REVIEW ARTICLE

The *vnd/NK-2* gene from *Drosophila melanogaster*: How relatively small molecular changes can lead to catastrophic genetic anomalies

Kae-Jung Hwang[1] and James A. Ferretti[2*]

Affiliations:

[1] Bruker Biospin Corp., 15 Fortune Drive, Billerica, MA 01821-3991 USA, E-mail: Kaejung.Hwang@bruker.com

[2] Laboratory of Structural Biophysics, National Heart, Lung & Blood Institute, National Institutes of Health, 50 South Drive MSC-8013, Bethesda, MD 20892-8013, USA, E-mail: ferrettj@mail.nih.gov

*** Corresponding author:** James A. Ferretti, E-mail: ferrettj@mail.nih.gov

Abstract

We summarize our knowledge on the *NK-2* class of homeobox-containing genes and their encoded proteins. These genes are characterized by the presence of a 180 base-pair segment of DNA called the homeobox. The corresponding 60 amino acid residue fragment of the encoded protein is called the homeodomain. The *NK-2* class of homeobox genes is defined by the encoding of tyrosine in position 54 of the homeodomain, which is responsible for the recognition of the unusual 5' – CAAGTG – 3' core consensus DNA sequence. We review structural results on the NK-2 homeodomains both in the free and DNA-bound states, thermodynamic properties and CAT assays of the wild type and selected single amino acid residue replacements. We demonstrate that a tyrosine to methionine mutation in the vnd/NK-2 homeodomain and a mutation from tyrosine to cysteine in the highly homologous human cardiac CSX/NKX-2.5 homeodomain do not alter the structures of their respective homeodomain-DNA complexes. Our transgenic data on a mutant gene that encodes for methionine in *vnd/NK-2*, a CNS gene, is lethal. Both the tyrosine to cysteine mutant in position 54 of the CSX/NKX-2.5 homeodomain that arises from a missense *UAC* to *UGC* mRNA codon change and a synonymous tyrosine to tyrosine change in position 14 from the rare *UAU* to the common *UAC* mRNA codon, also in the CSX/NKX-2.5 homeodomain, result in serious congenital heart defects. We have investigated potential roles of the mRNA in *vnd/NK-2* by carrying out preliminary transient transfection assays using RNA Affymetrix chip assay data to show that the altered mRNAs do not properly repress known downstream target genes. We suggest that alterations in the mRNA, as well as changes in individual amino acid residues in a protein, apparently can act as etiological agents to generate phenotypic alterations or genetic diseases in humans.

Keywords: Structure, Function, NMR, mRNA, Transgenic, Synonymous, Mutation, Disease

Introduction

The scourge of genetic diseases is one that scientists and clinicians hope will be resolved by the end of the 21st century. Any change (i.e., a mutation) in the DNA sequence of any gene might result in significant cellular alterations thereby damaging the corresponding organism and giving rise to a genetic defect. Mutations

even in regions of DNA outside a gene, namely in the so-called "junk" DNA [1], also can result in genetic disorders, although we do not specifically address this question here. Given that the genome of an organism can be made of many thousands of genes, one quickly realizes that the problem is enormously complicated and solutions (i.e., cures or control of the corresponding disorders) will require a superhuman effort. To reach the goal of at least managing if not curing genetic disorders, it is first necessary to break down the related scientific questions into separate but numerous individual parts. The goal here is to understand the functional consequences of any mutation by using a myriad of physical and biological techniques that can include structural studies on the corresponding proteins and mRNA, thermodynamic investigations, transgenic experiments on model systems and various cellular assays. We must recognize that the cascade of events starting with activation of a given gene at any stage in the life cycle is complex. Understanding what, how and why mutations in the DNA occur and the resulting effects on gene regulation is necessary. One hopefully manageable but restrictive approach involves selecting a single "interesting" gene, trying to understand its mechanism of activation and then examining the effects of individual mutations. Just the thought that humans have approximately 20000 genes and that spontaneous errors occur on average about every 1000 bases makes one realize the enormity of the problem [2]. At best one can hope to address the problem by studying a single gene or family of genes using a range of methods and this is the approach we have adopted.

The genes that have interested us for more than 25 years are the homeotic genes [3]. What are homeotic genes and why choose them? These genes were first characterized when it was realized that they were responsible for the replacement of the antennae of Drosophila melanogaster, the common fruit fly, with a second set of legs; i.e., antennapedia [3]. These genes are phenomenally interesting because they are master controlling genes that control the pattern of development of all organs in organisms ranging from single cell ones all the way up to and including humans (Homo Sapiens). To study these genes, one must first formulate relevant questions and then develop working hypotheses that can be tested. Our initial thoughts were that having the structure as well as other related biophysical properties of the homeodomain in both the free and receptor-bound states would permit us to better understand how the protein functions, how and why mutations alter function, and to design appropriate subsequent experiments.

In this review, we start by describing the structural, thermodynamic and binding experiments that we have performed both on wild type and mutant encoded proteins from these genes. We compliment these biophysical measurements with functional studies designed to tell us how any structural or thermodynamic changes that are observed might influence the ability of these genes or gene products to function properly. Based on the results we obtained, we carried out transgenic studies on wild type and mutant genes from *D. melanogaster.* [4] The transgenic experiments we have performed show how what might appear to be small molecular changes can lead to dramatic functional effects. [5] To compliment these transgenic results, we performed CAT (chloramphenicol acetyl transferase) assays using both the wild type and mutant vnd/NK-2 genes to study a segment of the gene that is found 5' – upstream from the coding region and is believed to be involved in auto-regulation (Nirenberg, M. private communication). We carried out transient transfection activation/repression studies on downstream genes that are targeted by: 1.

the wild type *vnd/NK-2* messenger RNA (mRNA), 2. the mRNA (*AUG*) that codes for M in position 54 rather than Y and 3. the mRNA that uses the rare *UAU* rather than the common *UAC* wild type codon to produce a synonymous mutation. The amino acid residue in position 54 is responsible for the binding of the homeodomain to the unusual 5' – CAAGTG – 3' DNA consensus sequence. The change in position 54 in the vnd/NK-2 homeodomain is analogous to a mutation found in the human CSX/NKX2.5 protein that results in congenital heart defects and will be described below. Some of the results we present in this review are part of our past studies on the *D. melanogaster vnd/NK-2* gene [6]. We present these studies along with some new results and further interpretations and conjecture.

The *vnd/NK-2* gene is responsible for the development of part of the central nervous system of the fruit fly. Furthermore, we compare our results on another gene in the *NK-2* class, namely human *CSX/NKX-2.5* [7] because of the high sequence homology of both the homeobox and the homeodomain. Thus, in effect we use the D. melanogaster *vnd/NK-2* gene as a model to help provide possible molecular-level explanations on how a single nucleotide polymorphism (SNP) could result [8] in congenital malformations found in the hearts of newborn babies. In addition, we have carried out various functional assays to evaluate the behavior of both the wild type and mutant proteins encoded from the respective *vnd/NK-2* genes. There is further a historical aspect to the study on the *Drosophila* analog in that these experiments were initiated before the human gene had been sequenced and before a mutation in position 54 of the human CSX/NKX-2.5 homeodomain had been reported [9]. A holistic appraisal of these together with other reported results [10] have lead us to question whether the protein is necessarily the etiological agent of the genetic disorder. We describe our rationalization in presenting this question along with some preliminary results of the hypothesis that the role of mRNA in the etiology of genetic disease and phenotypic alterations might be underappreciated.

Homeobox Genes

Homeotic genes are master controlling genes that regulate the development of the body plan in all organs [3,8,11,12]. A subset of the homeotic genes are ones that contain a homeobox, which is a 180 base-pair segment that is conserved throughout evolution in all species. Master controlling genes are very high up in the cascade of events that leads to embryonic development as well as continued development and control of function in adults [6,13]. They regulate development both at the embryonic as well as at the adult stage of life by activating and deactivating other genes both in time and in position. A simple example to visualize is to look at our own bodies and ask why all body parts seem to be appropriately placed. This observation appears to hold in all organisms. Each of us has two arms, two hands, two legs, five fingers on each hand and five toes on each foot. When we consider that each of us started out as amorphous cells at fertilization, it becomes natural to ask why, with few exceptions, our body plans are properly maintained from embryogenesis through adulthood.

Examples of the *NK* family of homeobox genes that are of high clinical interest are ones that are expressed in the heart, lungs, prostate and thyroid as well as part of the CNS. [14,15] Mutations in a specific homeobox gene found in the heart, i.e., *CSX/NKX-2.5* [16] or cardiac specific homeobox gene, lead to such anomalies as atrial septum defect and tetralogy of Fallot [17,18,19]. The *NKX-3.1* gene is supposed to play a role in the development of prostate

cancer [20,21]. Because the homeobox and, thus, its encoded homeodomain is highly conserved across all species, it is possible to investigate many of their structural and functional features at both the molecular and cellular level on systems from lower species, such as the *vnd/NK-2* gene from *D. melanogaster*. As in many such biochemical and biophysical studies, one is asked to accept the generalization of some of the conclusions drawn from studies in these lower species to humans. The molecular and cellular processes that follow protein encoding in the ribosome show many similarities across all species. We will further attempt to apply such generalizations to both single base changes in the homeobox and corresponding single amino acid residue replacements in the homeodomain. The fundamental question is what do we learn from knowing the structures of the wild type and mutant homeodomains in the free and DNA bound states, having the thermodynamic parameters for the folding and unfolding of the protein, as well as from the binding to DNA, from carrying out transgenic studies and from various assays? Having the three-dimensional structure means that we know the position of each atom in the molecule relative to all other atoms both in the molecule and of other molecules that interact at the atomic level.

These thoughts lead us to study the three-dimensional structure of a fragment of a protein that contained a homeodomain, the binding of the protein to DNA, the transgenic experiments, and ultimately to examine the role of mutations in the corresponding mRNA developed over many years. We will attempt to clarify our thoughts throughout this manuscript. The protein studied is the vnd/NK-2 homeo-domain from *D. melanogaster* (Fig. 1).

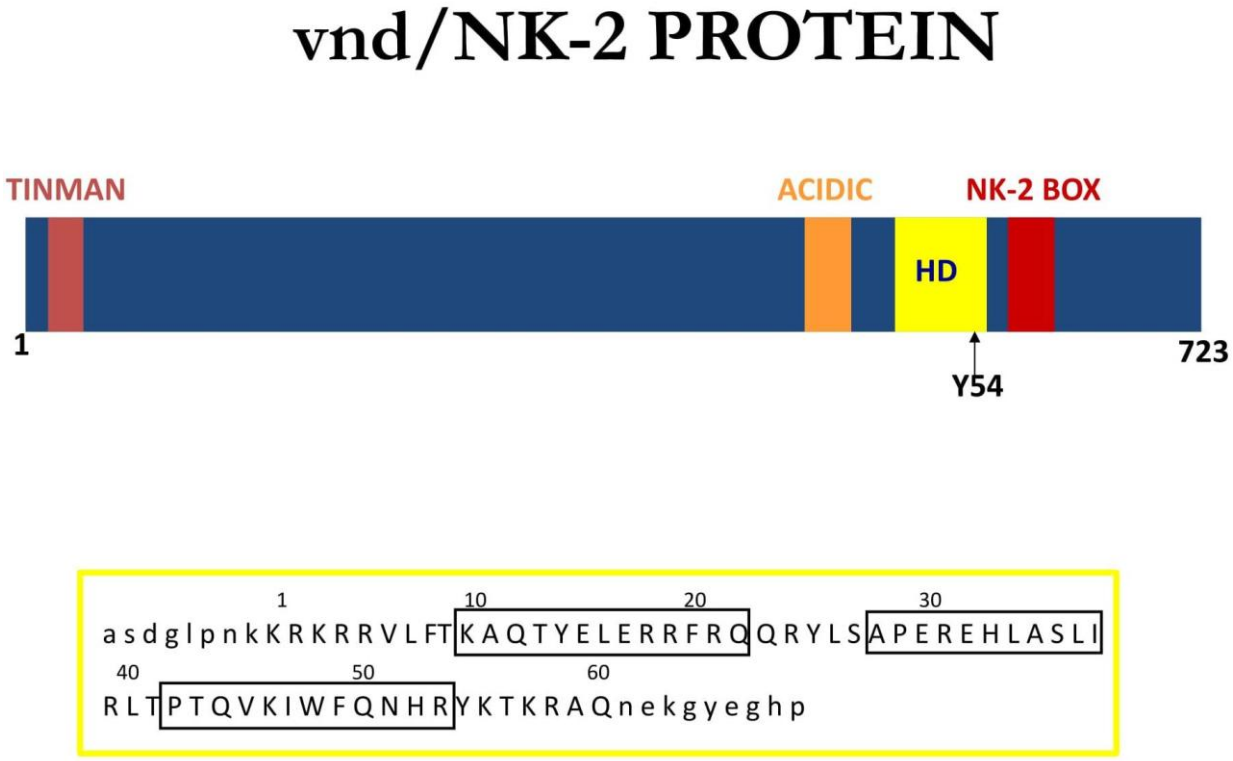

Figure 1. Block diagram of the 723 amino-acid residue vnd/NK-2 protein. The various colors depict the various conserved region of the protein. The position of Y54 is shown with the black arrow and the sequence of the homeodomain and the various helical segments are shown below the block diagram inside the yellow rectangle.

