

Direct Binding of Complementary Symmetric Ligands to DNA Nucleotide Pairs by Means of Transcription Factors

Charles D. Schaper, Ph.D.
Union City, CA 94587 USA
cschaper@transferdevices.com

September 14, 2020

Abstract

To activate gene expression, the initiation of transcription is a highly regulated process involving the interaction of proteins and DNA nucleotides at the promoter site, which consists of a small number of base pairs. As it involves interactions at the atomic scale, it is challenging to determine the mechanism of binding responsible for the great specificity between the amino acid residuals comprising the transcription binding protein and the DNA nucleotides comprising the promoter. Here, a new approach to characterize the transcription initiation process is developed and verified from analysis of comparative pharmacological efficacy data and elemental modeling. The newly developed description of a mechanism for transcription initiation involves the direct binding of small molecule ligands of approximately twenty carbon atoms, which are both structurally symmetric to DNA nucleotides, and also chemically complementary in its functional groups for interaction with the oxygen element at the carbon two position of thymine and with the phosphodiester chain. The results indicate that the activating ligands are transported to the DNA nucleotide promoter site by protein transcription factors, which serve as delivery vectors, for transfer of the ligand to the DNA nucleotide pairs. The ligands examined in this study include the steroid hormones, synthetic steroid molecules, derivatives of vitamin D, and prostaglandins, particularly PGJ2 and 15d-PGJ2. The transcription factors evaluated include glucocorticoid receptors, VDR, PPAR γ , and TBP. Through the developments, it is shown that because of the chemically complementary binding of the ligand to DNA nucleotide pairs, the resultant intermolecular complex produces three hydrogen bonds for the A-T and T-A configurations, which matches that of G-C and C-G. The orientation of the nucleotide base pairs is also seen to adjust as an inversion of the nominal position of the nucleobases to a dimer configuration presented via TBP transcription factor. The developments comprise a new approach to characterizing the initiation of the transcription process comprising the direct binding and interaction of ligands with DNA nucleotides as verified through comparative analysis of pharmacological activity and through perfect structural correspondence between the steroid hormone class as ligands with Watson-Crick DNA nucleotide pairings.

1 Introduction

To initiate the gene transcription process as a well controlled process, there is a very precise interaction at the so-called promoter region between DNA sequence elements and specific segments of amino acid residuals of a primary transcription factor, which itself is regulated by a myriad of other protein transcription factors and co-activator [1]. As the DNA nucleotides comprising the promoter regions for many genes have been identified, forming a large database [2], the corresponding transcription factors have also been extensively examined [3, 4, 5, 6]. However, even though there is significant understanding of the structure of the interacting proteins and promoter regions, which enable RNA polymerase to promote transcription, the actual binding configuration and associated mechanism of action comprising the highly regulated initiation steps is largely unresolved [7]. Because such binding mechanisms at the promoter region take place at atomic length scales consistent with hydrogen bonding, there is a lack of high resolution images to unequivocally confirm the interfacial interactions and molecular species associated with the inducement and initial steps of the transcription process, which explains why this fundamental and critical molecular binding association is unknown [8].

To identify the molecular mechanism that effectively couples the primary binding protein to the promoter region, recent results are applied that identify the molecular basis of DNA enable new insights into the binding processes associated with the initiation and regulation of the transcription process, as follows: In a series of recent preprints [9, 10] and a research monograph [11], a discovery on the molecular basis of DNA is developed to reveal that there is a seventeen element structure which is common to all DNA nucleotide Watson-Crick base pairings of adenine-thymine (A-T) and T-A, as well as guanine-cytosine (G-C) and C-G. Moreover, this newly identified structure is in perfect harmony with the four fused-ring structure of a steroid molecule, thereby establishing the potential for a symmetric relationship with ligands that also consist of four fused rings, or have similar positional arrangements of its composite functional groups. Moreover, not only is this relationship relevant to transcription, this discovery and concept forms the premise in the development of mechanisms for the design of DNA and the biological cell, as described in my recent research monograph on the origin of life [12].

In developing a transcription mechanism, the driving force must be established so as to overcome the stabilizing forces associated with the hydrophobic core of DNA to nitrogen base pair stacking [13]. Thus, it is required that there is both symmetric and complementary correspondence between DNA nucleotides and potential ligands for interaction. In the method described herein, it is shown that this consistency of symmetry and complementary interaction, corticosteroids are configured perfectly to hydrogen bond to thymine, which has an unpaired oxygen element; and androgen steroids are configured perfectly for cytosine. Thereby, intermolecular complexes are feasible between small molecule ligands and DNA nucleotides that are not only symmetric in terms of ring orientation, but also complementary in the formation of three internal hydrogen bonds and binding onto the

phosphodiester strands. Thus, the hydrophobic nature of the core of DNA is improved by coupling ligands similar to the steroid hormones which include other cholesterol derivatives and prostaglandins, both of which can be configured in a similar manner for direct integration with DNA nucleotides. This integration through binding directly onto the DNA nucleotides enables the very specific interaction with the promoter region of the gene transcription process. To facilitate the delivery of the ligand for direct binding onto the DNA nucleotide pairs, the transcription factors serve as delivery vectors for integration with DNA of the ligand, enabling its transport to the initiation site, and ultimate transfer to specific promoter base pairs.

As the development of a new mechanism for transcription requires experimental validation on the atomic scale, to confirm the approach, pharmacological efficacy data is utilized. Such data is readily available as often relative activity is measured for single atom differences, and it is relevant since the molecular structures of the ligands involved in transcription by directly binding onto DNA nucleotides are too small to image directly. This quantitative analysis is established by systematically by altering just a single functional element at a time and observing the response, and demonstrate through molecular modeling a consistency between the DNA nucleotide - ligand intermolecular complex. Thus, a predictable association of ligand and promoter site is established, which has significant potential in drug design, as well as enabling a fundamental analysis of altered transcription rates resulting in dysfunction inducing signs and symptoms of disease [14]. The use of pharmacological information is critical at identifying the binding mechanisms, as crystallographic information is unavailable at atomic level resolution [15], and the molecular structure analysis is subject to hypothesis testing and data fitting [16, 17], and thus requires a preconceived notion of what to look for, and therefore runs into deficiencies in the discovery process. It is acknowledged that the mapping of transcription factor proteins is difficult to understand under consensus transcription theories [18, 19, 20, 21, 22].

To develop this newly introduced mechanism of action for the interfacial interactions between ligands and DNA nucleotides, in which the ligands are delivered by transcription factors, the organization of the manuscript begins with a molecular analysis of DNA nucleotide pairs to develop classes of ligands that are compatible both in symmetry and complimentary in terms of hydrogen bonding of key elements of the molecular structure of DNA nucleotides, and the hydrophobic and hydrophilic characteristics of the core and periphery. The delivery vectors for transporting, aligning, and transferring the ligands to precise locations on the DNA are then examined. The transcription factor associated with the TATA box is assessed based on experimental data from the literature. Pharmacological efficacy and atomic-level molecular modeling is deployed to verify the approach.

2 Results

In this section, to enable the newly introduced mechanism for transcription, structural and molecular characteristics are defined for (1) ligands in terms of their ability to bind onto DNA nucleotides, and (2) transcription factors for their ability to serve as a delivery vector for such DNA-binding ligands. A critical aspect is that the ligands are both symmetric with the ring structure and arrangement of DNA nucleotides, as well as chemically compatible with DNA nucleotides with regard to their binding capability to Thymine, and to the phosphodiester backbone. Several examples of ligands and transcription factors are defined which are capable of this developed mechanism of action. These results thereby address the paradox of the well-defined specificity of the promoter as the transcription initiation site, and provide a basic mechanism of action for transcription initiation to start the central dogma of molecular biology.

2.1 Ligands Symmetric and Complementary for Binding onto DNA Nucleotide Pairs

To develop the new mechanism for transcription, a fundamental result begins the analysis on the basic structure of DNA nucleotides and the characteristics of ligands that are capable of directly coupling with the DNA nucleotide pairs. I first presented these results in the preprints [9, 10, 23] and my books pertaining to the design of DNA and the genetic codes, and the origin of life [11].

