairs. There was a large variation in the location of the footprints as determined by the five prime

boundaries of the footprints relative to the transcription initiation start site. The maximum upstream distance was -2708 base pairs and the maximum downstream distance was +5372 base pairs. The location of footprints across the ten studies from -1200 base pairs to the transcription initiation start site is displayed in a scatter plot (Figure 7.). The five prime location of footprints from MMTV, MOMSVLTR and chicken lysozyme (CLYSOZYM) DNA align approximately between -180 and -202 base pairs. MOMSVLTR footprints and the human metallothionein IIA footprint align from -265 to -271 base pairs. Otherwise there does not appear to be preferential footprint locations across these studies.

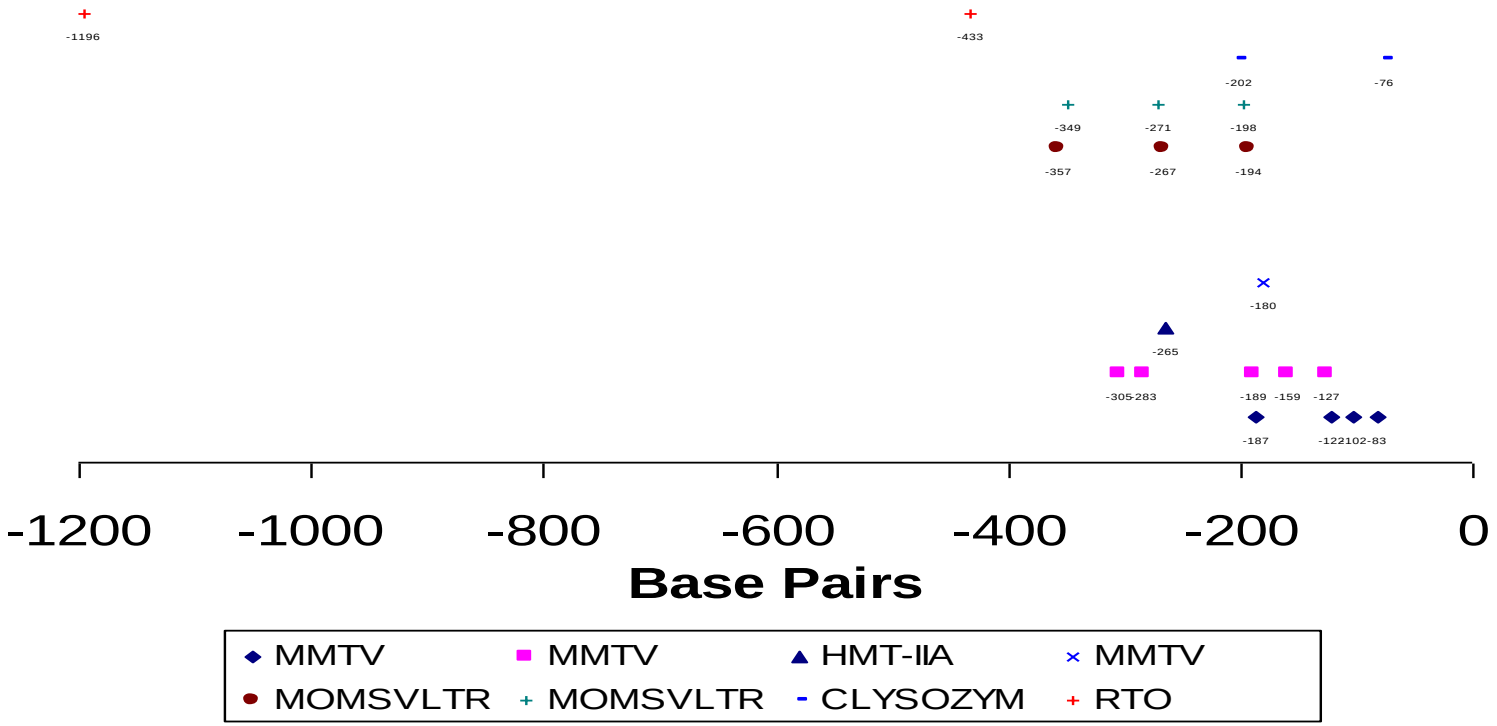

Figure 7. Upstream five prime GR footprint locations, -1200 base pairs to the transcription initiation start site (0).