Much of this work started before we had any appreciation of the genetics of the *vnd/NK-2* homeobox gene, a subject that will be discussed later in this review. After obtaining the structure of the free homeodomain containing protein, we then went on to study the protein bound to a consensus 16-base sequence of DNA that contained 5' – CAAGTG – 3' as its core (Fig. 2). This consensus sequence differs from the so-called canonical 5' – TAATGG – 3' sequence. This consensus core sequence was used exclusively in subsequent studies because it was found to be the sequence that had the highest affinity for the vnd/NK-2 homeodomain.

The term "homeobox" was first used by Walter Gehring when he and his group reported the isolation and characterization of the *Antennapedia* gene from *Drosophila melanogaster* (the common fruit fly). The *Antennapedia* gene, when heat activated [22] in *D. melanogaster*, produces fruit flies where the antennae are replaced by a second set of legs; i.e., a homeotic mutation. Since the isolation and identification of the *Antennapedia* gene, many other homeobox genes have been identified in numerous organs in various organisms from the fruit fly through humans. It is fair to say that the discovery and isolation of the homeobox rates near the top of list of significant genetic findings of the 20th century.

The homeodomain acts as a regulator of transcription, where proper regulation depends to an important extent on this homeodomain-DNA interaction in addition to other protein-protein interactions involving other parts of the full-length homeoprotein [15]. Numerous mutations in homeobox genes both in humans and in lower species have been identified that lead to genetic disease or phenotypic alterations and currently are being studied in various laboratories. Already here one must be more restrictive both with respect to the research and to the breadth of this review. Some of these mutants involve deletion of segments of the DNA, others involve frame shifts and some are single nucleotide polymorphisms (SNPs) [10]. We have chosen to study primarily mutations that involve SNPs and, thus, this review will reflect that choice. Some object to the term polymorphism because, in the past, it has been used to denote base changes that were believed to have no effect (i.e., silent mutations). However, the concept of a "silent mutation" is in general no longer accepted. We use the term single nucleotide polymorphism (SNP) to denote single base changes irrespective of functional consequences of the change. Our primary interest centers around the 180 base-pair homeobox itself and the encoded 60-amino acid residue homeodomain, although we have studied deletions of entire segments of the homeobox containing gene. Specifically, we have studied the effects of single base replacements in the homeobox that alter an amino acid residue in the encoded homeodomain. These mutations that change an amino acid residue often are called missense mutations. Since the homeobox itself is the most highly conserved region of the entire homeobox gene, one anticipates that these DNA mutations in the homeobox should, on average, be more deleterious than mutations elsewhere in the gene.

Homeodomain Structures in Solution – Free and DNA Bound

We started our work on the homeodomains by determining the structure of one of the NK family of homeodomains, namely vnd/NK-2, where vnd refers to ventral nervous system defective. The corresponding gene controls the development of part of the central nervous system of *D. melanogaster.* A block outline of the 723-amino acid residue homeodomain containing protein encoded by the *vnd/NK-2* gene is shown in Fig. 1. The homeodomain contains a helix-turn-helix DNA binding motif and

binds to specific DNA sequences. The NK-2 class of homeodomains is unusual in that it contains a tyrosine in position 54 of the sixty-amino acid residue homeodomain. In fact, the NK-2 class of homeodomains is defined by the presence of tyrosine in position 54. Methionine is the most common residue found in position 54 of all homeodomains described thus far. The DNA core consensus sequences recognized by tyrosine and methionine are show in Fig. 2. The sequences and helical regions of the vnd/NK-2 and CSX/NKX-2.5 homeodomains are shown in Fig. 3. A single amino acid residue mutation, for example, Y54M, or the corresponding mRNA codon change from *UAC* to *AUG*, in the vnd/NK-2 homeodomain or Y54C, or an mRNA codon change of *UAC* to *UGC* (a change in the 2nd base,) in the case of CSX/NKX-2.5 results in the genotypic alterations mentioned above.

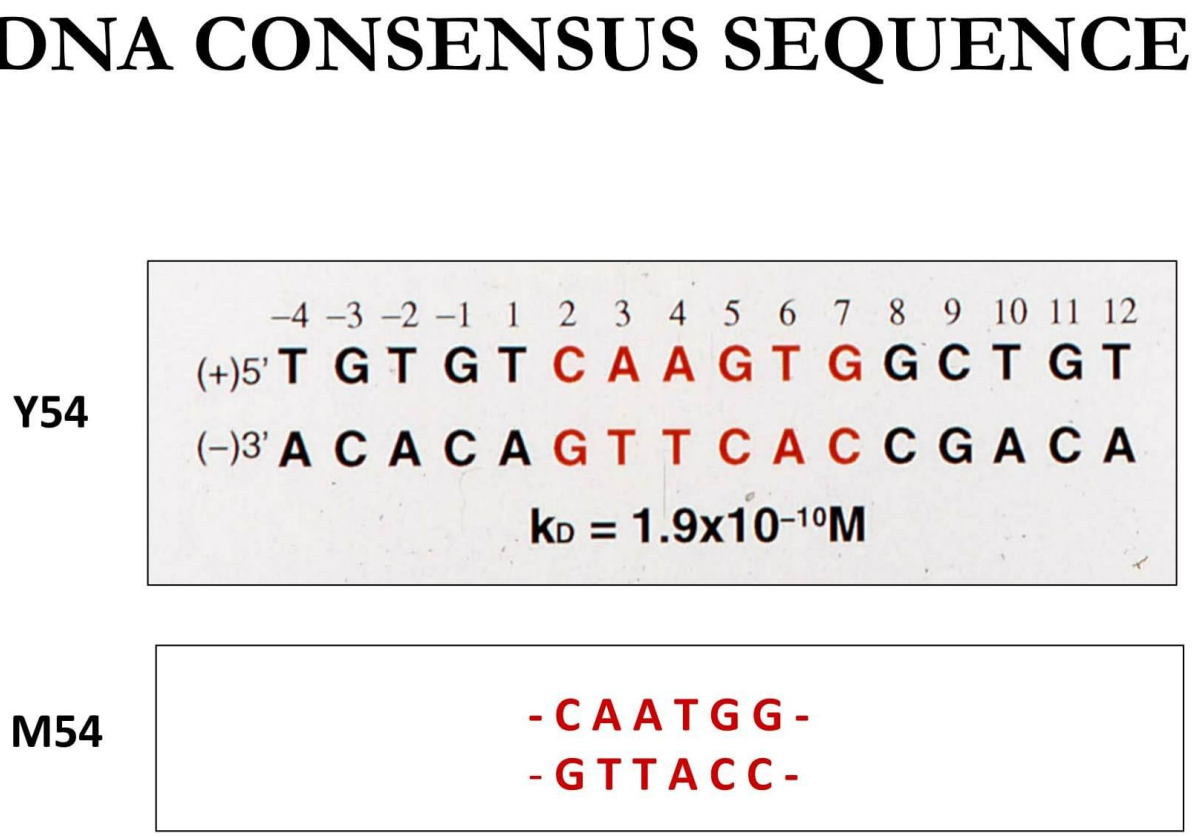

Figure 2. The consensus DNA recognition sequence for the NK-2 class of homeodomains along with the consensus DNA recognition sequence for homeodomains that contain methionine in position 54 of the homeodomain.

vnd/NK-2 & CSX/NKX-2.5 SEQUENCE HOMOLOGY

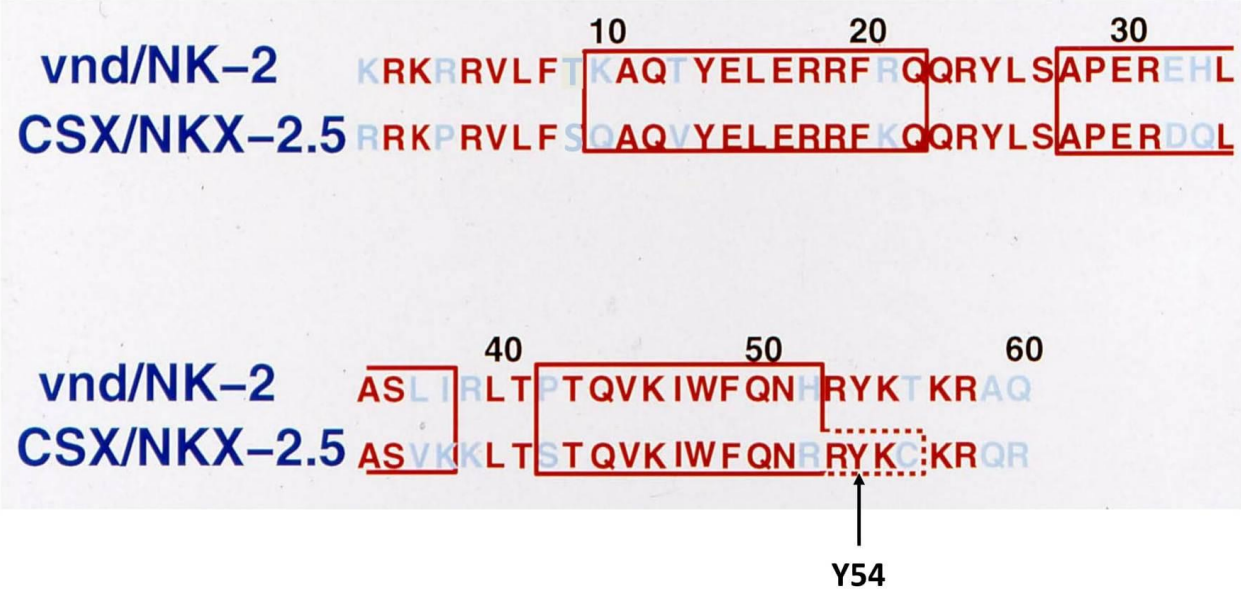

Figure 3. Amino acid residue sequences of the vnd/NK-2 and the CSX/NKX-2.5 homeodomains. Residues within the red boxes correspond to those that are helical in the folded state.

The primary technical questions are: 1. how does one go about obtaining the three-dimensional structure of a protein, 2. what experiments need to be done, 3. what type of information is required, 4. what protocols must be followed, 5. how precise is the structure and, 6. finally, how accurate is the structure? While going through all the details associated with an NMR structure determination is well-beyond the possibilities in this review, we will summarize the salient steps that are taken and try to give the reader an idea of what is involved, what are the advantages and what are the limitations. The basis of such structural determinations primarily is to obtain the through-space distances between atoms in the molecule such that the entire system (i.e., the coordinates of each atom in the protein) are overdetermined without having any pair of atoms at an unreasonably close distance or at an unacceptably long distance.