A fundamental result is presented in Figure 1. In Figure 1(a) the nucleotide pairing of cytosine to guanine indicates that there is hidden embedded structure of a steroid molecule, which is obtained by tracing 17 elements of carbon, nitrogen, or oxygen atoms. In Figure 1(b) the nucleotide pairing of thymine to adenine also has the hidden steroid molecular configuration obtain by precisely tracing 17 elements. Note for each the projections off of the C-G and T-A pairing which are not consistent are skipped in sketching the steroid structure. This is an important result as it will define a point for differentiation between the two structures. In Figure 1(c), the steroid structure of testosterone is indicated, there is a correspondence with the cytosine to guanine pairing, relative to Figure 1(d), which is of the steroid structure of cortisol, which has a correspondence to thymine to guanine. By coupling testosterone-like structures, which do not have a functional group on the C-ring, to C-G/G-C and by coupling cortisol-like structures, which have a functional structure on the C-ring, to T-A/A-T, there will develop through intermolecular binding, an equivalency of three hydrogen bonds for all of C-G/G-C/T-A/A-T if the hydroxyl group on the C-ring of cortisol-like structures is paired to the ketone group on Thymine. Thus the integration of testosterone to C-G will have a match to that of cortisol to T-A in terms of number of hydrogen bond coupling of the nucleotides. Therefore, the molecular basis of DNA is the steroid molecule.

A schematic of the structural and chemical relationship of this intermolecular coupling of steroid hormone as a symmetric and complementary ligand directly bound to DNA

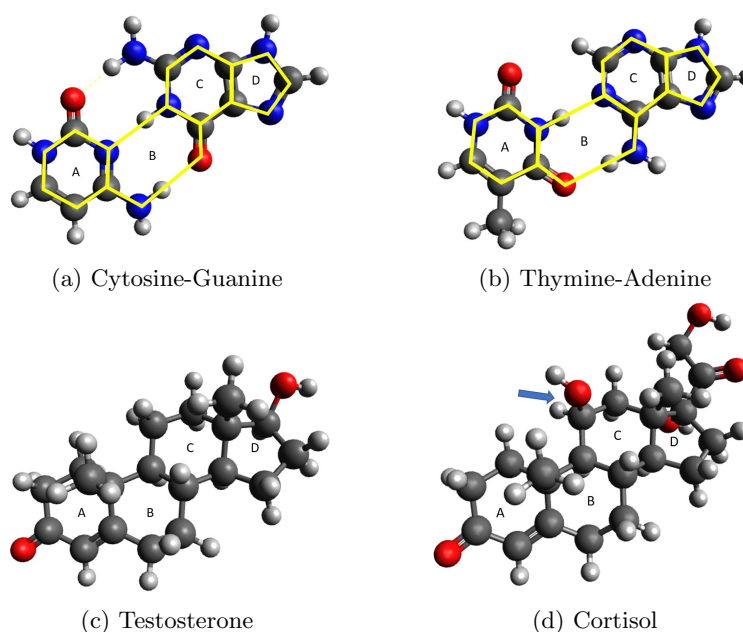


Figure 1: This structural depiction presents a fundamental result as it establishes a molecular basis for DNA, which I first presented in the preprints [9, 10, 23]. In (a) the nucleotide pairing of cytosine to guanine indicates that there is hidden embedded structure of a steroid molecule, which is obtained by tracing 17 elements of carbon, nitrogen, or oxygen atoms, as shown in yellow. The B ring is slightly stretched because hydrogen bonding forms the bond, while covalent bonding exists between the other atoms; (b) The nucleotide pairing of thymine to adenine also has the hidden steroid structure; (c) comparing a steroid structure of testosterone, there is a correspondence with the cytosine to guanine pairing; (d) comparing to a steroid structure of cortisol, there is a correspondence to thymine to guanine, which will result in an equivalency of three hydrogen bonds if the hydroxyl group on the C-ring is paired to the ketone group on Thymine. Thus the integration of testosterone to C-G will have a match to that of cortisol to T-A in terms of number of hydrogen bond coupling of the nucleotides. Therefore, the molecular basis of DNA is the steroid molecule.

Nucleotides is presented in Figure 2(a) and (b). The representative steroid hormones cortisol and testosterone are arranged such that each of the four nucleotide pairings are related to a steroid hormone to produce a set of three hydrogen bonds in an intermolecular complex. As indicated in the overlays, since the C-G and G-C pairing already have three internal hydrogen bonds, testosterone is paired; and since A-T and T-A only have two internal hydrogen bonds, cortisol is paired since in an amazing development the mid-section functional group on the C-ring of cortisol must couple with the unpaired oxygen element of Thymine at the carbon two position to form an intermolecular hydrogen bond.

Therefore, this relationship between steroid hormone defines not only a structural corre-

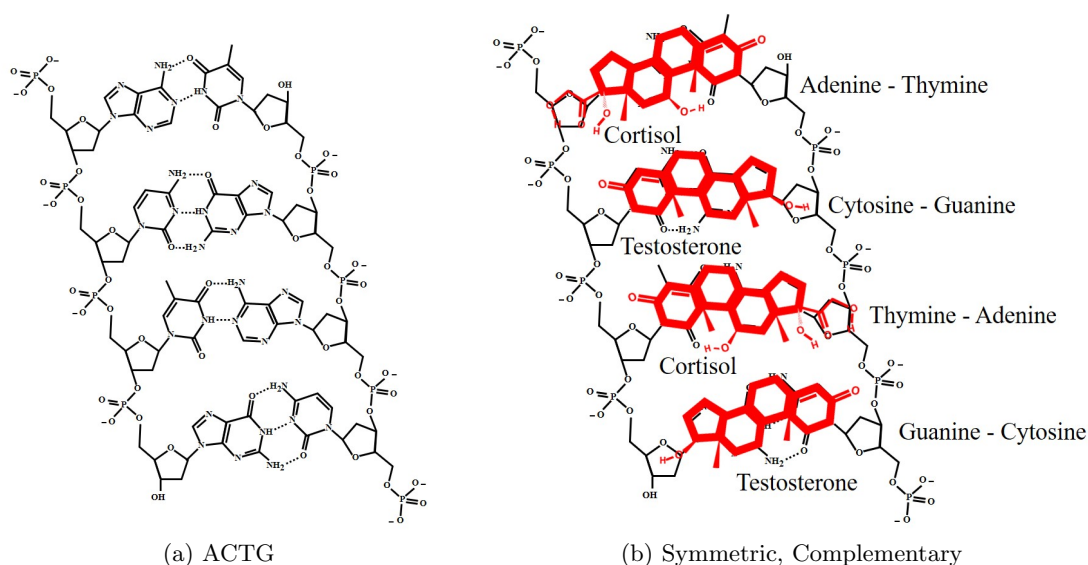


Figure 2: Schematic of direct binding of symmetric and complementary ligands onto DNA nucleotide pairs. There is a precise correspondence across 17 atoms of the cortisol steroid hormone with the A-T and T-A pairs and with testosterone steroid hormones with the C-G and G-C pairs. This association, including the intermolecular bonding, enables three internal hydrogen bonds for each of the DNA nucleotide arrangements as the mid-molecule C-ring OH group of cortisol-like molecules forms an intermolecular hydrogen bond with the unpaired oxygen of Thymine at the carbon two location. Moreover, to achieve the configuration for each of the four pairs, as indicated in that there is a front-side and back-side orientation in order to get the three hydrogen bonds, and to have the proper orientation of the methyl groups and hydroxyl groups such that there is an integration of ligand with DNA. This result is extensible to other steroid molecules with similar configurations of functional elements to cortisol or to testosterone.

spondence, but also a complementary correspondence, in terms of chemical compatibility. The binding to the unpaired functional group of Thymine will improve the interior hydrophobic configuration of the DNA molecule. The hydrophilic nature of the ends of the steroid molecule are available to bind to each of the phosphodiester backbone. Consequently, there is a complementary chemical correspondence between the steroid hormones, which will act as ligands, and the DNA nucleotide pairs, in addition to a symmetric structural correspondence.

Moreover, it is noted that in order to get the structural match for each of the pairings, the T-A and C-G pairings were required for the hormone to be matched on the “back” of the molecule to its normal presentation. As there is a front and back correspondence in order to structurally match the steroid hormone and the DNA Nucleotide base pairing, there is a defined orientation to the association. Thus, the combination of the orientation

(front/back) and class (capable or not capable of forming a hydrogen bond with Thymine) of steroid hormone forms a code for DNA sequences as shown in Table 1. This is of general interest, and will be important for assessing transcription initiation sites when coupled to ligands that are symmetric and complementary to DNA nucleotide pairs.