CHAPTER FOUR: DISCUSSION

The results of this study are consistent with the hypothesis that experimentally derived NR DNA response elements do not contain direct repeats, inverted repeats or everted repeats. This study details the methods used to generate the database and initial findings of literature searches of NR – DNA interactions. Examples for six experimental techniques from original literature were presented as validation of the process to identify NR DNA binding sites and footprints. The molecular interactions between NR amino acid residues and DNA bases were summarized from X ray crystallographic studies. Experimentally derived NR footprints for the GR were examined for consensus response elements. Finally, the five prime footprint locations relative to the transcription initiation start site were compared to determine if preferential footprint locations exist.

Explanation of Outcomes

Twenty nine footprints from seven different DNA sources are consistent with the hypothesis that naturally occurring NR response elements are not direct repeats, inverted repeats or everted repeats. Many of the GR footprints examined contain a consensus half site, but no consensus responses elements were observed. NR response elements are more complicated than the literature portrays. For example, two recent reviews describe the GR response element as a palindrome or direct repeat (Aranda & Pascual 2001, Claessens & Gewirth 2004). In fact, Claessens & Gewirth list the direct repeat 5'-AGAACAnnnAGAACA-3' as the consensus sequence for the GR

response element. This heximer 5'-AGAACA-3' was used to search the 29 GR footprints and no direct repeats where found. NR binding to DNA is a function of multiple factors rather than symmetric palindromic response elements.

The replication of results from MMTV and MOMSVLTR DNA across studies was reassuring. This was especially evident in the identification in two studies of the same footprint in MOMSVLTR DNA without a consensus sequence. Even techniques with different sensitivities were able to identify at least one GR footprint in common in MMTV DNA.

The X ray crystallographic summary indicates that while the relative position and identity of DBD amino acid residues involved in NR DBD-DNA interactions is limited, the pattern of interaction can be variable. The RXR NR receptor binds to the same half site with four different patterns of interaction (Zhao et al. 2000). Non-DBD amino acid residues are also involved in DNA base specific interactions such as the tyrosine five base pairs upstream of the DBD found in almost every NR. DNA minor groove base specific contacts also contribute to NR – DNA binding. These factors in combination with protein-protein interactions between NRs probably account for the selectivity of NR-DNA interactions.

CHAPTER FIVE: CONCLUSION

The results of this study support the hypothesis that NR response elements are more complicated that simple direct repeats, inverted repeats or everted repeats and validate the need for a database of nuclear receptor DNA binding sites. The findings demonstrate variability in the molecular mechanisms of NR – DNA binding and the

composition of GR response elements. Application of similar methods to all experimentally derived NR response elements will provide the information necessary to identify NR regulatory elements in genomic DNA.

Limitations

The current results are based on experimentally derived DNA footprints from the GR alone. A more comprehensive examination of experimentally derived footprints for all NRs would provide a more accurate estimate of naturally occurring response elements. In addition, the X ray crystallographic studies did not utilize any experimentally derived response elements. These studies used either palindromic or direct repeats. The use of these response elements limits the application of the findings to the experimentally derived response elements for the GR receptor. Moreover, the one GR X ray crystallography study was flawed by an incorrect inter-half site spacing.

Future Research

For the future, the first priority is to compile more NR DNA footprints and binding sites. Sixty one publications have been identified from 1983 through 1994 containing NR DNA binding sites and/or footprints. These publications and other publications beyond 1994 will provide DNA binding site data for many of the NRs. Once a sufficient number of footprints and binding sites have been compiled computational methods will be used to determine the minimum DNA sequence

requirements for NR binding.

The database design will also be changed to provide additional information necessary to understand NR-DNA interactions. For example, the database may be expanded to include a summary of NR-DNA X ray crystallographic data. Other data fields may be added as additional data requirements become evident. The database currently has additional data fields than as originally designed (e.g. the five and three prime location of DNA footprints relative to the transcription initiation start site). The database will also be converted to a more powerful relation database management system and made publicly available.

Summary

A database has been developed for nuclear receptor (NR) DNA binding sites and the corresponding NR DNA binding domain. DNA binding sites and NR DNA footprints are obtained from original peer reviewed research publications. The purpose of the database is to provide a resource for the study of NR-DNA interactions.