In any NMR study of the three-dimensional structure of a protein, a necessary first step is to synthesize or "express" the protein. Typically, the protein is expressed with the appropriate expression vector containing the relevant DNA in normal media (LB broth) and purified, the yield calculated. Then, the protein is examined to see if it is sufficiently soluble and well-behaved spectroscopically to proceed with the NMR structure determination. It is highly advisable to check the time stability of the protein to avoid any possibility of degradation over the period in which the experiments are performed. Before beginning the NMR experiments, often it is advisable to determine the thermal stability of the protein by carrying out CD measurements over a

range of temperatures. The helix content in the protein should be determined from the circular dichroism (CD) spectra. Simple one-dimensional spectra also should be collected to ensure that the protein is fully structured and stable under the temperature, pH, buffer and salt concentrations to be used. Once this information is successfully obtained, one can then express the protein in minimal media using $^{15}NH_4Cl$ as the sole nitrogen source. This uniform ^{15}N labeling of the protein is to obtain two-dimensional NMR experiments where one axis represents protein frequencies and the second axis represents ^{15}N frequencies (typically between directly bonded N – H atoms) so that the resonances simultaneously are separated both by ^{1}H and ^{15}N frequencies in a two-dimensional map. Often these data permit one to make preliminary resonance assignments, to obtain initial evidence indicating if protein has adopted a tertiary structure and to evaluate the feasibility of carrying out a full three-dimensional structure determination. Subsequently, the protein can be expressed in minimal media using both $^{15}NH_4Cl$ and ^{13}C labeled glucose. The purpose here would be to obtain three- and possibly even four-dimensional spectra for various technical reasons that include separation of the resonances into three- and four-dimensions to facilitate resonance assignments for high molecular weight proteins (greater than about 50 kDa) and to unambiguously determine distances between pairs of protons that are close in space but separated by many bonds. Also, such experiments are used to obtain spectra only of the protein when it is bound to another macromolecule such as a fragment of DNA. Some common questions include asking if there is an upper limit to the molecular weight of a protein or a protein complex whose structure can be obtained by NMR, are there significant structural differences between crystal structures and solution NMR structures and what will be the ultimate role of NMR in structural biology.

We determined the three-dimensional structure of a 77-amino acid residue protein that encompasses the homeodomain as shown in Fig. 4 (We used 80-amino acid residues in all subsequent studies on the vnd/NK-2 homeodomain.) [23,24,25]. The experiments were carried out at $12^{o}C$ at pH 4 in the absence of any added salt or buffer. A temperature of $12^{o}C$ was chosen based on both the CD and 1D NMR results that showed that the protein begins to unfold under the above conditions at higher temperatures. The structure of this 77-amino acid residue fragment turned out to be homologous to the structure of other homeodomain proteins that have been determined both in solution by NMR and in the crystal state by x-ray diffractometry. One important feature of this observation is that the conserved or invariant residues are in the appropriate sequential positions to produce the typical homeodomain fold with its hydrophobic core. The structure contains the three-helix bundle, where helix II and helix III constitute the well-known helix-turn-helix DNA binding motif found in all other homeodomains. This high degree of structural homology arises from similarities such as hydrophobicity or charge state maintained by amino acid residues in specific positions in the homeodomain that result in specific inter-residual contacts, thereby yielding a stable low energy structure. From the perspective of evolutionary conservation, we note that these low energy homologous structures exist across all species and use very similar control mechanisms to regulate function. Helix III also is called the recognition helix because of its residue-specific interactions with various sequences of DNA. The length of helix III in the absence of DNA varies among the homeodomains where helix III in vnd/NK-2 and fushi tarazu (ftz) [26] and Oct-2[27] extend from residues 42 – 52, whereas the antennapedia (antp) [28] homeodomain goes from residues 42 – 60. One diverse property of the various

homeodomains that are currently known is the temperature at which they melt; i.e., the midpoint of the folding-unfolding transition, T_c , in the absence of salt or added buffer. As examples, Antp unfolds around 48°C [29], whereas both the ftz [27] and vnd/NK-2 homeodomains unfold around 25°C in water. The primary reason for the difference correlates with the length of helix III and the variation in the values of T_c appears to depend heavily on the residue in positions 56, which in the vnd/NK-2 homeodomain is a threonine (T), whereas in antp it is a tryptophan (W). Mutation of this T to a W increases the stability of helix III and results in an increase in T_c. More importantly, a double mutation where position 52 is changed from a histidine (H) to an arginine (R) in addition to the T56R mutation resulted both in an elongation of helix III closer to residue 60 and an increase in T_c. Neither the single mutations H52R and T56W nor the double mutation H52RT56W alters the binding affinity of the homeodomain to the DNA [29]. A Y54M mutation reduced the affinity of the homeodomain for its cognate DNA by one order of magnitude but with no discernible alteration in the three-dimensional structure of the homeodomain (Fig. 5). In nature, such a mutation would be virtually impossible since it would require a 3-base codon change from TAC or TAT to ATG. The thermodynamic properties and thermal stability of the wild type and Y54M mutant vnd/NK-2 homeodomains, along with some transgenic experiments and preliminary functional will be described below.

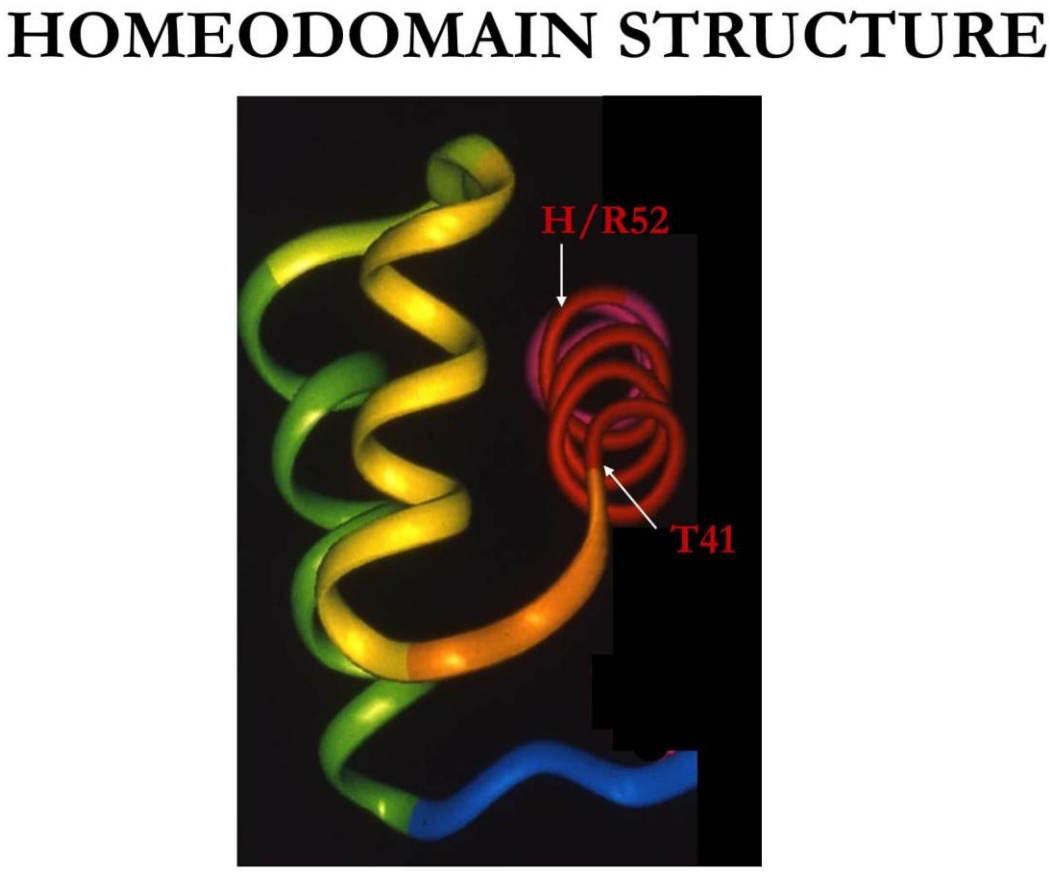

Figure 4. A depiction of the average tertiary structure of the vnd/NK-2 homeodomain determined by NMR. The first helix is shown in green, the second helix in yellow, the turn in orange and the third helix in red. The positions of residues 41 and 52 are shown.

The N-terminal region of this homeodomain appears to be unstructured and, thus, likely to be flexible, since very few distance restraints are not observed. However, one must consider that the remainder of the homeoprotein is absent, that other "factors" that might be present in a cell also are missing and that the structure was not determined under physiological conditions. While we fully anticipate that the three-helix bundle for the vnd/NK-2 protein would be the same in the cell, such structure determinations must be viewed with these caveats in mind.

One of the many remaining questions is the functional effect of any of these mutations. While there is no such information available on naturally occurring mutations in positions 52 or 56 for the vnd/NK-2 homeodomain, mutations in position 52 have been reported for two other homeodomains; even skipped, i.e., eve, and rough of *D. melanogaster*. These mutations in position 52 of the homeodomain are from R to H (the reverse of the mutation just described). For a mutation both in the eve and rough homeodomain proteins, normal development of the Drosophila embryos is observed at low temperatures (16 and 18 ^{o}C, respectively), whereas abnormal development occurs if the embryos develop at higher temperatures (30 and 25 ^{o}C, respectively). The amino acid residue found in position 56 of many homeodomains is typically hydrophobic and is believed to be important in determining the length of helix III. In fact, most members of the NK family of homeodomain proteins including vnd/NK-2 have the unusual characteristic of having S or T in position 56.

Some number of years ago, an intriguing observation was made that embryonically lethal Drosophila embryos were associated with a mutation in the homeobox of the *vnd/NK-2* gene that mapped to an A35T alteration in the homeodomain [30]. A study of the 80 - amino acid residue fragment containing the A35T mutant homeodomain showed that it was unable to adopt the folded three-helix bundle conformation in the absence of DNA. However, this A35T mutant binds to the 18 base-pair DNA sequence containing the 5' – CAAGTG – 3' consensus sequence but with an affinity that was 50 – fold lower than that of the wild type homeodomain. NMR spectra confirmed that the mutant protein adopted a stable conformation that contains the helix-turn-helix motif in the presence of the DNA [31]. The functional question that remains is whether the observed 50 – fold reduction in DNA binding affinity coupled with any distortions in the structure of the homeodomain in the DNA – bound state due to the presence of the larger T side chain in the hydrophobic core are sufficient to result in embryonic lethality.

The homeoproteins are known to be activators of transcription and it has been shown that this activation requires concerted protein-DNA as well as protein-protein interactions in the cell [3,11]. To help in the understanding of the mechanism of transcriptional control and to further appreciate what happens when individual bases have been mutated (i.e., phenotypic alterations or genetic disease), we felt it was useful to characterize the structural behavior of both the conserved as well as the variable amino acid residues in the homeodomain. It is now well-established that various genetic diseases and developmental abnormalities including embryonic lethality map to numerous single base mutations in the homeobox, the gene that encodes the homeodomain. The thinking has been that these mutations typically either alter the structure of the homeodomain, its specificity for a given DNA sequence or the homeodomain-DNA binding affinity. However, other possibilities involving the structure of the mRNA exist that do not involve a mutation in any amino acid residue and these will be discussed in a separate

section below. It is complex to relate changes in either the homeobox gene or the homeoprotein to phenotypic alterations or genetic disease. It is important to restate that the full-length homeobox gene (as well as its encoded homeoprotein) contains other conserved segments (Fig. 1), namely the tinman domain, the acidic domain and the NK-2 box. It is generally believed that mutations in the conserved regions of the homeobox gene are more likely to produce deleterious effects, although this belief has not yet been proved.

The role that each of the individual 60 - amino acid residues in the homeodomain plays in the structure and the sequence-specific DNA recognition as well as in the function of the homeodomain is not yet fully understood. Structural studies complemented by data on the free energies of association and binding have provided a useful approach to answer some of the questions. If one examines disease or phenotypic causing mutations across all known homeodomains, one finds them in many but not all positions in the homeodomain at least thus far. The first question that arises is: are mutations in certain positions more deleterious? Alternatively, one can ask if the evolutionary pressure to conserve or maintain invariant specific positions is sufficiently strong to maintain such behavior [32].