Pairing	Orientation	Class
A–T	\mathcal{F}	\mathbb{H}
T–A	\mathcal{B}	\mathbb{H}
C–G	\mathcal{B}	\mathbb{S}
G–C	\mathcal{F}	\mathbb{S}

Table 1: DNA Base Pairing Code indicating the location of the ligand structure, \mathcal{F} is front and \mathcal{B} is back, and the corresponding class of structural positioning of the hydroxyl group representing, \mathbb{H} signifies the class of cortisol-like molecules, which has a functional group positioned to interact with the available functional group of Thymine, and \mathbb{S} signifies the class of testosterone-like molecules, which do not have a functional group positioned that can interact with the available functional group of Thymine. This general capability of class and orientation will be extended to other ligands.

In Figure 3 other types of ligands are introduced that are synthetic in nature, but contain the same basic structural requirements. The structures separate the steroid molecules into two types: Class \mathbb{S} which interacts with the C–G and G–C nucleotides, and Class \mathbb{H} which interacts with the A–T and T–A nucleotides. Class \mathbb{S} has functional groups on either for binding onto the phosphodiester backbone and have a hydrophobic interior; whereas, Class \mathbb{H} have functional groups on either end, as well as a functional group in the interior for binding with Thymine.

Thus, the ligands that have capability of binding to DNA nucleotides need not be steroid hormones, and can be synthetic agents. In the following subsection, different types of ligands will be developed based on their pharmacologic activity of inducing transcription. It will be seen that the same basic requirements exists of hydrophobic ends, hydrophilic internal section consistent with thymine or cytosine, and the structural symmetry such that the hydrophobic characteristics are aligned with the ring structure of the DNA nucleotides.

2.2 Pharmacological Activity of Ligands

To identify ligands that are symmetric and complementary to DNA nucleotides, as well as to support its premise, comparative pharmacological efficacy is evaluated. One of the requirements for a ligand that directly bind onto the DNA nucleotide pairs is that it must have a potency, and thus to validate the methodology it is possible to examine molecules which are close to in structure, but differ by one or two atoms and have pronounced differences in potency. This differences are examined in the ability of the ligand to bind onto the phosphodiester backbone, and to Thymine. By showing molecules that differ

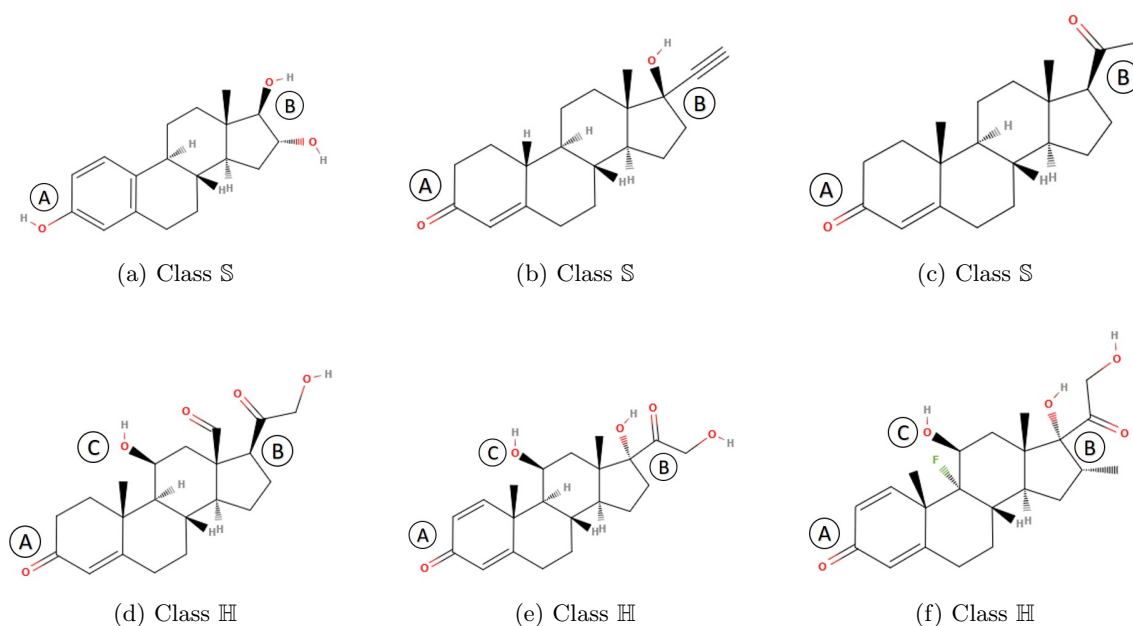


Figure 3: Separating the steroid molecules into two types: Class S which interacts with the C-G and G-C nucleotides, and Class III which interacts with the A-T and T-A nucleotides, in which Class S has A and B functional groups for binding, and Class III has A and B functional groups, as well as C functional group for binding with Thymine.

by just a single element, but if done in the appropriate position, a very large difference in transcription rates result, then it identifies the ligand, as well as provide confirmatory evidence of the approach developed in this paper.

For example, a clear-cut experimental agreement of this mechanism to experimental response comes from comparing the efficacy of prednisone to prednisolone, which are synthetic molecules introduced to have a more powerful effect than cortisol. Prednisone as seen in Figure 4(a) can not form a hydrogen bond with Adenine-Thymine, specifically the oxygen element of Thymine at the carbon two location; thus prednisone is not active, and has to be converted to prednisolone in order to be active, and it is five times more active than cortisol. As seen in Figure 4(b), Prednisolone can form an intermolecular hydrogen bond with the oxygen element of Thymine. These results are extendable to synthetic steroids, including corticosteroids such as dexamethasone and hydrocortisone, which, due to their complementary nature of binding onto Thymine with a functional group on the C-ring, and therefore will bind onto A-T. It is also consistent with the anabolic steroids, which do not have a functional group on the C-ring, and thus will bind onto C-G pairs. The relative potency of the steroids, both corticosteroids and anabolic, can be examined based on the quality of the binding with the phosphodiester backbone and with Thymine, in the case of

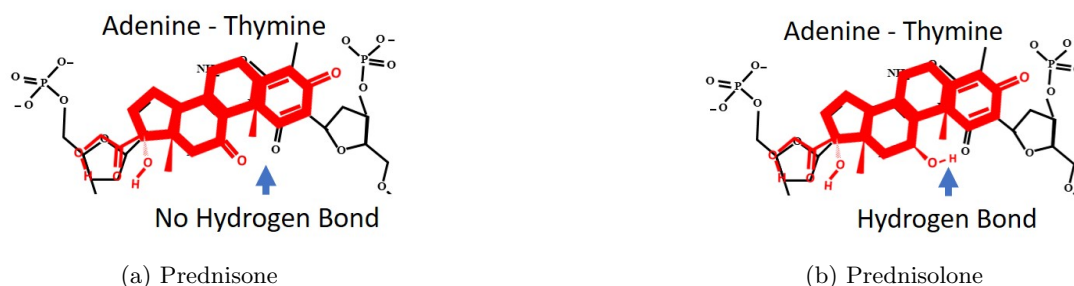


Figure 4: Potency of synthetic steroids indicating the importance of the binding of the functional group of the C-ring to Thymine: (a) Synthetic steroid Prednisone is not able to form a hydrogen bond of an intermolecular nature with the oxygen element of Thymine, and is thus inactive. In (b) synthetic steroid Prednisolone is able to form an intermolecular hydrogen bond with Thymine, and is thus active. Prednisone must be converted to Prednisolone in order for it to work. Moreover, as further indication of the importance and predictability of the interaction, prednisolone is five times more powerful than cortisol, likely because of the double bond in the A-ring enables a stronger bond to Thymine.

corticosteroids.

Now turning the attention of a different type of steroid, cholesterol, which is the most abundant steroid in the human body, which is shown in Figure 5(a). Related to cholesterol is Vitamin D, which is shown in Figure 5(b). Both cholesterol and Vitamin D are inactive in terms of inducing transcription. This can be explained in that although it is symmetric with the C-G and G-C pairs, it is not complementary to binding with the phosphodiester backbone on both strands. However, the an active form of Vitamin D: when exposed to ultraviolet light, a result is Vitamin D3, which then is converted to calcitriol, as indicated in Figure 5(c), through metabolic pathways. As shown, the main conversion site is in applying a functional group such that both sides are capable of binding with both strands of DNA, while retaining their lipid, hydrophobic characteristic in the center. However, the four fused ring of the steroid molecule is broken, and thus the ability to induce transcription is less effective than it would be for testosterone-like materials for binding onto C-G and G-C. Because of a lack of functional group in the interior, it would thus be not complementary with the A-T and T-A pairs.