REFERENCES

Achermann, J.C. & Jameson, J.L. (2003). Human Disorders Caused by Nuclear Receptor Gene Mutations. In *Pure Applied Chemistry* (vol. 75, no. 11, pp. 1785--1796).

Adams, M.D., et al. (2000). The Genome Sequence of Drosophila Melanogaster. In *Science* (vol. 287, pp. 2185-2195).

Aranda, A. & Pascual, A. (2001). Nuclear Hormone Receptors and Gene Expression. In *Physiological Reviews* (vol. 81, no. 3, pp. 1269--1304).

Branden, C. & Tooze, J. (1999). Chapter 10 Specific Transcription Factors Belong to a Few Families. In *Introduction to Protein Structure, Second Edition* (p. 182): Garland Publishing.

Cato, A.C.B., Geisse, S., Wenz, M., Westphal, H.M. & Beato, M. (1984). The Nucleotide Sequences Recognized by the Glucocorticoid Receptor in the Rabbit Uteroglobin Gene Region are Located Far Upstream from the Initiation of Transcription. In *The EMBO Journal* (vol. 3, no. 12, pp. 2771--2778).

Claessens, F. & Gewirth, D.T. (2004). DNA Recognition by Nuclear Receptors. In *Essays in Biochemisty* (vol. 40, pp. 59—72).

Danesch, U., Gloss, B., Schmid, W., Schutz, G., Schule, R. and Renkawitz, R. (1987). Glucocorticoid Induction of the Rat Tryptophan Oxygenase Gene is Mediated by Two Widely Separated Glucocorticoid-Reponsive Elements. In *The EMBO Journal* (vol. 6, no. 3, pp. 625--630).

DeFranco, D. & Yamamoto, K.R. (1986). Two Different Factors Act Separately

or Together to Specify Functionally Distinct Activities at a Single Transcriptional Enhancer. In *Molecular and Cellular Biology* (vol. 6. no. 4, pp. 993—1001).

Deroo, B.J. & Korach, K.S. (2006). Estrogen Receptors and Human Disease. In *The Journal of Clinical Investigation* (vol. 116, no. 3, pp. 561—570).

Duarte, J., Perriere, G., Laudet, V. & Robinson-Rechavi, M. (2002). NUREBASE: Database of Nuclear Hormone Receptors. In *Nucleic Acids Research* (vol. 30, no. 1, pp. 364—368).

Evans, R.M. (1988). The Steroid and Thyroid Hormone Receptor Superfamily. In *Science* (vol. 240, pp. 889—895).

Fasel, N., Pearson, K., Buetti, E. & Digglemann, H. (1982). The Region of Mouse Mammary Tumor Virus DNA Containing the Long Terminal Repeat Includes a Long Coding Sequence and Signals for Hormonally Regulated Transcription. In *The EMBO Journal* (vol. 1, no. 1, pp. 3--7).

Freedman, L.P. & Luisi, B.F. (1993). On the Mechanism of DNA Binding by Nuclear Hormone Receptors: A Structural and Functional Perspective. In *Journal of Cellular Biochemistry* (vol. 51, pp. 140-150).

Geisse, S., Scheidereit, C., Westphal, H.M., Hynes, N.E., Groner, B. & Beato, M. (1982). Glucocorticoid Receptors Recognize DNA Sequences In and Around Murine Mammary Tumour Virus DNA. In *The EMBO Journal* (vol. 1, pp. 1613--1619).

Glass, C.K. (1994). Differential Recognition of Target Genes by Nuclear Receptor Monomers, Dimers, and Heterodimers. In *Endocrine Reviews* (vol. 15, no.

3, pp. 391—407).

Govindan, M.V., Spiess, E. & Majors, J. (1982), Puridied Glucocorticoid Receptor-Hormone Complex from Rat Liver Cytosol Binds Specifically to Cloned Mouse Manaary Tumor Virus Long Terminal Repeats *in Vitro*. In *Proceedings of the National Academy of Sciences of the United States of America* (vol. 79, pp. 5157--5161).

Gorski, J., Toft, D., Shyamala, G., Smith, D., & Notides, A. (1968). Hormone Receptors: Studies on the Interaction of Estrogen with the Uterus. In *Recent Progress in Hormone Research* (vol. 24, pp. 45--80).

Green, S., & Chanbon, P. (1991). The Oestrogen Receptor: From Perception to Mechanism. In *Nuclear Hormone Receptors* (pp. 15--38): Academic Press.