Valid questions that can be asked at this point are: 1. Is the consensus DNA sequence (the one with the highest affinity) the functional sequence? 2. Does the homeoprotein bind to more than one DNA sequence in the cell? 3. Are other 'factors' involved in the protein-DNA interaction? 4. Are histones and chromatin granules involved in limiting accessibility of the protein to DNA? 5. What can be said about dynamic features of the protein on the DNA? These and other analogous questions are beyond the scope of the structural and functional studies being described here.

Important advantage that we had in our study of the vnd/NK-2 homeodomain bound to its cognate DNA was that we had the structure of the homeodomain in the absence of DNA also determined in solution for comparison. The structure of the vnd/NK-2 homeodomain-DNA complex contains the same three-helix bundle as is found in the free homeodomain (Fig. 6). Helix III, the DNA recognition helix is inserted into the major groove of the DNA and makes most of the specific intermolecular homeodomain-DNA contacts (i.e., close distances). The primary contacts are made by Q50, N51 and Y54. N51 is invariant across all homeo-domains and its interaction with the 2nd adenine in the CA**A**GTG consensus sequence explains why this adenine also is necessary for high affinity binding. The N-terminal arm interacts with the minor groove of the DNA and becomes ordered upon binding. The L in position 7 of the homeodomain appears to make base-specific contacts with the DNA. Further, it has been established that mutations in position 7 alter the binding affinity. Also, mutations in position 7 of other homeodomains correlate with genetic alterations. A comparison of the part of the structure of both the wt and Y54M (codon change from *UAC* to *AUG*) vnd/NK-2 homeodomain bound to the consensus DNA sequence is shown in Fig. 5. Other than the slightly shorter side chain for the M54 vnd/NK-2 homeodomain and thus its inability to "contact" the DNA, no notable differences are observed for the mutant. An analogous set of experiments were carried out on the wild type and Y54C mutation for a CSX/NKX-2.5 homeodomain containing protein. Since the wt analog of this protein also has Y in position 54, it likewise has 5' – CAAGTG – 3' as its cognate DNA recognition sequence. Mutating the Y to C lowers the affinity of the homeodomain by about one order of magnitude also without any significant structural change in the homeodomain except removing the 'contact' that the Y was

making with the DNA. The question of any functional alteration induced by this single amino acid residue replacement now becomes crucial. Does the change in the homeodomain protein represent the etiology of the phenotypic and genotypic alterations observed for homeodomains in the NK-2 class or must we look elsewhere?

Having the three-dimensional structures of various homeodomain containing proteins both free and bound to DNA permit us to start trying to understand how mutations, specifically in the homeobox or in the encoded homeodomain in this review, lead to altered embryonic development or genetic disease. The information concerning atom-atom interactions from such studies provide a partial understanding of the role of individual amino acid residues. Although there is a high degree of tertiary structural homology across the homeodomains, variations due to individual amino acid residues are important. A visualization of the role of Y54 in determining the consensus DNA recognition sequence and the effect of altering position 54 to M in vnd/NK-2 and C in CSX/NKX-2.5 are summarized in Fig. 5 and Fig. 7, respectively. Since the homeodomain is capable of binding to various DNA sequences with varying affinities, the possibility that the homeodomains in fact to bind to a set of DNA sequences in the cell becomes crucial.

HOMEODOMAIN-DNA COMPLEX

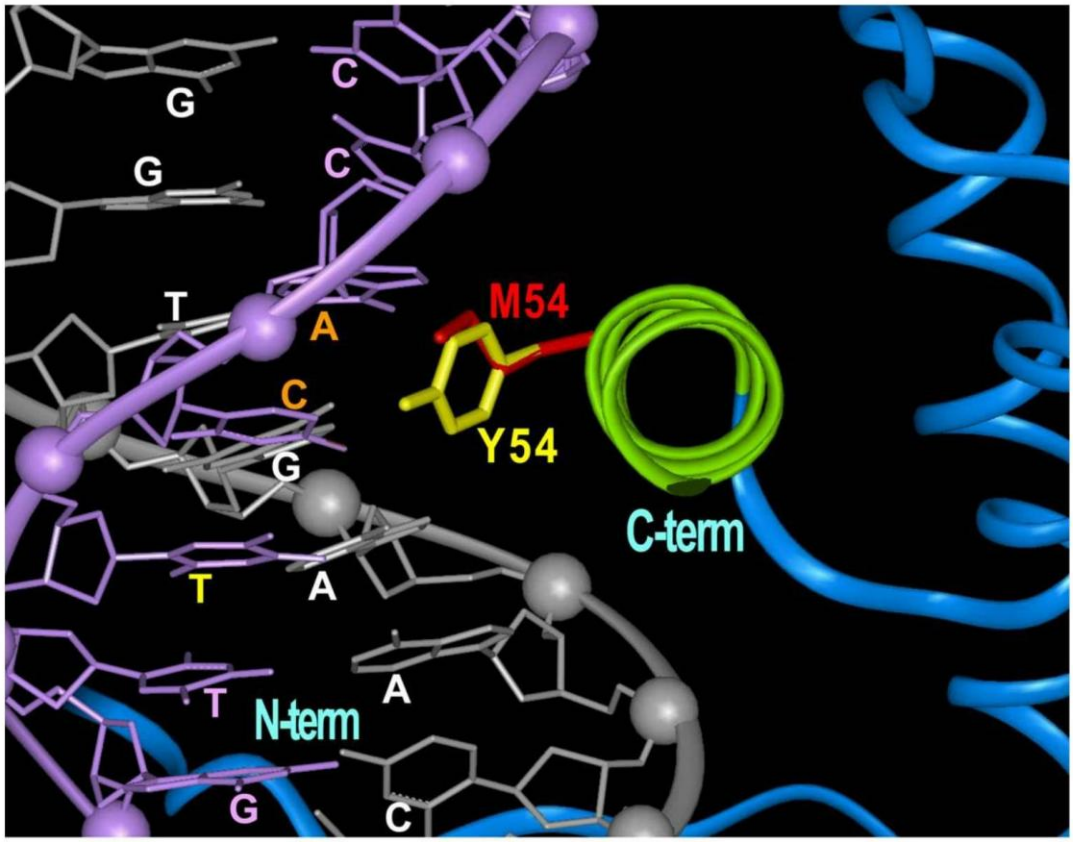

Figure 5. The orientation of residue 54 in helix III for the vnd/NK-2 homeodomain is shown for both tyrosine and methionine when the homeodomain is bound to its consensus DNA sequence as determined by NMR.

HOMEODOMAIN-DNA COMPLEX

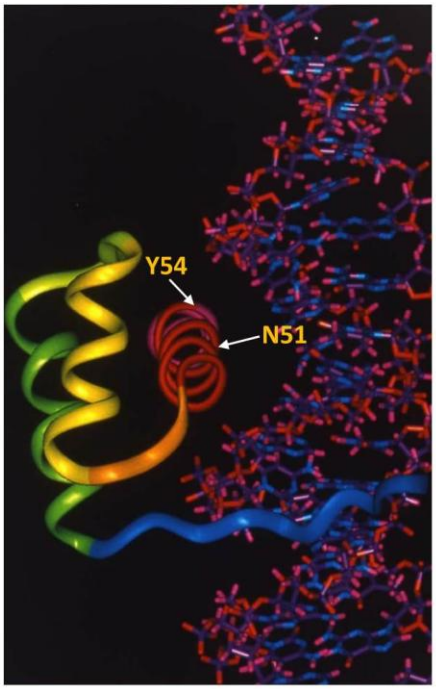

Figure 6. A representation of the average structure of the vnd/NK-2 homeodomain bound to its consensus DNA sequence as determined by NMR. The orientation of N51 and Y54 are shown.

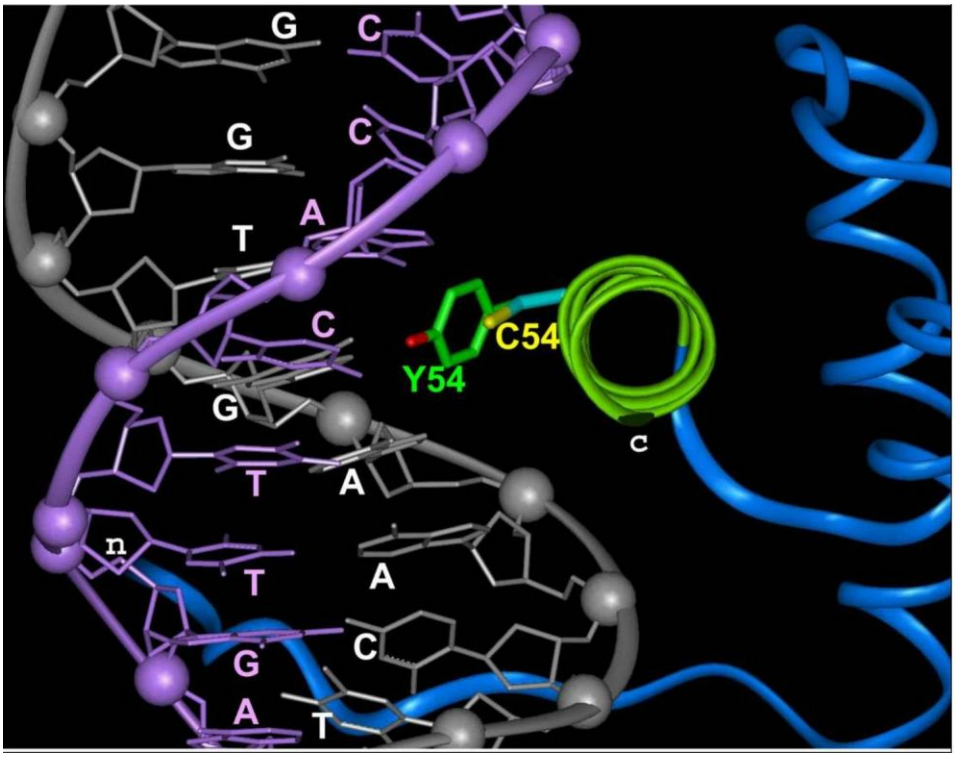

Figure 7. The orientation of residue 54 in helix III for the CSX/NKX-2.5 homeodomain is shown for both tyrosine and cysteine when the homeodomain is bound to its consensus DNA sequence as determined by NMR.

To further evaluate the stability of the homeodomain-DNA complex we measured the exchange lifetimes of the backbone amide protons by dissolving lyophilized samples of the wt and Y54M mutant 80 amino acid residue proteins containing the homeodomain bound to 16 base-pair fragments of DNA containing 5' – CAAGTG – 3' and 5' – CAATGG – 3', respectively. The two DNA fragments contain the putative core consensus sequences for homeodomain proteins that contain Y and M, respectively, in position 54 (Fig. 8). The exchange times are longest for the wt homeodomain containing Y in position 54 bound to 5' – CAAGTG – 3', which is in qualitative accord with the relative affinities measured in gel shift assays [30]. However, these NMR measurements provide additional information since one can measure the individual exchange lifetimes for each amide proton residue in the protein. These exchange lifetimes are not uniform over either the wild type or Y54M 60 residue homeodomain bound to either of the two DNA fragments most notably in helix III, the DNA recognition helix. Such behavior is a result of internal motions of the individual residues in helix III so that in solution under the experimental conditions used, the binding behavior of the protein-DNA complexes cannot be described using a single affinity constant. Any further interpretation of this behavior would be speculative at this time.