Unique among the hormones are the prostaglandins, which are capable of inducing transcription, but in unique ways relative to the steroid hormones. Moreover, slight alterations in structure can produce different effects. The reason for this within the context of the developments described herein is as follows: In Figure 6 (a) the structure of Prostaglandin J2 (PGJ2) is indicated, which has functional groups of A, B, and C. Therefore, PGJ2 is capable of interaction with T-A and A-T nucleotides, but due to the non-fused nature of the internal molecules, the binding would be weaker than that of cortisol. In Figure 6(b), the structure of 15-deoxy Prostaglandin J2 (15d-PGJ2) which except for a hydroxyl group,

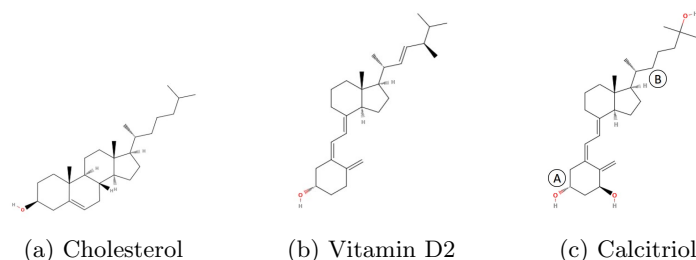


Figure 5: The structures of (a) Cholesterol and (b) Vitamin D2, which are inactive in terms of transcription. Both have just a single functional group which can only interact with one phosphodiester strand even though it has symmetry with the C-G nucleotide pair; (c) To activate Vitamin D, it must convert to Calcitriol, which in our analysis indicates that there is an A and a B functional group which will enable its interaction with the C-G nucleotide, but not the A-T nucleotide because it lacks a functional group that can interact with Thymine.

has functional groups A and B. 15d-PGJ2 is thus capable of interacting with C-G and G-C nucleotides, and the response would be expected to be different than PGJ2, which is examined in the following subsection in comparison to cortisol.

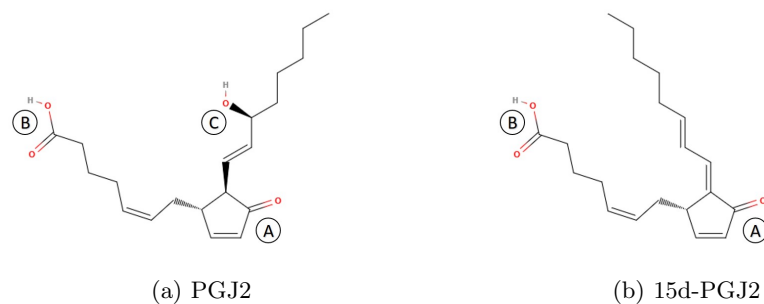


Figure 6: (a) The structure of Prostaglandin J2 (PGJ2) has functional groups of A, B, and C and thus is capable of interaction with T-A and A-T nucleotides; (b) The structure of 15-deoxy Prostaglandin J2 (15d-PGJ2) only has functional groups A and B, and is thus capable of interacting with C-G and G-C nucleotides.

In Table 2, examples of ligands are presented and organized as associated of Class S which couples with C-G and of Class H which associates with A-T. The ligands need not have the four fused structure of a steroid hormone, but has to have certain requirements in terms of capability of not only symmetry but also complementary to the Thymine, via a hydroxyl group, or Cytosine, via an absence of a hydroxyl group. The internal structure will require a driving force that will make the internal portion of DNA less hydrophilic, that is more hydrophobic. Using this characterization, it is thus possible to design synthetic

ligands to perform unique transcription functions. C-

Class S	Class H
Testosterone	Cortisol
Estriol	Aldosterone
Progesterone	Dexamethasone
Calcitriol	Prednisolone
Deoxy-15 PGJ2	PGJ2

Table 2: Examples of ligands that are associated of Class S which couples with C-G and of Class H which associates with A-T. Note that the ligands need not be a steroid hormone, but has to have certain requirements in terms of capability of not only symmetry but also complementary to the Thymine or Cytosine. The internal structure needs to have a driving force that will make the internal portion of DNA less hydrophilic, more hydrophobic.

2.3 Cytosolic Receptors as Delivery Vectors for Binding

In addition to the requirement of compatibility, both symmetry and complementary, of the ligand with the DNA nucleotide pairing, it is also necessary for the ligand to be compatible with a delivery vector that can transport and enable the binding of the ligand onto the DNA nucleotide pairs. In this subsection, protein structures that are nominally considered as receptors of the cytosol are assessed in their capability for delivery of the ligand to the nucleotide pairs. In the standard consideration, the ligand, such as a steroid hormone, never depart the receptor during the transcription process; however, in this analysis, I will consider the situation in which the receptor is used as a temporary storage device for the ligand, and after associating with the DNA strands, presents the ligand for transfer to the DNA nucleotide strands for binding.

In Figure 7(a), cortisol is depicted surrounding by the amino acid groups comprising a ligand binding domain, with hydrogen bonding in combination with an ion as the mechanisms to keep the unit secured, which I originally developed in evaluating competitive inhibition of cortisol at the ligand binding domain [14]. With the introduction of the calcium ions, a significant shift in the energy levels ensues, with Cortisol- Ca^{2+} -LBD stabilizing by 113% to $-1,888$ kJ/mol. It is sensible for cortisol-LBD to be less energetically favorable than the LBD alone, as it makes it possible for cortisol to release from the LBD. The electrostatic potential of the overall structure of the cortisol- Ca^{2+} -LBD complex is shown in Figure 7(b). In addition to the stability enabled by calcium ions within the cortisol-LBD complex, it also provides electrostatic potential organization, which is useful for interaction and alignment with other biomolecules. After association with ions, the steroid hormone-receptor complex is ready to proceed with interaction with the DNA molecule. After passage into the nucleoplasm, the steroid hormone-LBD receptor complex aligns through electrostatic interaction and steric effects with the DNA molecule using the

phosphate groups as the alignment mechanism.

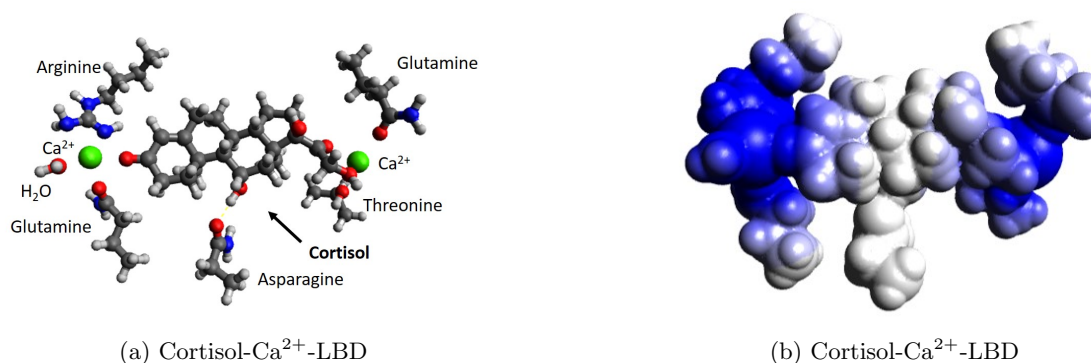


Figure 7: For a ligand binding domain of cortisol with the glucocorticoid receptor, (a) the configuration is indicated which is just the product of hydrogen bonding among the five functional groups of cortisol and ionic coupling with calcium (green) at each end with the ligand binding domain, and with the functional group of the C-ring; (b) The electrostatic potential of the coupling indicates a positive charge at each end due to the coupling with Ca²⁺.