Guille, M.J. & Kneale, G.G. (1997). Methods for the Analysis of DNA-Protein Interactions. In *Molecular Biotechnology* (vol. 8, pp. 35—52).

Horn, F., Vriend, G. & Cohen, F.E. (2001). Collecting and Harvesting Biological Data:The GPCRDB and NucleaRDB Information Systems. In *Nucleic Acids Research* (vol. 29, no. 1, pp. 346—349).

Jensen, E.V. & DeSombre, E.R. (1973). Estrogen-Receptor Interaction. In *Science* (vol. 182, pp. 126-134).

Jensen, E.V. & Jordan, V.C. (2003). The Estrogen Receptor: A Model for Molecular Medicine. In *Clinical Cancer Research* (vol. 9, pp. 1980—1989).

Jensen, E.V., Suzuki, T., Kawashina, T., Stumpf, W.E., Jungblut, P.W., & DeSombre, E.R. (1968). A Two-Step Mechanism for the Interaction of Estradiol with

Rat Uterus. In *Proceedings of the National Academy of Sciences of the United States of America* (vol. 59, pp. 632--638).

Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H. M., & Beato, M. (1984). Characterization of DNA Sequences through which Cadmium and Glucocorticoid Hormones Induce Human Metallothionein-II_A Gene. In *Nature* (vol. 308, pp. 513—519).

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. & Sigler, P.B. (1991). Crystallographic Analysis of the Interaction of the Glucocorticoid Receptor with DNA. In *Nature* (vol. 352, pp. 497—505).

Malo, M.S., Pushpakaran, P. & Hodin, R.A. (2005). A 'Swinging Cradle' Model for in Vitro Classification of Different Types of Response Elements of a Nuclear Receptor. In *Biochemical and Biophysical Research Communications* (vol. 337, pp. 490—497).

Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. & Evans, R. M. (1995). The Nuclear Receptor Superfamily: The Second Decade. In *Cell* (vol. 83, pp. 835—839).

Martinez, E., Moore, D.D., Keller, E., Pearce, D., Vanden Heuvel, J.P., Robinson, V., Gottlieb, B., MacDonald, P., Simons, S., Jr., Sanchez, E. & Danielsen, M. (1998). The Nuclear Receptor Resource: A Growing Family. In *Nucleic Acids Research* (vol. 26, no. 1, pp. 239—241).

McGrath, C. M. (1971). Replication of Mammary Tumor Virus in Tumor Cell Cultures: Dependence on Hormone-Induced Cellular Organization. In *Journal of the*

National Cancer Institute (vol. 47 pp. 455--467).

Meinke, G. & Sigler, P.B. (1999). DNA-Binding Mechanism of the Monomeric Orphan Nuclear Receptor NGFI-B. In *Nature Structural Biology* (vol. 6, no. 5, pp. 471—477).

Miksicek, R., Heber, A., Schmid, W., Danesch, U., Posseckert, G., Beato, M. & Schutz, G. (1986). Glucocorticoid Responsiveness of the Transcriptional Enhancer of Moloney Murine Sarcoma Virus. In *Cell* (vol. 46, pp. 283—290).

Nuclear Receptors Nomenclature Committee (1999), A Unified Nomenclature System for the Nuclear Receptor Superfamily. In *Cell* (vol. 97, pp 161-163).

Payvar, F., Wrange, O., Carlstedt-Duke, J., Okret, S., Gustafsson, J-A. & Yamamoto, K. R. (1981). Purified Glucocorticoid Receptors Bind Selectively in Vitro to a Cloned DNA Fragment Whose Transcription is Regulated by Glucocorticoids in Vivo. In *Proceedings of the National Academy of Sciences of the United States of America* (vol. 78, pp. 6628-6632).

Payvar, F., DeFranco, D., Firestone, G.L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J-A. & Yamamoto, K.R. (1983). Sequence-Specific Binding of Glucocorticoid Receptor to MTV DNA at Sites within and Upstream of the Transcribed Region. In *Cell* (vol. 35, pp. 381—392).

Peterkofsky, B. & Tomkins, G.M. (1968). Evidence for the Steroid-Induced Accumulation of Tyrosine-Aminotransferase Messenger RNA in the Absence of Protein Synthesis. In *Proceedings of the National Academy of Sciences of the United States of America* (vol. 60, pp. 222-228).