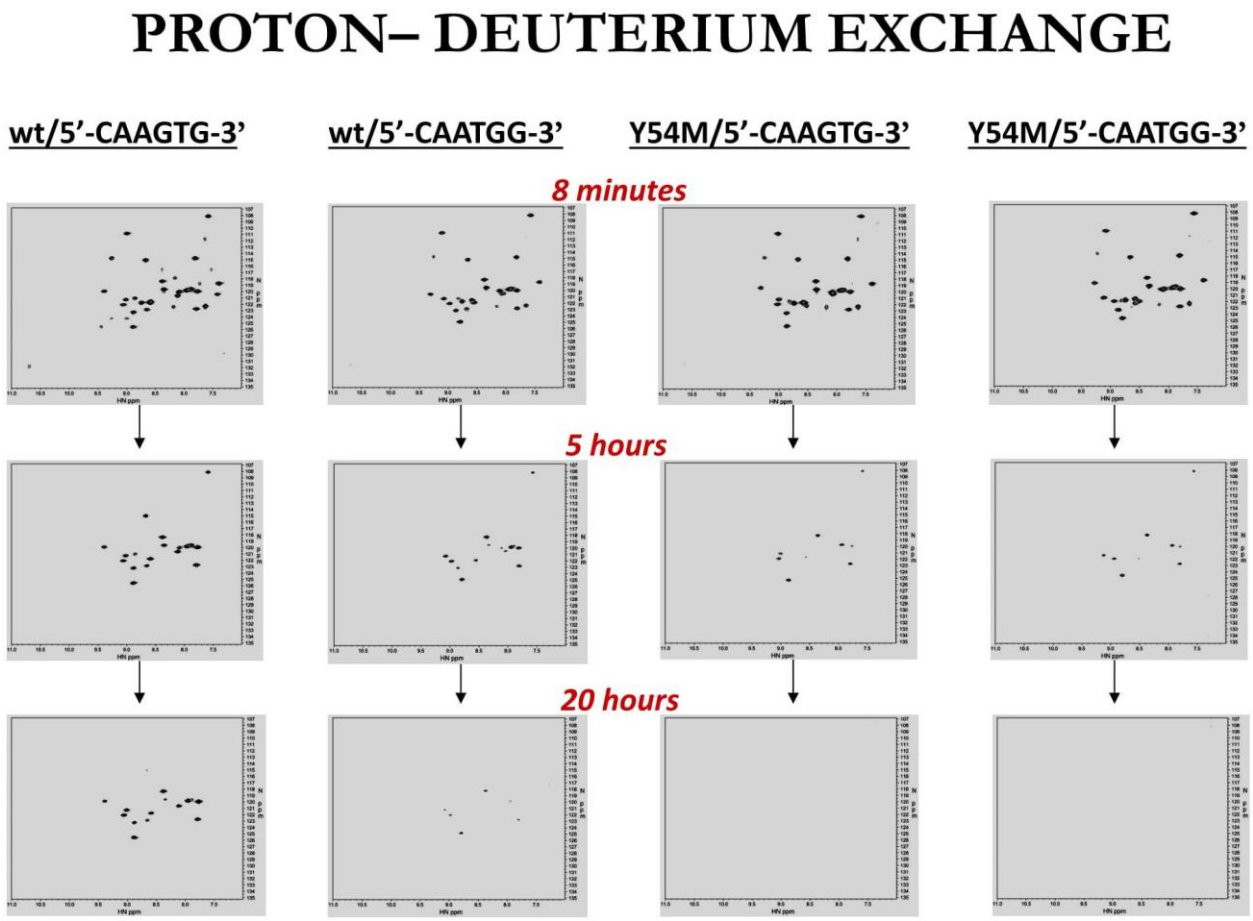

Figure 8. Two-dimensional 15N – 1H HSQC spectra for the wild type and Y54M mutant vnd/NK-2 homeodomain bound to the respective core consensus DNA sequences.

Thermodynamic Properties of the Homeodomain

Solution differential scanning calorimetry (DSC) measures the difference in heat energy uptake between a sample and reference with increasing temperature thereby producing what is called an endotherm [33]. As in the cases we studied, this heat energy uptake is measured (i.e., the endotherm is recorded) during the folding-unfolding transition of the homeodomain from the ordered three-helix bundle to a disordered state or random coil. These results are then used to calculate, based on an appropriate model of the folding-unfolding process to give thermodynamic parameters such as ΔH_{cal}, the calorimetric enthalpy [34]. Likewise, the Van't Hoff equation, which relates the change of the equilibrium constant for the folding-unfolding process to what is called ΔH_{vH} (Eq. 1).

$$d \ln K_{eq}/dT = \Delta H_{vH}/RT^2 \qquad [1]$$

This determination of ΔH_{vH}, which uses the Gibbs relation (Eq. 2) assumes the folding-unfolding process is two-state and further that the enthalpy and entropy, ΔH and ΔS, are temperature independent over the temperature range of the experiment.

$$\Delta G = -RT\ln K_{eq} = \Delta H - T\Delta S \qquad [2]$$

Eq. 2 assumes a two-state process. This approximation formally is incorrect for folding-unfolding of the homeodomain simply because many amino acid residues must change configuration during the folding or unfolding process. However, the two-state approximation might be sufficiently good if only the fully folded and fully unfolded states are populated [35]. One test of the two-state approximation can be obtained from the ratio $\Delta H_{cal}/\Delta H_{vH}$. If the ratio is unity in any given situation, then the two-state approximation might be acceptable for that purpose.

Thermodynamic studies of protein unfolding as well as their binding to their respective receptors, which is DNA in the case of the homeodomain, have been carried out [36,37]. Such studies measure the forces involved either in unfolding or in protein binding and provide information that is complementary to that obtained in the corresponding three-dimensional structure determinations. The structural studies provide information on the local short-range interactions that lead to the overall global fold of the protein, whereas the thermodynamic parameters include the local intramolecular interactions as well as the intermolecular interactions especially those involving the solvent, which often is water.

We first investigated the thermodynamics of folding/unfolding and binding to DNA of the vnd/NK-2 together with two mutants (H52R and H52RT56W) by DSC and isothermal titration calorimetry (ISC) [38]. Our prime reason for undertaking these studies was to gain additional understanding of the shortened length of helix III in the unbound state. Even though helix III is not shortened in the presence of DNA and might not be shortened in the full-length homeoprotein in the presence or absence of DNA or other factors that might interact with this region of the protein, our reasoning was that the energetics associated with helix III might ultimately be useful in understanding the protein-DNA binding in more detail. The thermal unfolding of the homeodomain was carried out at pH 7.4 in the presence of 50 – 500 mM NaCl and either phosphate or Hepes buffer. Addition of NaCl along with either phosphate or Hepes buffer increased the thermal stability of the wild type and the two mutant homeodomains, H52R and H52RT56W vnd/NK-2. The order of thermal stability was H52RT56W vnd/NK-2 > H52R vnd/NK-2 > wt vnd/NK-2. What surprised us in these studies is that the cooperative ratio ($\Delta H_{cal}/\Delta H_{vH}$) was significantly less than unity under most

conditions. The ratio varied from 0.27 to 0.64 for the wt vnd/NK-2, from 0.42 to 0.73 for the single mutant H52R vnd/NK-2 and from 0.61 to 0.80 for the double mutant H52RT56W vnd/NK-2. These results on the cooperative remain perplexing. Prabhu and Sharp [39] conclude that there is no sound theoretical reason to expect ΔH_{cal} and ΔH_{vH} to differ. However, based on the number of studies we performed as well as their reproducibility, it is difficult to attribute these differences in ΔH_{cal} and ΔH_{vH} simply to measurement error. The changes in the cooperative ratio by mutating positions 52 and 56 suggest that the behavior of helix III somehow is responsible. One possibility would be that the unfolding process deviates significantly from the two-state approximation and that the population of intermediate states differs between the wild type and the two mutant homeodomains. This observation seems to imply that the helix-coil transition that helix III undergoes is not fully correlated with the global unfolding of the three-helix bundle of the homeodomain. Further clarification here awaits additional investigation.

The thermodynamic parameters of the sequence-specific binding of the 80-amino acid residue fragment containing the wt vnd/NK-2 as well as both the single and double mutants with an 18 base-pair segment of DNA containing the 5'-CAAGTG-3' core segment by ITC demonstrated that, in all three cases, the protein-DNA interactions were enthalpically controlled. The binding stoichiometry was found to be 1:1. One interesting feature that emerged from these studies was that the observed value for the change in heat capacity**,** ΔC_p, was found to be small and positive for the wt vnd/NK-2 homeodomain over the temperature range 10^oC – 30^oC. Typically, large negative heat capacity changes usually are associated with site-specific protein-DNA binding. The two mutant vnd/NK-2 homeodomains, H52R vnd/NK-2 and H52RT56W vnd/NK-2 were small and negative of the range 10^oC – 30^oC [40]. Thus, for the wt vnd/NK-2, this exceptional behavior of ΔC_p with temperature appears to occur because of a decrease in the solvent-accessible hydrophobic surfaces produced by the DNA binding is offset by solvent rearrangement, burial of polar surfaces, and possibly a strain in the complex of the DNA duplex. Also, the entropic changes we observed for the homeodomain-DNA binding were much smaller than we would have expected based on analogous protein-DNA binding studies. It seemed rather remarkable to us that the observed entropy changes for the binding were so small considering the large structural changes we believed were involved. Here again we hope that future studies will clarify the questions raised here.

Given that the length of helix III for wt vnd/NK-2 is about 8 amino acid residues shorter than that observed for many other unbound homeodomains, we were somewhat surprised that neither the thermodynamic parameters of DNA binding nor the binding affinities themselves differed significantly relative to those homeodomains where helix III was already elongated in the unbound state. In fact, we had expected that there would be a measurable entropic cost to elongate helix III upon binding to DNA. While the measurements do show that $T\Delta S$ is positive for the wt vnd/NK-2 as well as the single and double mutant homeodomains, we are unable to discern any correlation with the structural behavior of helix III. The reason for the apparent evolutionary conservation of a homeodomain protein with limited stability of the DNA recognition helix could be the subject of a separate study. Specifically, why is the helix destabilizing amino acid residue, H, conserved in position 52 of the vnd/NK-2 rather than the more commonly found amino acid residue R? Would the ability of a mutant vnd/NK-2 homeobox

gene with the codon for R in position 52 rather than the codon for H allow *D. melanogaster* embryos to develop at a higher temperature as the antithesis of the codon for R being replaced by the codon for H in both eve and rough, and would such mutations lead to a cascade of undesirable events? Would such a study provide information on how and why the homeodomains proteins have evolved and what selective pressures are involved?

We carried out an analogous set of DSC and ITC experiments on the C56S CSX/NKX-2.5 homeodomain [41]. For this homeodomain, the cooperative ratio was close to unity for the wild type protein under the conditions studied. However, a detailed mutational analysis of residues 52 and 56 was not performed and, thus, these studies do not shed further light on discrepancies between ΔH_{cal} and ΔH_{vH} observed for the vnd/NK-2 homeodomain.

Transgenic Alterations

To compliment the structural and thermodynamic studies were carried out on single amino acid residue replacements in the vnd/NK-2 homeodomain, we were curious what would happen if we made various changes in the *vnd/NK-2* gene including one that changed the codon for the amino acid residue in position 54 from the one that encodes Y to the one that encodes M in a more biological experiment. We were motivated to do so in part because the structural, binding and thermodynamic changes observed for the Y54M vnd/NK-2 homeodomain appeared to us to be quite subtle and we asked ourselves what changes in the gene are required to produce observable functional alterations. Would what appears to be a subtle structural change when replacing Y by M in the vnd/NK-2 protein have any functional effect and, if so, how and why? A fundamental question that we ask here is whether and, if so, under what circumstances do the changes resulting from this single amino acid reside replacement constitute the molecular basis for any observed phenotypic or genotypic alterations? Is the etiology of the alteration due to the single residue replacement in the protein or might it be found elsewhere involving transcription or translation?

The experiments [5] that we designed were transgenic ones, where ectopic expression patterns of wild type and mutant homeobox genes could be examined. In addition, since two downstream genes that are targeted by the vnd/NK-2 protein, a study of the expression of the *vnd/NK-2* homeobox gene should answer whether mutating the Y54 in the homeodomain alters functionality (i.e., Y54M or UAC to UAG for the mRNA in vnd/NK-2 and Y54C or UAC to UGC in humans). The downstream genes that are targeted by the vnd/NK-2 protein are *ind* (intermediate neuroblast defective) and *msh* (muscle specific homeobox). The *msh* gene is regulated by vnd/NK-2 both directly and indirectly through *ind* (Fig. 9).