Upon alignment of the steroid hormone complex with the DNA complex, the LBD motions away from the hormone complex presentation to the DNA complex, and in this study, the cortisol molecule and A-T configuration is assessed to exemplify the discussion. After interaction of cortisol with Ca²⁺ ions with the PO₄⁻ ions of the A-T base pairing, thus forming a strong ionic bond, the result is shown in Figure 8(a). The structural integration is noted that the steroid hormone molecule is attached laterally, bridging and strengthening the overall complex, with the hydrogen bonding still retained for A-T. The calcium ions have stabilized at the mid-way point of the phosphate groups and the lateral separation is similar to what was present with the LBD prior to integration. In Figure 8(b) the electrostatic potential map is indicated of the combined structure, and it is noted that there is a charge normalization, and a smaller positive charge at the calcium ions is present on one side of the molecule. The lateral distance between the calcium ions is 16.733 Å; the upper phosphate ions is 16.036 Å; and the lower phosphate ions is 17.039 Å. According to the molecular dynamic simulation, the energy is -7,158 kJ/mol, compared to -1,162 kJ/mol without the cortisol/Ca²⁺ complex, indicating significant stabilization.

To assess the stability and configuration of the Ion-Ligand-DNA complex, the attachment points for the cortisol example are analyzed. In Figure 9(a), the left end group is shown where the calcium ion is positioned at the mid point of adjacent phosphate groups and also interacts with the oxygen-based functional group that is attached to cortisol in Ring A, thus connecting and stabilizing the group at one side. At the mid-point, as shown in Figure 9(b), hydrogen bonding connecting the unmatched Thymine functional group to

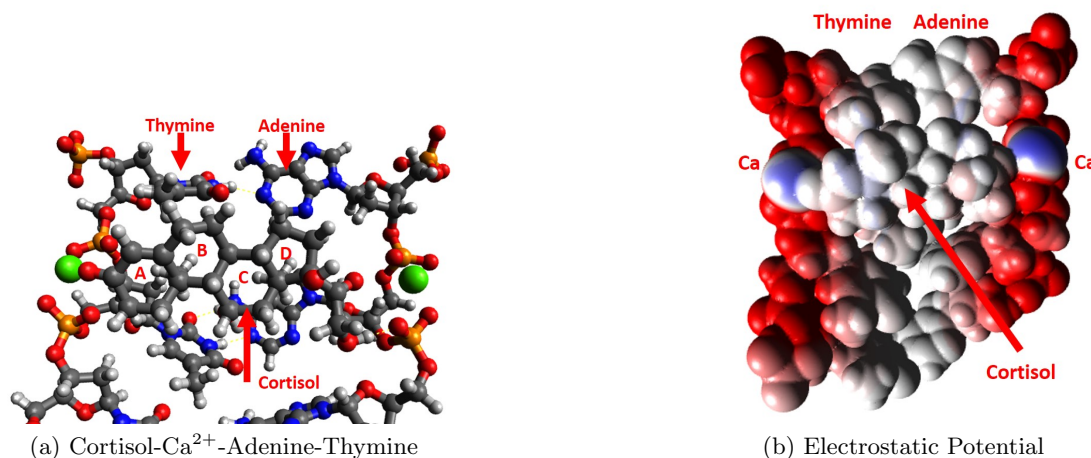


Figure 8: After integration of the cortisol molecule with calcium ions, and release from the LBD, the interaction with the Adenine-Thymine pairing is shown in (a). It is noted that the calcium ions have stabilized at the mid-way point of the phosphate groups (shown in orange) and the lateral separation is similar to what was present with the LBD prior to integration. In (b) the electrostatic potential map is indicated of the combined structure in which a small positive charge from the calcium ions is maintained, and overall the charge is distributed at the point where the integration was made.

the oxygen-based functional group on the C-ring of the cortisol molecule. The distance is dependent upon the orientation of the hormone and DNA structures, including the phosphate groups. The position of the hydrogen bonding between the steroid and the thymine ranges between 2 Å and 3 Å in the simulations conducted, and is 2.744 Å in the configuration shown, which is larger than desired for a strong hydrogen bond, but can be closed with the addition of the water molecule available from the ligand binding domain entry. The distance between the oxygen elements is 3.562 Å. The distance between adjacent phosphate groups is 4.860 Å at the right side and 4.734 Å at the left; The distances between the cortisol functional group on the left side to the calcium, upper and lower phosphate ions, are 2.495 Å, 4.216 Å, and 4.513 Å, respectively; and on the right side 3.137 Å, 2.876 Å, and 3.192 Å. Before the association of the cortisol/Ca²⁺ complex, the distance of the phosphate groups of left and right was 7.075 Å and 7.110 Å, with lateral distances of 16.883 Å upper, and 17.540 Å lower. This indicates a compression of the phosphate ions perpendicular to the nucleotides to accommodate the steroid molecule, creating a strong but constrained structure, which can be examined further in Figure 9(c), the right-side end group is indicated where the second calcium ion is also positioned in the mid-point of the phosphate groups of adjacent and is also positioned in close proximity to a functional group of cortisol at the A-ring of cortisol. It is also noted that the alcohol group of cortisol, that is O1 at C1, is in hydrogen bonding with one of the phosphate groups on a row, thus

adding to the integration positioning and stability of the overall complex.

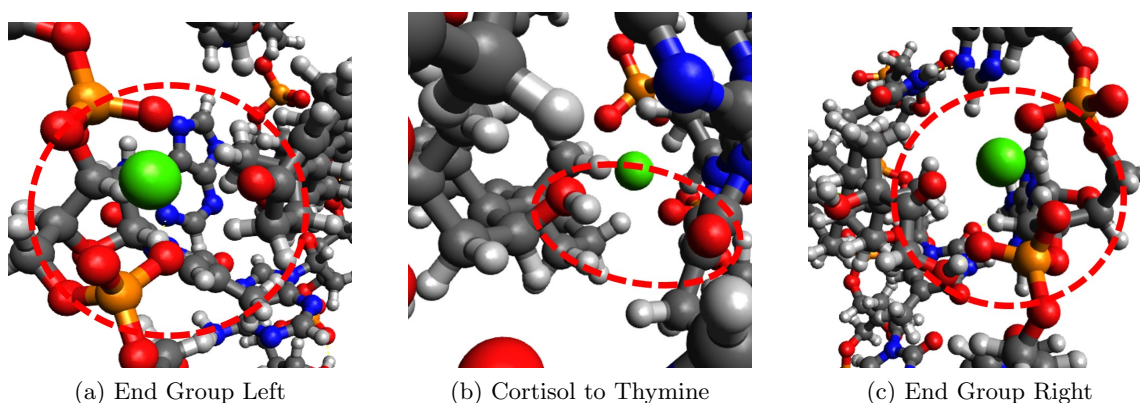


Figure 9: The attachment points for the cortisol- Ca^{2+} complex is shown in (a) at the left end group where the calcium ion is positioned at the mid point of the phosphate groups and is interacting with the oxygen-based functional group attached to cortisol; (b) the mid-point where the unmatched functional group of the A-T pairing is associated by hydrogen bonding with the middle functional group of cortisol; (c) the right end group where the second calcium ion is also positioned in the mid-point of the phosphate groups and is also in proximity to a functional group of cortisol. It is also noted that the alcohol group of cortisol is in hydrogen bonding with one of the phosphate groups.

In this approach, it is interesting to determine the interaction of adjacent calcium along the DNA backbone. It is first noted that the TATA Box configuration, which is commonly associated with gene expression, is feasible as represented in Figure 10. The attachment points of the steroid hormone with the ion coupling element to the phosphate backbone is stable. While the TATA box is feasible since there is separation of four base-pairing units between the binding of the steroid hormone, if two ligands were to bind next to each other with a positive two charge coupling agent, ionic repulsion would result and binding would not be possible. This result is consistent with the base sequences experimentally determined for transcription regulatory regions. The calculations and developments were performed for the calcium ion, but it is noted here that the magnesium ion is also suitable for interaction. Calcium was selected because of its reserves are significant, it is large enough to permit multiple element interaction, the cellular pumps available to keep calcium at relatively low concentrations in the cell and thus enable its removal from the cell after gene expression, its lower hydrated radius than magnesium. However, magnesium ions also produce favorable energetic stabilization, a bit better than calcium and the size of the element is consistent with the other elements comprising DNA. Thus it is feasible that in addition to calcium, other positive two charged ions may participate in a similar manner in the transport and interaction of the steroid hormone with DNA. Similar results are obtained for other positive two charged ions Mg^{2+} , Zn^{2+} , Mn^{2+} .