Pfahl, M. (1982). Specific Binding of the Glucocorticoid-Receptor Complex to the Mouse Mammary Tumor Proviral Promoter Region. In *Cell* (vol. 31, pp. 475-482).

Rastinejad, F., Perlman, T., Evans, R.M. & Sigler, P.B. (1995). Structural Determinants of Nuclear Receptor Assembly on DNA Direct Repeats. In *Nature* (vol. 375, pp. 203—211).

Rastinejad, F., Wagner, T., Zhao, Q. & Khorasanizadeh, S. (2000). Structure of the RXR-RAR DNA-Binding Complex on the Retinoic Acid Response Element DR1. In *EMBO Journal* (vol.19, no. 5, pp. 1045—1054).

Ringold, G. M. (1985). Steroid Hormone Regulation of Gene Expression. In *Annual Review of Pharmacology and Toxicology* (vol. 25 pp. 529--566).

Robins, D.M., Paek, I., Seeburg, P.H. & Axel, R. (1982). Regulated Expression of Human Growth Hormone Genes in Mouse Cells. In *Cell* (vol. 29, no. 2, pp. 623—631).

Robinson-Rechavi, M., Carpenter, A-S., Duffraisse, M. & Laudet, V. (2001). How Many Nuclear Hormone Receptors are there in the Human Genome. In *Trends in Genetics* (vol. 3, no. 10, pp. 554—556).

Scheidereit, C., Geisse, S., Westphal, H.M. & Beato, M. (1983). The Glucocorticoid Receptor Binds to Defined Nucleotide Sequences Near the Promoter of Mouse Mammary Tumour Virus. In *Nature* (vol. 304, pp. 749--752).

Scheidereit, C. & Beato, M. (1984). Contacts between Hormone Receptor and DNA Double Helix within a Glucocorticoid Regulatory Element of Mouse Mammary

Tumor Virus. In *Proceedings of the National Academy of Sciences of the United States of America* (vol. 81, pp. 3029-3033).

Schmid, W., Scherer, G., Danesch, U., Zentgraf, H., Matthias, P., Strange, C.M., Rowekamp, W. & Schutz, G. (1982). Isolation and Characterization of the Rat Tryptophan Oxygenase Gene. In *The EMBO Journal* (vol. 1, no. 10, pp. 1287—1293).

Schwabe, J.W.R., Chapman, L., Finch, J.T. & Rhodes, D. (1993). The Crystal Structure of the Estrogen Receptor DNA-Binding Domain Bound to DNA: How Receptors Discriminate between their Response Elements. In *Cell* (vol. 75, pp. 567—578).

Seiler-Tuyns, A., Walker, P., Martinez, E., Merillat, A-M., Givel, F. & Wahli, W. (1986). Identification of Estrogen-Responsive DNA Sequences by Transient Expression Experiments in a Human Breast Cancer Cell Line. In Nucleic Acids Research (vol. 14, no. 22, pp. 8755—8770).

Shaw, P.E. & Stewart, A.F. (2001). Identification of Protein-DNA Contacts with Dimethyl Sulfate. In *Methods in Molecular Biology* (vol. 148, pp. 221—227).

Singleton, P. & Sainsbury, D. (2006). Coding Strand. In *Dictonary of Microbiology and Molecular Biology, Third Edition Revised.* (p. 177): John Wiley & Sons Ltd.

Slater, E.P., Rabenau, O., Karin, M., Baxter, J.D. & Beato, M. (1985). Glucocorticoid Receptor Binding and Activation of a Heterologous Promoter by Dexamethasone by the First Intron of the Human Growth Hormone Gene. In

Molecular and Cellular Biology (vol. 5, no. 11, pp. 2984—2992).

Sluder, A.E. & Maina,C.V. (2001). Nuclear Receptors in Nematodes: Themes and Variations. In *Trends in Genetics* (vol. 17, no. 4, pp. 206-213).

Tata, J. R. (2002), Signalling through Nuclear Receptors. In *Nature Reviews Molecular Cell Biology* (vol. 3, pp. 702--710).

Tenbaum, S. & Baniahmad, A. (1997) Nuclear Receptors: Structure, Function and Involvement in Disease. In *International Journal of Biochemistry and Cell Biology* (vol. 29, no. 12, pp. 1325—1341).