The technical details of the ectopic expression experiments are beyond the scope of this review. The various constructs used each contained an upstream activating sequence (UAS) linked to a reporter gene *Gal4*. After modification, each of the vnd/NK-2 cDNAs was subcloned into the *Eco*R1 site of puast, P-element of the UAS transcript. The corresponding DNA was microinjected into the Drosophila embryos to generate the transgenic fly lines. *In situ* hybridization was then carried out using a well-known protocol. The embryos were stained with a *vnd/NK-2*, *ind* or *msh* RNA probe for visualization of the expression of the respective RNA.

DOWN-REGULATION of *ind* and *msh* by vnd/NK-2

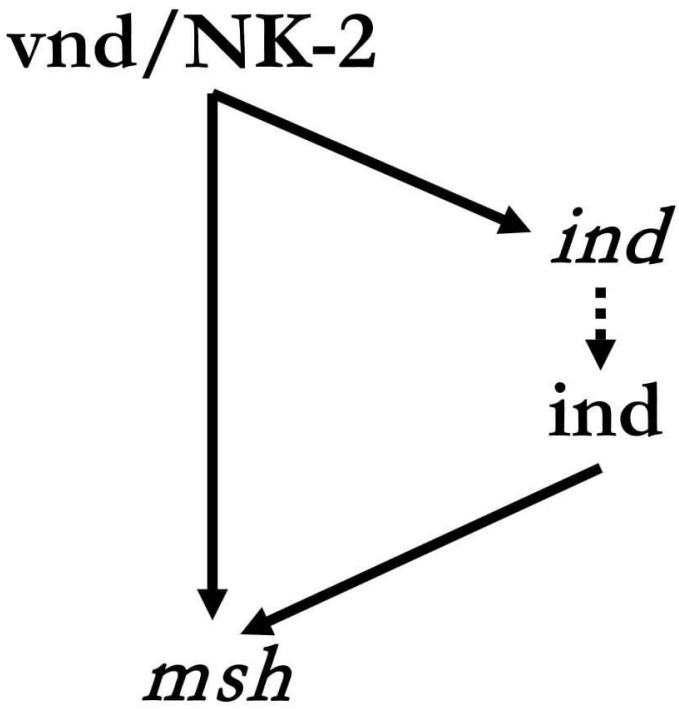

Figure 9. Depiction of the down regulation of the *intermediate neuroblast defective (ind)* and *muscle specific homeobox (msh)* genes by the vnd/NK-2 homeodomain containing protein.

The ectopic expression patterns of wt and four mutant vnd/NK-2 full-length genes were analyzed together with expression of *ind* and *msh*. Three of the mutants of deletions of conserved regions of the vnd/NK-2 gene, namely the tinman motif, the acidic motif and the *NK-2* box. The fourth mutant was the codon for the Y54M residue in the homeodomain, *TAC* to *ATG*, where the rest of the vnd/NK-2 gene was left unchanged. One might suspect *a priori* that removing entire conserved regions of a gene would be more deleterious than changing a single codon. The results of the various ectopic expression studies are depicted in Fig. 10 - 11 and have been described in detail previously. To summarize the results, reductions in the mRNA expression were observed for the acidic domain and the mRNA that encodes the single amino acid Y54M replacement. Further, all the mutants including the one that encodes the Y54M mutation are embryonically lethal. The result on the Y54M mutant arguably is the most surprising one given that the structural alteration of the homeodomain-DNA complex is minimal. For the tinman and NK-2 box, no significant reductions in the mRNA expression were observed.

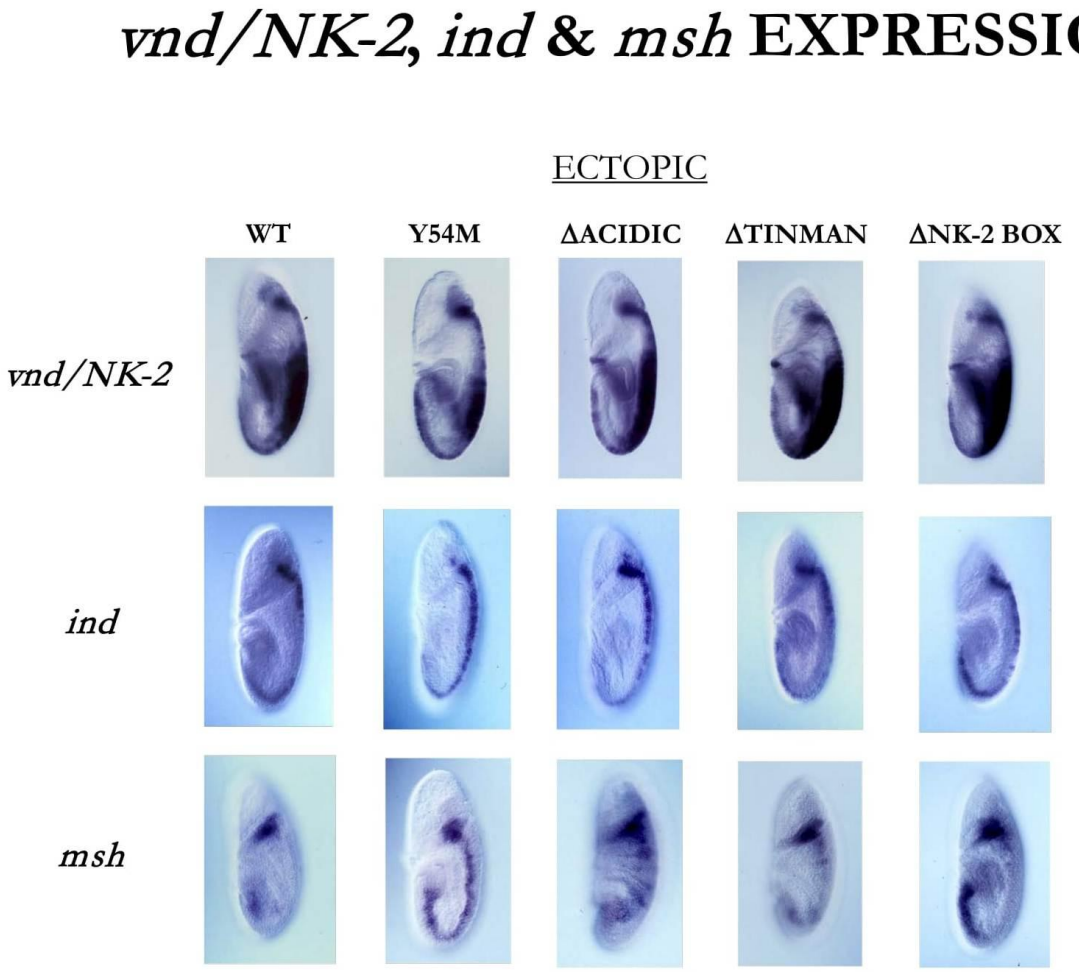

Figure 10. Transgenic ectopic expression of the wild type and four mutants of the mRNA produced by the *vnd/NK-2* gene and observation of the effect on the downstream target *ind* and *msh* genes. The effect of the mutation that would encode Y54M in the homeodomain is to completely disable the repression of *ind* and *msh*.

Based on both the transgenic results obtained on the vnd/NK-2 gene and from observations made on the human cardiac gene CSX/NKX2.5, several questions arise. Both the transgenic results on the Y54M mutation (or the *UAC* to *AUG* codon replacement in the mRNA) for the *vnd/NK-2* homeobox gene and the mutant protein encoded by the transgene, and the Y54C mutation (or the *UAC* to *UGC* codon replacement in the mRNA) for the *CSX/NKX-2.5* homeobox gene mapped from infants born with congenital heart disease show no significant structural changes in either the unbound or DNA bound states, and only small difference in their binding affinities to their cognate 5' – CAAGTG – 3' DNA. The first question to ask is: Do these small differences represent the etiology of the lethality that results from mutating the codon for Y to that of M in the *vnd/NK-2* gene and from the alteration of the codon for Y to that of C in the *CSX/NKX-2.5* gene? A second question would be to ask if other factors present in the Drosophila cells or the human cardiac cells that interact with the respective homeobox genes or their protein products somehow amplify the effects of the individual mutations and, thus, that one would need to investigate additional possibly protein-protein interactions that involve the homeodomains? A third question considers

the effect that the mutations would have given that homeodomains most likely bind to an ensemble of DNA sequences with a distribution of both affinities and populations and that mutations involving a critical amino acid residue such as Y54 significantly modifies the affinity and population distribution perhaps even in a cooperative way. One might argue that such a cooperative change might be the origin of the observed overall lower mRNA expression observed in the transgenic experiments. In addition, one might even imagine other factors that are present in cells that critically require Y in position 54 of the homeodomain. The critical question to be asked at this point is: can one conclude that the homeodomain protein is the etiological agent for the phenotypical or genotypical alterations described above? What about the changes in the DNA itself? Because of the crucial nature of this question, we would like to "offer" another possible explanation that, to date, has received little attention. That is, can changes in the messenger RNA, at least in some cases, be the etiological agent of observed phenotypic alterations or genetic disease? We attempt to address this question in the following sections.

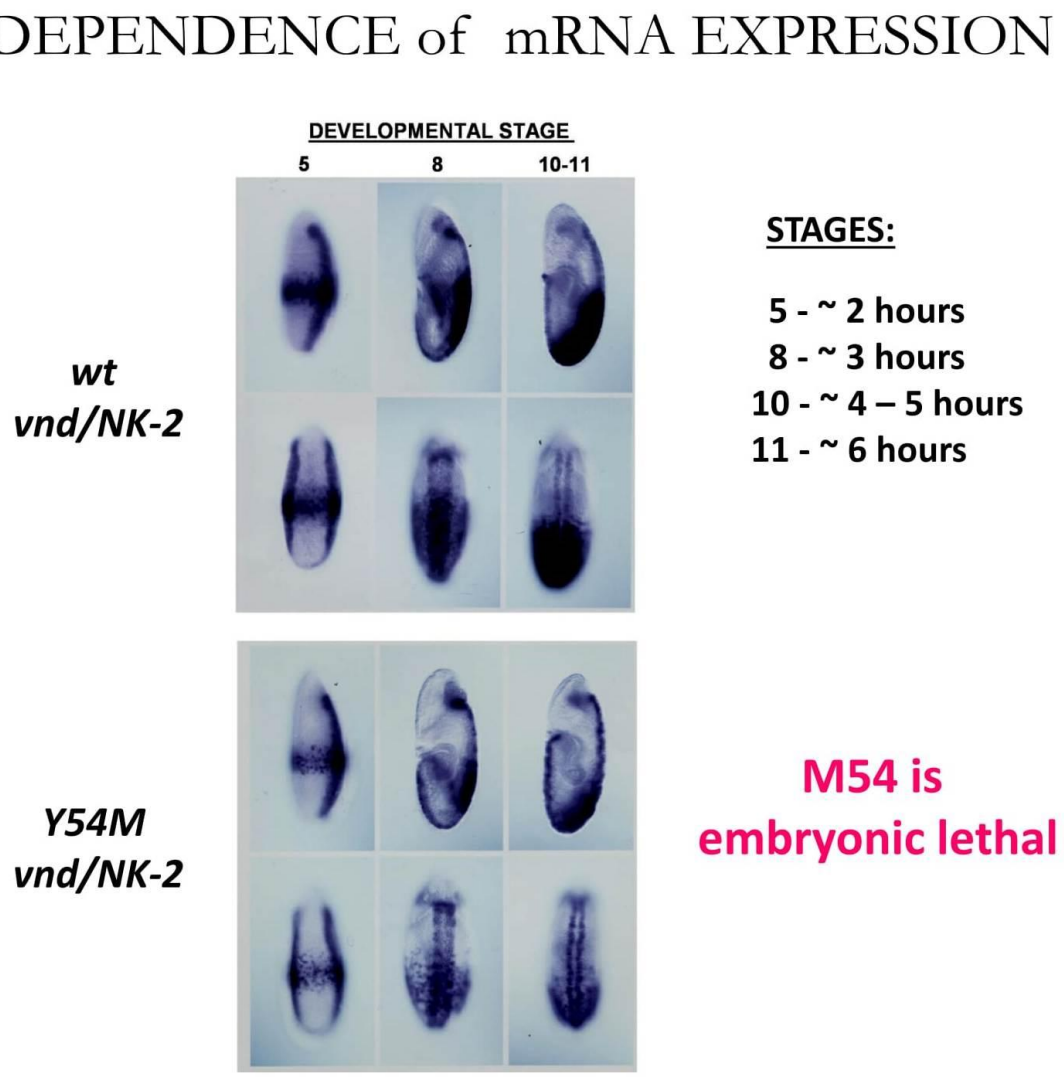

Figure 11. Time dependence of the mRNA of both the wild type code for the Y54M of the homeodomain of *vnd/NK-2*. The levels of expression of the mutant mRNA are significantly lower than those of the wild type.