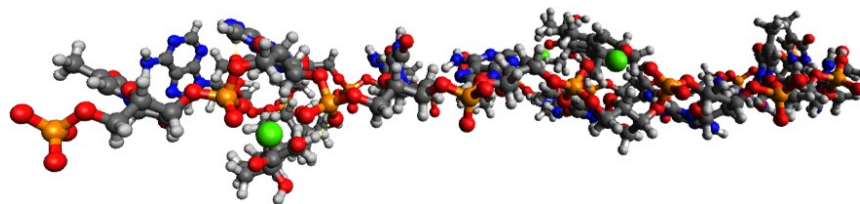


Figure 10: In the TATA configuration with 2 base pair (bp) gap, the arrangement and binding of cortisol is feasible as the calcium ions are stable, securing the steroid hormone to the phosphodiester backbone, as well as coupling the cortisol molecule to Thymine, and achieving symmetry with the purine-pyrimidine configuration.

In addition to ionic binding to stabilize the structure of charge positive two, other ions are feasible at position one charge, including sodium, as indicated in Figure 11(a) and (b). In this case, the density of attachment sites for the promoter can increase to each base pair associated with an ionic charged element to stabilize along multiple sites. This eliminates the issue of ionic repulsion seen with positive two charge ions such as calcium, magnesium or zinc, however, it is a more complex solution involving the positioning of multiple ligands and ions at adjacent positions.

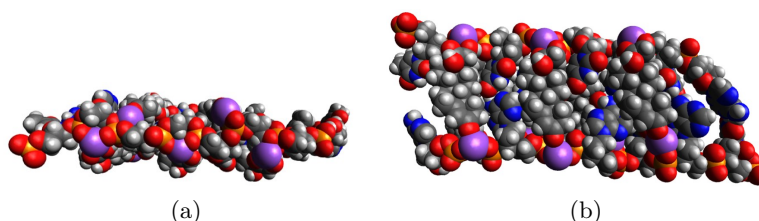


Figure 11: (a,b) Arrangements of sodium ions arranged to secure cortisol for each base pair of the ATATA chain, including internal complementary intermolecular binding with thymine.

Turning the attention now to C-G pairing, in Figure 12(a), the binding configuration is depicted of Calcitriol with the C-G pairing using Mg^{2+} on either side to enhance the coupling. In Figure 12(b) the C-G binding of testosterone onto the C-G pair is indicated. Calcitriol is transported to the DNA nucleotide by the VDR (Vitamin D receptor) and testosterone by the AR (androgen receptor). The binding to the DNA nucleotides with testosterone is significantly greater affinity than that of calcitriol, and thus is a stronger effect in transcription rate and duration of testosterone. The procedure of transport, delivery and separation described for the glucocorticoid receptor would be consistent with that of VDR and AR to deliver their ligands.

For PGJ2 and 15d-PGJ2, which is associated with the peroxisome proliferator-activated receptor- γ (PPAR γ), a comparative analysis is presented first with receptor to cortisol in

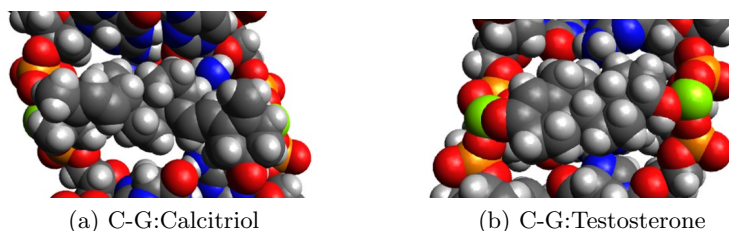


Figure 12: (a) Calcitriol binding onto C-G pairing with Mg^{2+} on either side to enhance the coupling. (b) This can be compared with testosterone binding onto the C-G pair. Note that the appearance of the binding is more condensed with testosterone than with calcitriol.

terms of structure and chemical complementary. In Figure 13(a) cortisol is presented for analysis with Figure 13(b) PGJ2 and Figure 13(c) 15d-PGJ2. The black arrows indicate that the functional group and two extending carbon sections are equivalent between three molecules. The chains are at the equivalent location and thus would expect to align in similar ways to pyrimidine five-carbon group of either adenine or guanine. The blue arrow indicates and equivalency in positioning of the functional group for association with the phosphodiester backbone. The green arrow indicates an equivalency of positioning of the hydroxyl group for interaction with cortisol between 15d-PGJ2 and cortisol. Note that there is exceptional matching between the three structures in terms of the interaction potential with DNA nucleotide pairs. Therefore, PGJ2 would be expected to interact with C-G and G-C, and 15d-PGJ2 would be expected to interact with A-T and T-A.

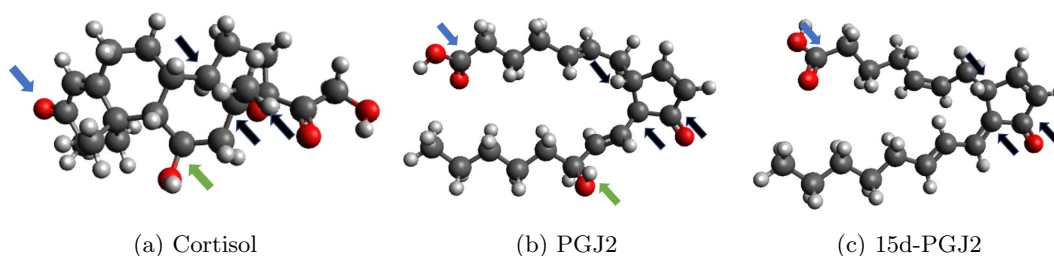


Figure 13: (a) Comparative analysis of cortisol, (b) PGJ2 and (c) 15d-PGJ2. Black arrows indicate that the functional group and two extending carbon sections are equivalent between three molecules. Blue arrow indicates and equivalency in positioning the functional group. Green arrow indicates an equivalency of positioning of the hydroxyl group for interaction with cortisol between 15d-PGJ2 and cortisol. Hence PGJ2 would be expected to interact with C-G and G-C, and 15d-PGJ2 would be expected to interact with A-T and T-A.

The direct coupling, after transport from the receptor, of PGJ2 is shown in Figure 14(a) paired with Mg^{2+} on either side interacting with the C-G nucleotide pairing. In Figure 14(b) because of the functional group able to interact with Thymine, the association of

PGJ2 with T-A pairing is presented. The mechanism for strand separation via endogenous binding is examined further in the preprint [9].

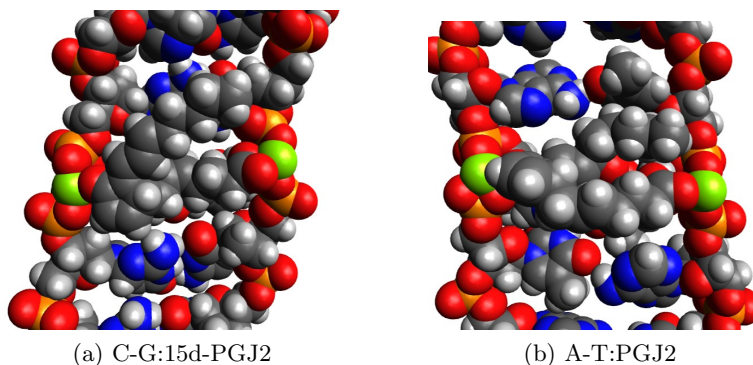


Figure 14: (a) 15d-PGJ2 pairing with Mg^{2+} on either side interacting with the C-G nucleotide pairing. (b) because of the functional group able to interact with Thymine, the association of PGJ2 with T-A pairing.

2.4 Transcription Factors as Delivery Vectors

In addition to the transport of cytosolic receptors, such as the glucocorticoid receptor, serving as delivery vector for the ligand, in this subsection, the use of general transcription factors is examined, starting with the TBP (TATA box Binding Protein). The TATA box involves configurations of A-T and thus symmetric complementary ligands will have the basic structure of cortisol, in which there is a functional group on the C-ring positioned to interact with Thymine. Thus, to examine this potential, the ligand of cortisol is selected for evaluation, which is considered to be recruited for binding onto the DNA nucleotide in the subsequent analysis.