Toft, D., & Gorski, M. A. (1966). A Receptor Molecule for Estrogens: Isolation from the Rat Uterus and Preliminary Characterization. In *Proceedings of the National Academy of Sciences of the United States of America* (vol. 55, pp. 1574--1581).

Umesono, K. & Evans, R. M. (1989). Determinants of Target Gene Specificity for Steroid/Thyroid Hormone Receptors. In *Cell* (vol. 57, pp. 1139—1146).

von der Ahe, D., Janich, S., Scheidereit, C., Renkawitz, R., Schutz, G. & Beato, M. (1985). Glucocorticoid and Progesterone Receptors Bind to the Same Sites in Two Hormonally Regulated Promoters. In *Nature* (vol. 313, pp. 706-709).

von der Ahe, D., Renoir, J-M., Buchou, T., Baulieu, E-E., & Beato, M. (1986). Receptors for Glucocorticosteroid and Progesterone Recognize Distinct Features of a DNA Regulatory Element. In *Proceedings of the National Academy of Sciences of the United States of America* (vol. 83, pp. 2817-2821).

Wasserman, W.W. & Sandelin, A. (2004). Applied Bioinformatics for the Identification of Regulatory Elements. In *Nature Reviews Genetics* (vol. 5, pp. 276—

287).

Westphal, H.M. & Beato, M. (1980). The Activated Glucocorticoid Receptor of Rat Liver. In *European Journal of Biochemistry* (vol. 106, pp. 395—403).

Wrange, O., Carlstedt-Duke, J. & Gustafsson, J.-A. (1979). Purification of the Glucocorticoid Receptor from Rat Liver Cytosol. In *Journal of Biological Chemistry* (vol. 254, no. 18, pp. 9284—9290).

Zhang, Z., Burch, P.E., Cooney, A.J., Lanz, R.B., Pereira, F.A., Wu, J., Gibbs, R.A., Weinstock, G. & Wheeler, D.A. (2004). Genomic Analysis of the Nuclear Receptor Family: New Insights into Structure, Regulation and Evolution from the Rat Genome. In *Genome Research* (vol. 14, pp. 580-590).

Yamamoto, K.R. (1985). Steroid Receptor Regulated Transcription of Specific Genes and Gene Networks. In *Annual Review of Genetics* (vol. 19, pp. 209—252).

Zhao, Q., Khorasanizadeh, S., Miyoshi, Y., Lazar, M.A. & Rastinejad, F. (1998). Structural Elements of an Orphan Nuclear Receptor-DNA Complex. In *Molecular Cell* (vol. 1, pp. 849—861).

Zhao, Q., Chasse, S.A., Devarakonda, S., Sierk, M.L., Ahvazi, B. & Rastinejad, F. (2000). Structural Basis of RXR-DNA Interactions. In *Journal of Molecular Biology* (vol. 296, pp. 509—520).

APPENDICES

Appendix A: Database Table Structures

Table: NR- DNA BINDSITE

Columns

Name	Type	Size
Journal	Text	30
Volume	Integer	2
FirstPage	Integer	2
Year	Integer	2
BindSite	Text	15
DNASourc	Text	10
InterSite	Integer	2
Method	Text	25
NRID	Text	10
Authors	Text	125

Table: NR-DNA FOOTPRINT

Columns

Name	Type	Size
Journal	Text	30
Volume	Integer	2
FirstPage	Integer	2
Year	Integer	2
FivePrime	Integer	2
ThrePrime	Integer	2
Footprint	Text	50
DNASourc	Text	10
NRID	Text	10

Table: NR-PROTEIN

Columns

Name	Type	Size
NRID	Text	8
DESCRIP	Text	24
NRSource	Text	50
SWISPROT	Text	6
AASEQRH	Text	75

Appendix A: Database Table Structures

Table: DNA Source

Columns

Name	Type	Size
DNASourc	Text	10
DESCRIP	Text	50

Table: DNA Methods

Columns

Name	Type	Size
Method	Text	16
DESCRIP	Text	50

Curriculum Vitae

Personal Data

Name: Timothy E. Breen

Date of Birth: May 16, 1951

Home Address: 6125 Maren Drive
Indianapolis, IN 46224

Phone: (317) 291-1535

Military Service: Branch - United States Army
Period of Service - June 1971 to November 1974
Discharge - Honorable
Rank at Discharge - Specialist 5th Class

Education

B.S., Psychology, Purdue University, Indianapolis, 1978
M.S., Physiology, Indiana University, Indianapolis, 1986
Ph.D., Physiology, Indiana University, Indianapolis, 1991
Current – M.S., Bioinformatics, Indiana University, Indianapolis

Current Position

Biostatistics Project Coordinator
Division of Biostatistics
Department of Medicine
Indiana University School of Medicine
1050 Wishard Boulevard, RG4101
Indianapolis, IN 46202-2872

Supervisor – Beverly Musick, M.S.