Figure 12. 5' upstream region of the 351 base from the coding region of the vnd/NK-2 gene that was fused to a CAT reporter gene together with the various modifications made to examine the effects of cotransfection of a construct designed to express the vnd/NK-2 homeodomain.

CAT Assay

A chloramphenicol acetyltransferase (CAT) reporter assay was carried out where a set of 351 base-pair upstream putative enhancer elements (Fig. 12) were fused to the reporter. Both the *wild type* and *Y5M mutant vnd/NK-2* expression vectors were transiently co-transfected with the reporter gene construct. The wild type enhancer element contains several recognition sites for the homeodomains. Two of these sites contain the 5' – CAAGTG – 3' consensus vnd/NK-2 homeodomain recognition sequence (i.e., Y54) and one site contains the 5' – CAATGG – 3', which is the consensus site for homeodomains that contain M54 such as antennapedia. In addition, there are several sites that contain 5' – CAAGXX – 3' and, thus, are expected to bind to the homeodomains but with reduced affinities.

For the transient transfection Schneider 2 (S2) Drosophila cells were used. A critical feature of these cells is that they contain no endogenous *vnd/NK-2*. Since it is well-established that the invariant N in position 51 of all the homeodomains interacts with the 2nd A of the 5' – CA**A**G – 3', we chose replacement of this A by C as a means of "inactivating" this DNA binding site as shown in Fig. 12. The results of the CAT transient transfection assay are presented in Fig. 13.

CAT Assay of Reporter Gene Cotransfected with a vnd/NK-2 Expression Vector

Reporter gene	NK-2 Protein		
	-NK-2	+WT NK-2	+Y54M NK-2
pNK-2-B1 (3 high affinity sites, wt-mt-wt)	83.6	0	16.1
pNK-2-B2 (only 1 of the 3 sites, wt)	44.0	0	43.5
pNK-2-B3 (only 1 of the 3 sites, mt)	29.2	0	37.7
pNK-2-B4 (none of the sites)	4.2	0	61.6

wt : 5'-CAAGTG-3'
mt : 5'-CAATGG-3'
*CAT activity normalized by total cell count

Figure 13. CAT assay of the reporter gene containing the 351 base-pair DNA putative enhancer fragment cotransfected with the vnd/NK-2 homeodomain expression vector.

Once we obtained the results of the CAT assays, the difficulty that confronted us was the interpretation. In all our studies, the wild type vnd/NK-2 homeoprotein completely suppressed the activity of the chloramphenicol transferase, even when the various sites indicated in Fig. 12 were "inactivated". Conversely, the Y54M vnd/NK-2 showed activation or repression that varied as a function of the sites that were inactivated. These results are quite difficult to interpret especially in the light of the structural behavior of the Y54M vnd/NK-2 bound to the 18 base-pair DNA fragment containing the cognate 5' – CAAGTG – 3' as its core. These results do tell us that at least interpretation of the CAT assay is more complex than simply considering inability of M54 to contact various possible DNA recognition sequences.

Affymetrix Microchip mRNA Transient Transfection Assay

Another approach we took to investigate if the genes we have studied are able to up or down regulate *ind* and *msh* as was seen in the transgenic studies described above.

A standard protocol used to transiently transfect Schneider 2 (S2) *D. melanogaster* cells. The S2 cells were transiently transfected with the wt, Y54M and Y54Y vnd/NK-2 genes coupled to a reporter gene for assay on Affymetrix microchips in separate transient transfections experiments. The cells were transfected transiently to avoid integration of the exogenous genetic into the Drosophila genome for both 24 and 72 hours. The results are summarized in Table I for the 72-hour transfection. It is important to point out that the experimental results obtained here are highly preliminary and should not yet be used to draw hard conclusions. For technical reasons and other limitations, these experiments were performed only once and, although there is an internal reference that suggests the results are reliable. However, currently these data must be considered as tenuous at best and should be used only to contemplate further experimentation. What these results suggest is that the wt *vnd/NK-2* gene is capable of fully repressing (in this transient transfection assay) *ind* and *msh* in accord with the transgenic expression results. The data imply that the Y54M *vnd/NK-2* gene can repress ind and msh but at only about 30% of the efficiency of the wt analog. The data further suggest that the synonymous Y54Y *vnd/NK-2* is incapable of any measurable suppression of *ind* and *msh*. These results are compatible with the observation for Y14Y [45] in the human CSX/NKX-2.5, namely that synonymous mutations can produce either phenotypic alterations or genetic disease and that it is possible that the associated etiology is related to the change in the mRNA.

Table I: RELEVANT RESULTS – MICROCHIP ASSAY

GENE	wt vnd	Y54M vnd	Y54Y vnd*
ind	1.01	3.11	60.2
msh	1.90	4.23	53.7

* This synonymous change significantly alters the expression of two downstream target genes, ind and msh, and thus should produce a phenotype.

Synonymous and Nonsynonymous Mutations

The purpose of this section is to introduce the hypothesis that the etiology of the observed lethality in the Y54M mutation of the vnd/NK-2 homeodomain protein might reside, at least in part, in the mRNA transcribed from the DNA of the homeobox gene rather than in the structure, binding and/or the thermodynamic properties of the protein. How might the mRNA be the etiological agent and how might we test that hypothesis? What we know is that to change

Y54 to M requires altering the respective codon from *UAC* to *AUG* in *vnd/NK-2* or from Y54C from *UAC* to *UGC* in *CSX/NKX-2.5* (Table II). Firstly, we should point out that single nucleotide polymorphysms (SNPs) that lead to genetic disease always involve the change of only one base such as *UAC* to *UGC* that lead to the observed congenital heart disease. The change we made from Y to M, the most common amino acid residue found in position 54 of all homeodomains and which involves mutating all three bases of the codon, was done solely for studying the role of M in position 54. This change from Y to M would be virtually impossible to ever be found in nature.

Table II: mRNA Mutation

vnd/NK-2 (CNS)		CSX/NKX 2.5 (Cardiac)	
Y54M	Y54Y	Y54C	Y14Y
UAC ->AUG	UAC -> UAU	UAC -> UGC	UAU -> UAC

To develop this hypothesis, what we need to remember is that the mRNA is single stranded (i.e., not paired to a complimentary strand of mRNA) before entering the ribosomal tunnel. Thus, any base pairing that occurs in the mRNA is intra-strand may involve many RNA bases. Once a single base is changed such as in the case of Y54C from *UAC* to *UGC*, the 2nd base of the codon, G, will show different base pairing that the original *UAC*. An example of such single stranded RNA base pairing is shown in Fig. 14. Before the single stranded mRNA can enter the ribosomal tunnel and find its complimentary tRNA to encode the next amino acid residue in the nascent protein chain, it must first linearize, where the non-covalent base-paired bonds are broken. For each base in the RNA chain this will require free energy that can be quantified on a per base-pair basis and the total free energy difference will depend on the total number of base pairs broken in preparation for passing into the ribosomal tunnel. Any alteration in any base likely will change the set of base-paired structures adopted by the mRNA thereby changing the total free energy difference between the codon encoding the wild type protein and the codon encoding the mutant protein. Admittedly quantifying this free energy difference would be quite difficult but, in principal, the difference could be large and, thus, consequential. So, what other approaches might provide useful relevant information? One approach might be to study synonymous mutations; that is, replace a wild type codon with it synonym and examine the consequences of such a replacement. If, indeed, the same amino acid residue is encoded by a synonym, any phenotypic or genotypic alteration then could not be ascribed to the structure, binding affinity for DNA or the thermodynamic properties of the protein. The question that arises at this point is whether synonymous mutations alter function. If the hypothesis suggested here has merit, it would imply the rules of natural selection apply and that synonymous mutations are capable of influencing the viability of an organism.

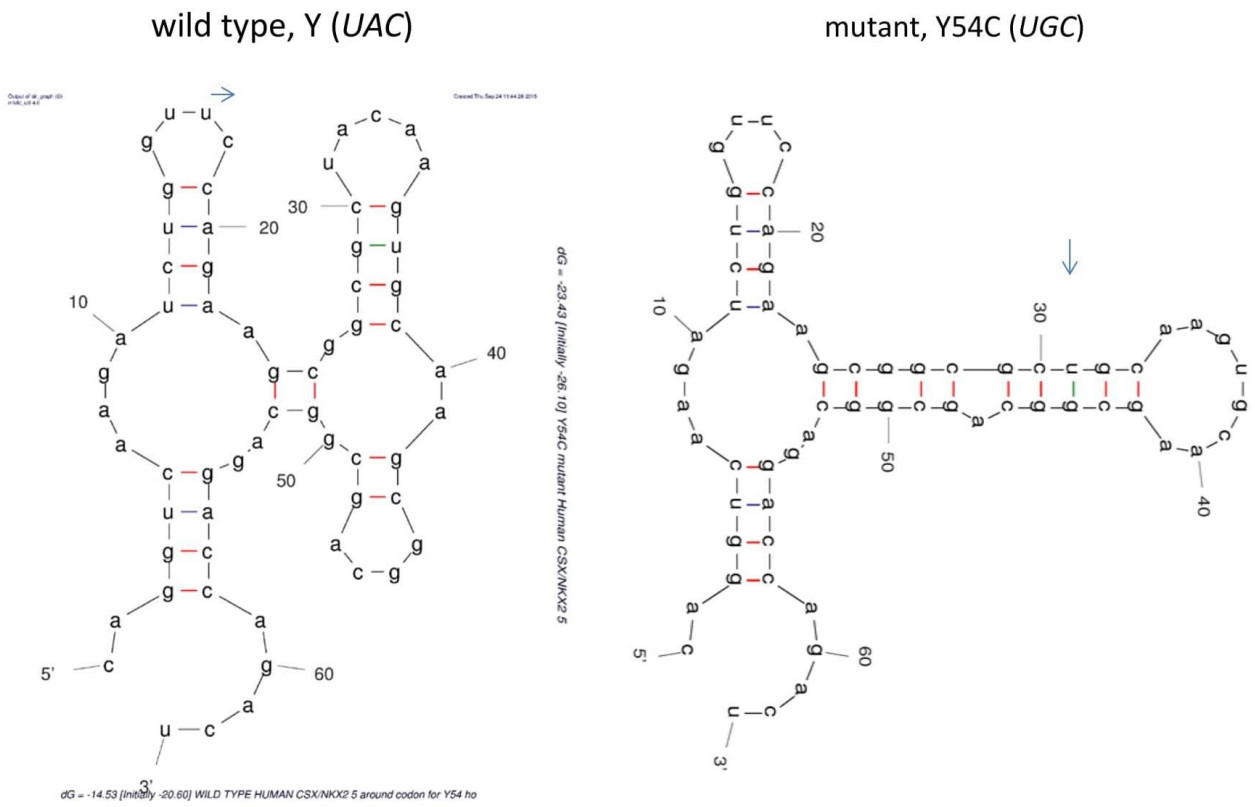

Figure 14. Possible ssmRNA structural alterations when the wild type CSX/NKX-2.5 mRNA is altered to encode the Y54C mutant.