In the paper of [24], the crystal structure was presented of a negative cofactor 2 that was in association with the TBP-DNA transcription complex, of which an identification was estimated from the data of the amino acid positioning with respect to the TATA box binding site. While [24] assumed that the amino acid residuals interacted directly with the DNA nucleotide pairs, here, because of the newly introduced mechanism of interaction, the analysis is re-done considering an intermediate cortisol molecule as a go-between the TBP and DNA complexes. Molecular modeling is performed to show that an equivalent results could be obtained, but from a different perspective and mechanism.

The analysis of Reference [24] estimated of the position of the amino acids of the binding protein transcription factor relative to the TATATTTT initiation site, which had a repressive function in this case. The relative positioning of the amino acids that are active, which is a subset of all of the amino acids surrounding the base pairs, is presented in Figure 15(a).

As it is an objective to assign a receptor like function to general transcription factors, the association of the ligand, here selected as cortisol, is approximated in Figure 15(b). Note that the overall distribution is consistent with the positioning configuration of the glucocorticoid receptor, and the general positioning structure of the perceived positioning of the TBP amino acids relative to the DNA strands.

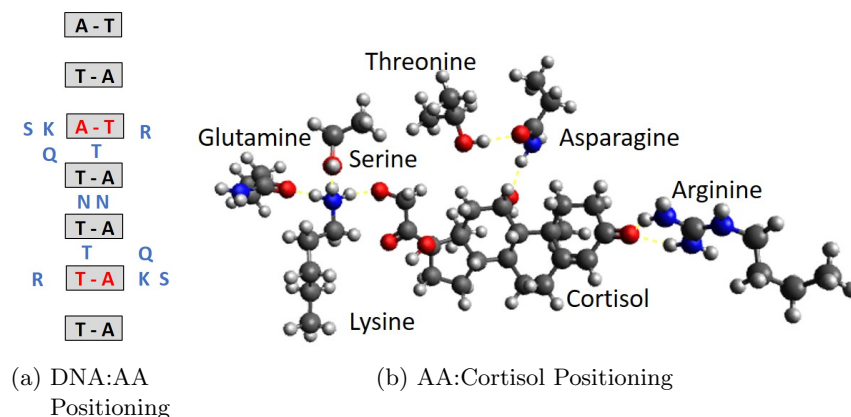


Figure 15: (a) Based on the experimental analysis of Reference [24], the position of the amino acid groups of the binding protein transcription factor relative to the TATATTTT binding site. (b) Using the general positioning of the amino acid groups, I indicate the likely positioning of the amino acid residuals relative to cortisol, used in this example as the binding molecule.

So as to position all amino acids, it is necessary to employ two binding sites, and thus two cortisol molecules, as indicated in Figure 16(a). This results in a precise relative orientation of the amino acid residuals relative to the estimated positioning of the crystal study. Note that in this configuration, the cortisol molecule has been transferred to bind onto the DNA nucleotide pairs (here shown with Mg^{2+} ions) rather than its initial site, which is the binding pocket of the TBP. Moreover, in order to get this configuration, it is necessary to flip the T-A pairing (denoted as $T^{-1}-A^{-1}$) so as to couple with the intermolecular bond of cortisol to Thymine, as indicated in the molecular model of Figure. In Figure 16(b) the side opposite to that of the binding sites by the two molecules of cortisol is shown to indicate, and in Figure 16(c) the space filling model of the same side binding of the two molecules of cortisol and amino acid residuals of the transcription factor is presented.

The electrostatic potential is presented on the sides opposite of the binding sites as indicated in Figure 17(a). It indicates neutral core of the DNA molecule, in which it is hydrophobic. This is maintained on the side at which the binding takes place, as indicated in Figure 17(b).

With the direct binding of ligands to DNA molecules via transcription factors, including

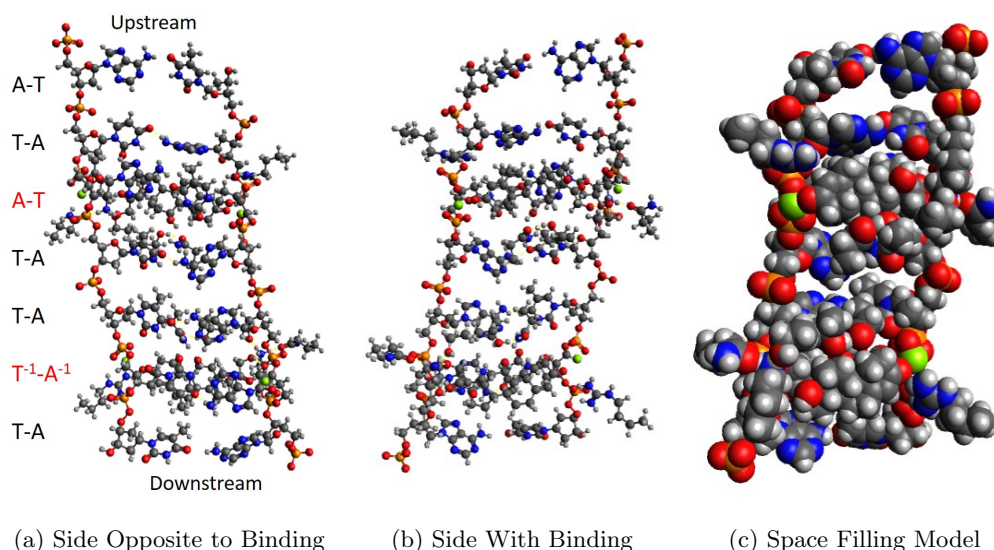


Figure 16: (a) In order to achieve the proper configuration, it is necessary to flip the T-A pairing ($T^{-1}-A-1$) so as to couple with the intermolecular bond of cortisol to Thymine; (b) The side opposite to that of the binding sites by the two molecules of cortisol; (c) The space filling model of the same side binding of the two molecules of cortisol and amino acid residuals of the transcription factor.

steroid and other cytosolic receptors, to initiate gene expression, a mechanism of action can be developed involving a sequence of steps. The procedures described above add the critical step of linking a ligand (steroid, vitamin, prostaglandin, and others) to the transcription factor, which is then transferred to the DNA nucleotide pair in order to activate transcription. Consequently, the transcription factors effectively act as a receptor, which interacts with a ligand and brings it into close proximity with the DNA nucleotide pair. It is then aligned and separated from the transcription factor, binding onto the phosphodiester backbone and if available, the thymine functional group, to initiate the transcription process. The transcription process will largely be dictated by the quality of the bond between the transferred ligand and the nucleotide pairing. This quality may be influenced by the transcription factor.

3 Discussion

Strong support for a transcription process involving the direct binding of ligands that are both symmetric and complementary to DNA nucleotide pairs comes from the predictive capability of pharmacological efficacy. In particular, validity is shown in that the synthetic

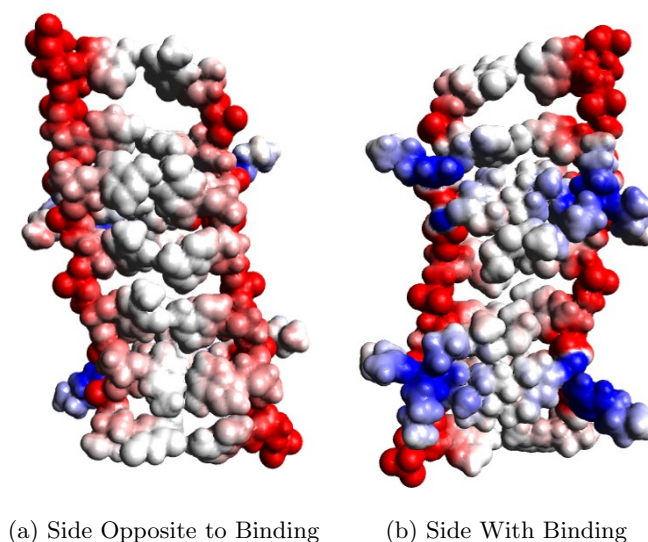


Figure 17: (a) Electrostatic potential on the side opposite to that which the cortisol molecule bonds onto the TATATTTT nucleotides, indicating neutral internal. (b) Electrostatic potential on the side in which the cortisol molecule bonds onto the TATATTTT nucleotides indicating neutrality down the midline.

steroid of prednisone converting to its active form of prednisolone by simply changing a ketone to a hydroxyl, which is done at a position that is exactly complementary to hydrogen bind to thymine. This analysis can be extended to other synthetic steroids including dexamethasone, which due to the position of a fluoride atom enables a stronger bond to thymine. The synthetic corticosteroids can be compared to the quality of the bonds than the natural hormone cortisol forms, which I have assessed in other preprints [10]. Because atomic level resolution is not available in imaging technologies, the use of pharmacological comparative data in efficacy provides an independent method of evaluating theoretical developments.