Phone: (317) 274-2693

Professional Experience

1999-2003 Data Manager II
Division of Biostatistics
Department of Medicine
Indiana University School of Medicine

1997-1999 Data Manager I
Division of Biostatistics
Department of Medicine
Indiana University School of Medicine

1995-1997 Histocompatibility Technologist
Transplant Immunology Laboratory
Department of Surgery
Indiana University School of Medicine

1994-1995 Research Associate
Department of Medicine
Institute of Psychiatric Research
Indiana University School of Medicine

1990-1994 Postdoctoral Fellow
Department of Medicine
Institute of Psychiatric Research
Indiana University School of Medicine

1987-1990 Graduate Research Assistant
Department of Physiology and Biophysics
Indiana University School of Medicine

1986-1987 Senior Research Technician
Department of Medicine
Krannert Institute of Cardiology
Indiana University School of Medicine

1981 - 1986 Graduate Research Assistant
Department of Physiology and Biophysics
Indiana University School of Medicine

1978 - 1981 Statistician IV
Indiana Department of Mental Health

Awards and Honors

Sigma Pi Alpha Scholastic Honor Society - 1977
School of Science Dean's List - 1975, 1976, 1977
Student Government Senator of the Year Award –1979
Circle City Circuit Outstanding Service Award -1978
Doctoral Student Grant-In-Aid of Research -1985
Predoctoral Fellowship American Heart Association, Indiana Affiliate- 1988-89,1989-90.
Postdoctoral Fellowship, National Institute on Alcohol Abuse and Alcoholism - 1990-94
National Institute on Alcohol Abuse and Alcoholism/ Research Society on Alcoholism Junior Meeting Investigator Award – 1995

Membership in Professional Organizations

International Society for Computational Biology

Research Interest

Identification of Eukaryotic Nuclear Receptor DNA Binding Sites

Teaching Experience

Department of Physiology and Biophysics, Indiana University School of Medicine 1982 - 1986
Lecturer Cell Biology and Neurophysiology - Human Physiology, Physical Therapy Program
Laboratory Instructor - Medical Physiology
Laboratory Instructor - Dental Physiology
Seminar Series - Molecular Biology and Biophysics of Ion Channels

Extracurricular Activities

Certified Baseball and Softball IHSAA Umpire – 1997- present

Publications

1. Ochs,S., Jersild, R.A. Jr., Breen, T., Peterson, R.
Comparison of Calcium and Strontium Sequestration in Nerve Axons in Relation to Axoplasmic Transport.
Soc. Neurosci. Abst. 9:149,1983.

2. Ochs, S., Jersild, R.A. Jr., Breen, T., Morey, K., McKltrick, L.
The Maintenance of Axoplasmic Transport in Nerves by Strontium and Its Localization in Nerve Fibers.
J. Neurobiol. 17(l):55-61, 1985.

3. Breen, T.E., Ochs, S.
Longitudinal Loss of Excitability in Peripheral Nerves.
Soc. Neurosci. Indianapolis Chapter 3:1, 1986.

4. Breen, T.E., Pressler, M.L.
Alpha-Adrenergic Stimulation and Phorbol Esters Alter Intracellular pH in Cardiac Purkinje Fibers.
Clin. Res. Abst. 36(3):266A, 1988.

5. Breen, T.E., Pressler, M.L.
Modulation of Contractility by Alpha Agonists and Phorbol Esters in Cardiac Purkinle Fibers.
Clin. Res. Abst. 36(6):816A, 1988.

6. Pressler, M.L., Lovelace, D.E., Breen, T.E.
Analysis of Amiloride-Sensitive H^+ Efflux During Alpha-Adrenoreceptor Stimulation of Cardiac Purkinje Fibers.
Biophys. J. Abst. 55:392A, 1989.