Synonymous mutations are changes in the mRNA codon that encode for the same amino acid residue [42]. This phenomenon occurs because there are degeneracies in the genetic code for all amino acid residues except M. In the past, synonymous mutations; that is, use of a degenerate codon in the mRNA was thought to be "silent" or, in other words, synonymous change of mRNA codons had no effect. The possibility that synonymous mutations are, in fact, not "silent" and that they may produce phenotypes was introduced about 10 years ago. Supporting evidence currently is mounting that synonymous codons are not silent; that is, synonymous replacements can result in phenotypic alterations. For most organisms whose genome has been sequenced, synonymous codons are not found in equal amounts; that is, there is codon bias. For example, in the case of Y in *D. melanogaster*, *UAC* is found more often than *UAU* and *UAC* is referred to as the common codon whereas *UAU* is referred to as the rare codon. Furthermore, synonymous codons are not used in equal frequency; that is, there exists codon usage bias and this bias does not necessarily correlate with the relative occurrence of the two degenerate codons. Furthermore, the amount of the corresponding tRNAs also differ. In fact, a synonymous mutation for the tyrosine in position 14 of the CSX/NKX-2.5 has been reported. The usual or wild

type tyrosine in position 14 is encoded by the conserved rare *UAU* codon in the *NK-2* class of homeobox genes. Mutation of this codon to the common *UAC* codon leads to congenital heart defects [43] even though tyrosine is encoded in both cases.

Conclusions

Over the years we have carried out many structural, binding and functional studies on the wild type and various mutant vnd/NK-2 homeodomains and on their corresponding RNAs. What have we learned, what conclusions can be drawn and where does this lead us? Firstly, it is important to point out that all the structural and biophysical studies are carried out in vitro with all the accompanying limitations. Also, in the transgenic studies an additional gene is expressed ectopically in nascent embryos rather than by homologous replacement of the endogenous wild type gene. While the data have turned out to be interesting, one must still consider this limitation when examining the results and drawing conclusions.

The structural studies of the Y54M vnd/NK-2 homeodomain show that there is no significant change when Y is replaced by M. Yet both the ectopic expression transgenic results and the CAT assay data show that this amino acid replacement is functionally significant. Further, we know that in CSX/NKX-2.5 the Y54C replacement also is functionally catastrophic. A detailed molecular explanation for these observations remains elusive, although the information available on synonymous mutations does indicate an important role associated with changes in the mRNA that requires further investigation.

Acknowledgements

We thank Drs. Edward D. Korn and Alan Peterkofsky for many helpful discussions and comments. We acknowledge the Intramural Research Program of the NHLBI at the NIH for funding and we thank Dr. Nico Tjandra for his continued interest and support.

References

1. Pennisi, E., "Genomics, ENCODE project writes eulogy for junk DNA." Science, 2012, 337, 1159-1161.

2. Zia, A & Moses, A. M. "Ranking insertion, deletion and nonsense mutations based on their effect on genetic information." BMC Bioinformatics. 2011, 12, 299-313.

3. Gehring, W.J. "Homeo boxes in the study of development." Science, 1987, 236, 1245-1252.

4. Kim, Y. & Nirenberg, M. "Drosophila NK-homeobox genes." Proc. Natl. Acad. Sci. USA, 1989, 86 7716-7720.

5. Koizumi, K., Lintas, C., Nirenberg, M., Maeng, J.-S., Ju, J.-H., Mack, J. W., Gruschus, J. M., Odenwald, W.F., Ferretti, J. A. "Mutations that affect the ability of the vnd-NK-2 homeoprotein to regulate gene expression: Transgenic alterations and tertiary structure." Proc. Natl. Acad. Sci. USA, 2003, 100(6), 3119-3124.

6. Nakayama, K., Nakayama N., Kim, Y. ,Weber, K. O., Lad, R., Nirenberg, M. "The NK-2 homeobox gene of Drosophila" Ann. N. Y. Acad. Sci. 1995, 758, 224-242.

7. Kasahara, H., Lee, B., Shott, J.-J., Benson, D.W. Seidman, J. G., Seidman, C.E., Izumo, S. "Loss of function and inhibitory effects of human CSX/NKX2.5 homeoprotein mutations associated with congenital heart disease." J. Clin. Invest., 2000, 106, 299-308.

8. Scott, M. P., Tamkun, J. W. & Hartzell, G. W. III "The structure and function of the homeodomain." Biochim. Biophys. Acta., 1989, 989, 25-48.

9. Benson, D.W.,Silberbach, G.M., Kavanaugh-McHugh, A., "Mutations in the cardiac transcription factor NKX2.5affect diverse cardiac developmental pathways." J Clin. Invest., 1999, 104, 1567-1573.

10. Plotkin J. B. and Kudia G. "Synonymous but not the same: the causes and consequences of condo bias" Nat. Rev. Genet. 2011, 12(1), 32-42.

11. Gehring, W. J., Affolter, M. & Burglin, T. "Homeodomain proteins." Annu. Rev. Biochem. 1994, 63, 487-526.

12. McGinnis, W. & Krumlauf, R. "Homeobox genes and axial patterning." Cell, 1992, 68, 283-302.

13. Mellerick, D. & Nirenberg, M "Dorsal-ventral patterning genes restrict NK-2 homeobox gene-expression to the ventral half of the central-nervous-system of Drosophila embryos." Dev. Biol., 1995, 171, 306-316.

14. Apergis, G. A., Crawford, N. Ghosh, D. Steppan, C. M., Vorachek, W. R., Wen, P. & Locker, J. "A novel NK-2 related transcription factor associated with human fetal liver and heptatocellular carcinoma." J. Biol. Chem., 1998, 273, 2917-2925.

15. Kimura S. " Thyroid-Specific Transcription Factors and Their Roles in Thyroid Cancer" Journal of Thyroid Research, 2011, Article ID 710213.

16. Ouyang P., Saarel E., Bai Y., Luo C., Lv Q., Xu Y., Wang F., Fan C., Younoszai A., Chen Q., Tu X., Wang Q.K., "A de novo mutation in NKX2.5 associated with atrial septal defects, ventricular noncompaction, syncope and sudden death." Clin. Chim. Acta., 2011, 412(1-2),170-175.

17. Harvey, R. P. "NK-2 homeobox genes and heart development", Dev. Biol., 1996, 178, 203-216.

18. Lyons, G.E. "Vertebrate heart development." Curr. Opin. Genet. Dev., 1996, 6, 454-460.

19. Bartlett, H., Veenstra, G. J. C. and Weeks, D. L. " Examining the cardiac NK-2 genes in early heart development" Pediatr. Cardiol. 2010, 31(3), 335-341.

20.Ornstein D.K., Cinquanta M., Weiler S., Duray P.H., Emmert-Buck M.R., Vocke C.D., Linehan W.M., Ferretti J.A., "Expression studies and mutational analysis of the androgen

regulated homeobox gene NKX3.1 in benign and malignant prostate epithelium." J Urol. 2001,165,1329–1334.

21. Zheng S. L., Ju, J-H.,Chang B-L., Ortner E., Sun J.,Isaacs S. D., Sun J.,Wiley K. E., Liu W.,Zemedkun M., Walsh P. C., Ferretti J., Gruschus J., Isaacs W., Gelmann E. P., Xu J. " Germ-Line Mutation of NKX3.1 Cosegregates with Hereditary Prostate Cancer and Alters the Homeodomain Structure and Function." Cancer Res. 2006, 66(1), 69-77.

22. McGinnis W., Garber R.L., Wirz J., Kuroiwa A., and Gehring W.J. "A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans." Cell 1984, 37 (2), 403-408.

23. Tsao D. H., Gruschus, J. M., Wang L-H., Nirenberg M., and Ferretti J. A. "The Three-dimensional Solution Structure of the NK-2 Homeodomain from Drosophila" J. Mol. Biol. 1995, 251, 297-307.

24. Gruschus J.M., Tsao D.H., Wang L.H., Nirenberg M., Ferretti J.A. "The three-dimensional structure of the vnd/NK-2 homeodomain-DNA complex by NMR spectroscopy1" J. Mol. Biol., 1999, 289, 529-545.

25. Gruschus J., Tsao,D., Wang,L., Nirenberg,M. and Ferretti,J. " Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity." Biochemistry, 1997, 36, 5372–5380.

26. Qian Y., Furukubo-Tokunaga,K., Resendez-Perez,D., Muller,M., Gehring,W. and Wuthrich,K. "Nuclear magnetic resonance solution structure of the fushi tarazu homeodomain from Drosophila and comparison with the Antennapedia homeodomain." J. Mol. Biol., 1994, 238, 333–345.

27. Dekker N., Cox,M., Boelens,R., Verrijzer,C., van der Vliet,P. and Kaptein,R. "Solution structure of the POU-specific DNA-binding domain of Oct-1." Nature, 1993, 362, 852–855.

28. Qian Y., Billeter,M., Otting,G., Muller,M., Gehring,W. and Wuthrich,K. "The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors." Cell, 1989, 59, 573–580.

29. Weiler S., Gruschus J.M., Tsao D.H., Yu L., Wang L.H., Nirenberg M., Ferretti J.A. "Site-directed mutations in the vnd/NK-2 homeodomain. Basis of variations in structure and sequence-specific DNA binding" J. Biol. Chem. 1998, 273(18),10994-1000.

30. Xiang B., Weiler, S., Nirenberg, M. and Ferretti, J.A. "Structural basis of an embryonically lethal single Ala → Thr mutation in the vnd/NK-2 homeodomain." Proc. Natl Acad. Sci. USA, 1998, 95, 7412–7416.

31. Hwang, K-J., Xiang, B., Gruschus, J. M., Nam, K-Y., No, K. T., Nirenberg, M., Ferretti, J. A., " Distortion of the Three-Dimensional Structure of the vnd/NK-2 Homeodomain Bound to DNA Induced by an Embryonically Lethal A35T Point Mutation " Biochemistry, 2003. 10(42), 12522-12531.

32. Chi, Y-I. " Homeodomain Revisited: a Lesson from Disease-causing Mutations" Hum. Genet. 2005, 116(6), 433-444.

33. Cooper, A. "Thermodynamic analysis of biomolecular interactions." Curr.Opinion Chem.Biol., 1999, 3, 557-563.

34. Naghibi, H., Tamura, A. and Sturtevant, J. M. "Significant discrepancies between van't Hoff and calorimetric enthalpies." Proc. Natl. Acad. Sci. USA 1995, 92, 5597-5599.

35. Zwanzig R. "Two-state models of protine folding kinetics" Proc. Natl. Acad. Sci. USA 1997, 94, 148-150.

36. Ghosh K. and Dill K.A. "Theory for Protein Folding Cooperativity: helix-bundls" J. Am. Chem. Soc. 2009, 131(6),2306-2312.

37. Carra, J. H. and Privalov, P. H. "Energetics of folding and DNA binding of the MATa2 Homeodomain." Biochemistry, 1997, 36, 526-535.

38. Gonzalez, M., Weiler, S., Ferretti, J.A., Ginsburg, A. "The vnd/NK-2 homeodomain: thermodynamics of reversible unfolding and DNA binding for wild-type and with residue replacements H52R and H52R/T56W in helix III." Biochemistry, 2001,40(16),4923-4931.

39. Prabhu, N. V. and Sharp, K. A. "Heat Capacity in Proteins" Annu. Rev. Phys. Chem. 2005, 56, 521-548.

40. Fausti, S., Weiler, S., Cuniberti, C., Hwang, K-J., No, K.T., Gruschus, J.M., Perico, A., Nirenberg M., Ferretti, J.A. " Backbone dynamics for the wild type and a double H52R/T56W mutant of the vnd/NK-2 homeodomain from Drosophila melanogaster." Biochemistry, 2001,40(40),12004-12012.

41. Fodor, E., Mack, J.W., Maeng, J.S., Ju, J.H., Lee, H.S., Gruschus, J.M., Ferretti, J.A., Ginsburg, A. "Cardiac-specific Nkx2.5 homeodomain: conformational stability and specific DNA binding of Nkx2.5(C56S)." Biochemistry, 2005,44(37),12480-12490.

42. Sauna, Z. E. and Kimchi-Sarfaty, C. "Understanding the contribution of synonymous mutations to human disease." Natural Reviews-Genetics, 2011, 12, 683-691.

43. Chung, I-M. and Rajakumar, G. , "Genetics of Congenital Heart Defects: The NKX2-5 Gene, a Key Player." Genes, 2016, 7(6), 1-12.