7. Lovelace, D.E., Breen, T.E., Pressler, M.L.
A Mathematical Model for pH Regulation in Purkinje Fibers.
Blophys. J. Abst. 57:193A, 1990.

8. Breen, T.E.,
Alpha-1 Adrenergic Receptors, Protien Kinase C, and Regulation of Intracellular pH in Cardiac Purkinje Fibers.
Indianapolis, Indiana: Indiana University.
Dissertation, 1990.

9. Morzorati, S., Breen, T.E., Lumeng, L., Li, T.-K.
Is There a Relationship Between Peak Theta Frequency and Ethanol Preference.
Alcohol Clin. Exp. Res. Abst. 17:474,1993.

10. Breen, T.E., Morzorati, S., Lumeng, L., Li, T.-K.
Medial Septal Single Cell Activity and The Hippocampal Theta Rhythm in Ambulating P and NP Rats.
Alcohol Clin. Exp. Res. Abst. 17:474,1993.

11. Morzorati, S., Breen, T.E., Lumeng, L., Li, T.-K.
Comparison of Innate EEG Parameters in Rat Lines Selected for Ethanol Preference
Alcohol 11(3):253-258,1994.

12. Morzorati, S., Breen, T.E.
Intragastric Ethanol Lowers the Peak Frequency of Type I Theta.
Soc. Neurosci. Abst. 20(2):1617,1994.

13. Duemler, S., June, H.L., Breen, T.E., Lewis, M.J., Murphy, J.M.
The Benzodiazepine Inverse Agonist R019-4603 Exerts Prolonged Antagonism of the Locomotor Depressant Actions of Ethanol in Alcohol-Preferring (P) Rats.
Soc. Neurosci. Abst. 20(l):499,1994.

14. Breen, T.E., Lumeng, L., Morzorati, S.L.
A Comparison of Medial Septal Area Burst Activity in Ambulating P and NP Rats.
Alcohol Clin. Exp. Res. Abst. 19(2):61A, 1995.

15. Breen, T.E., Morzorati, S.L.
Innate Differences in Medial Septal Area Burst Firing Neurons and the Hippocampal Theta Rhythm During Ambulation in Selectively Bred Rat Lines.
Brain Res. 714:156-64,1996.

16. Breen, T.E., Katz, B.P., Fortenberry, J.D., Orr, D.P.
A Data Management System for Multidisciplinary Adolescent STD Research.
Int. J. STD AIDS. Abst. 12(supp 2):198, 2001.

17. Khan, A., Van der Pol, B., Breen, T.E., Fortenberry, J.D., Orr, D.P., Batteiger, B.E., Sexually Transmitted Infections Frequently Differ Between Partners in Sexual Dyads.
Abst. ISSTDR 2003.

18. Orr, D.P., Fortenberry, J.D., Batteiger, B.E., Breen, T., Katz, B.P., Shier, L., Braslins, P., Schillinger, J.
Event-Specific Correct Condom Use.
Abst. ISSTDR 2003.

19. Blank, S., Chapin, J., Hutchins, K., Braslins, P., Shrier, L., Madico, G.E., Rice, P.A., Van Der Pol, B.J., Breen, T., Katz, B.P., Orr, D., Batteiger, B.E., Papp, J.R., Markowitz, L.E.
Transmission of Chlamydia Trachomatis Between Heterosexual Sex Partners; Preliminary Results from a Concordance Study.
Abst. ISSTDR 2003.

20. Madico, G.E., Rice, P.A., Batteiger, B.E., Van der Pol, B.J., Braslins, P., Shrier, L., Breen, T.

**Trichomonas Vaginalis Infection in Sexual Partners.
Abst. ISSTDR 2003.**

21. **Batteiger, B.E., Fortenberry, J.D., Breen, T., Madico, G.E., Stothard, D.R.
Molecular Epidemiology and Transmission of Chlamydia Trachomatis Infection Based on omp1 Genotyping within Sexual Dyads.
Abst. ISSTDR 2003.**

22. **Brown,D.R., Shew,M.L., Qadadri,B., Neptune,N., Vargas,M., Tu,W., Juliar, B.E., Breen,T.E., Fortenberry,J.D.
A Longitudinal Study of Genital Human Papillomavirus Infection in a Cohort of Closely Followed Adolescent Women.
J. Infect. Dis. In press.**