It was determined that both symmetry and complimentary mechanisms of interaction between ligands and DNA nucleotide pairs was required for binding in order to achieve efficacy. The symmetric arrangement is achieved through alignment of the ring structures of the ligands in relation to the ring structures of the DNA nucleotide pairs. This was necessary in order achieve hydrophobic characteristic of the inner core of DNA in between the phosphodiester strands. The complementary relationship was achieved through the alignment of the functional groups on both the external phosphodiester backbone, as well as the internal interaction with the unpaired functional group of Thymine, or of the paired oxygen group of Cytosine. Hydrogen bonds of internal number three will result for both the A-T/T-A and the G-C/C-G pairings if an intermolecular bond with the ligand is achieved,

and apparently this is very important and understandable as the driving force would be to make the internal core more hydrophobic.

As a further aid in the newly developed predictive capability for analysis of ligands, it is profound that there is an association, both structurally and complementarily, of cortisol-like molecules with the A-T and T-A pairing, and of testosterone-like molecules with the G-C and C-G pairing. Therefore, it is possible to determine from the nature of the preinitiation complex, that is the promoter site, the type of influence that the hormone will have on the transcription rate. For example, it would be expected that increases in cortisol levels would have a stronger impact on gene transcription rates associated with promoter sites similar to the TATA box, that is rich in A and T, while testosterone would have a stronger impact on genes associated with promoter sites similar to CpG islands, or rich in C and G. The organization of the A/T versus C/G in terms of separation will also permit binding by positive two charged ions, and thus will influence the quality of the bond between the ligand and DNA nucleotides. Moreover, the fact that there is a perfect structural correspondence between the steroid hormones and the A/T and C/G Watson-Crick DNA nucleotide pairings provides very strong evidence that direct binding of ligands onto DNA nucleotides induces transcription (otherwise, this perfect match would be effectively coincidental and useless, which seems unlikely.)

Although vitamin D and prostaglandins are considered hormones, their influence over the transcription process is much less pronounced than that observed for hormones such as cortisol and testosterone biomaterials. However, there is an impact, and insight can be drawn on analysis of the efficacy of these materials. The weaker intermolecular structures that will result of calcitriol and 15d-PDJ2 relative to the four fused rings of the steroid hormones are indicative that the stabilization processes of the DNA nucleotides will not be as significant, and thus the relative degree of transcription would be less accordingly in duration and rate. The stabilization of the binding site, and destabilization of the upstream to promote unwinding and strand separation, may be further enhanced through the flipping of the DNA nucleotide pairs, that is the formation of A^{-1}/T^{-1} or due to a backside and frontside orientation of the binding sites.

In this analysis, the transcription factors function as a delivery vector in a manner similar to the glucocorticoid receptor, rather than as DNA binding mechanism. The molecular modeling indicates that the ligand is transported out of a ligand binding domain and comes into association with the DNA nucleotide pairs to activate transcription, separated slightly from the transcription factor. Alternatively, it is conceivable that a complementary ligand freely available in the nucleoplasm is recruited to a binding domain associated with the delivery vector, and then transitioned to bind onto the DNA nucleotide pairs. (It is noted that the charged residuals of the amino acids comprising the transcription factors would have to enter the core of the DNA nucleobases in between the phosphodiester strands, which would be counterintuitive to its hydrophobic characteristic.)

For applications using this newly found mechanism of action for transcription, it is possible to precisely design pharmacological ligands. There is a role for the end groups in

terms of binding affinity in its interaction with the phosphodiester. The ring structure and maintaining symmetry to achieve a hydrophobic interior is critical. Apparently of the most significance is the intermolecular hydrogen bond with Thymine, or the absence of it. The content of A/T versus C/G will thus define the ligand steroid class, \mathbb{S} or \mathbb{H} , that will have the most effect on the corresponding gene expression. Other areas where influence over the transcription process takes place is in the nature of the ion that can secure the stabilizing ligand to the DNA nucleotide, which is also possible to be achievable without an ion, and through hydrogen bonding alone; but it would be expected that the transcription duration would be less since it would lose association quicker.

As to the reasons for this mechanism of action not having been developed before, it may be explainable due to the difficulty in obtaining atomic level measurements, and because the ligand is complementary and symmetric to the DNA nucleotide pair, which may explain why this interaction between ligand and DNA has been hidden. Imaging techniques have difficulty picking up a signal from smaller molecules like steroid hormones [25]. In addition, my prerequisite analysis involving the discovery of a molecular primer as to a shape of the steroid molecule embedded in DNA structure was important in this regard in developing the mechanism of action for transcription, and this result is only recent. In light of these molecular modeling results in combination with the pharmacological efficacy data supporting a mechanism of action involving ligand binding onto DNA nucleotides, it seems likely that direct binding of a ligand onto the DNA nucleotide strands induces transcription.

4 Methods

To develop the case, molecular modeling was applied, and the literature was consulted to validate the approach using published data, particularly that related to pharmacological efficacy. above descriptions provide the basis for the common shape of a steroid molecule for each of the DNA nucleotide pairings, and the assignment of a class of steroid hormones to each of the nucleotide pairings based on the capability to produce a hydrogen bond with the unmatched functional group of thymine.

- **Molecular Dynamics:** To build three dimensional molecular models and to evaluate the interaction of ligands and nucleotides, the software program Avogadro was used. Using the software program, the atoms were graphically arranged according to the preferred chemical configuration of the molecular structure. The force field MMFF94 [26] was applied. To determine bond length and 3D orientation, the optimization program was used based upon steepest descent available in the software that minimizes the energy associated with the complex.
- **Docking ligands to DNA nucleotides:** The ligands were prepared atom by atom for the steroid hormones cortisol and testosterone, and for calcitriol, PDJ2 and 15d-PDJ2.

The DNA nucleotides were prepared graphically atom by atom, and the software program used to minimize energy. The ligands were then graphically placed in the vicinity adjacent to the targeted A-T/T-A or G-C/C-G pair, and arranged such that the binding to the unpaired thymine was feasible. Ions of positive two charge for calcium or magnesium, or positive one charge for sodium, were then spatially positioned in the vicinity of adjacent phosphor atoms and the functional end group of the ligand. The software optimization routine was utilized to allow for docking of the overall structure. If necessary, the molecules were graphically perturbed so as to avoid local minimum and to direct the final configuration in which the overall molecule was stabilized relative to the overall energy of the complex.

- Docking ligands to transcription factor as the glucocorticoid receptor or the TBP: The amino acid residues of its ligand binding domain, including the α carbon, were positioned in proximity to the functional groups in accordance with the approximate layout presented in [27], which used crystal studies for identification of the amino acid residues of the ligand binding domain of a conserved glucocorticoid receptor. A cortisol molecule was then set in the vicinity of the residuals, and calcium or magnesium ions were added with positive two charge. The optimization program was then run until the overall stability was established. The dimensions of the spacings were determined at the central location of the carbon, oxygen, hydrogen, nitrogen, calcium or magnesium element, depending upon whether it was the size of the ligand binding domain or the length of the hydrogen bonding. A similar procedure was utilized for the TBP, whose amino acid residuals and appropriate positioning relative to the TATATTTT configuration was utilized [24].
- Transferring corticosteroid ligand from the transcription factor TBP and docking to DNA nucleotide pairs A-T and T⁻¹-A⁻¹: After assembling the cortisol molecule in the amino acid residuals of TBP at two sites, the units were brought in close vicinity to the TATATTTT DNA nucleotide pairs. The magnesium ions were brought in between lateral phosphate groups at the appropriate nucleotide. One of the A-T nucleotide pairs were inverted, that is swiveled along the single bond between the base and sugar structures. Both corticosteroids were then placed such that it was feasible for binding the hydroxyl group on the C-ring of the cortisol molecule to the ketone group at the carbon two position of thymine. The TBP residues were then motioned slightly away from the cortisol molecule functional groups. The software program was then run to achieve a stable complex, docking the two cortisol molecules, bound to the DNA nucleotide strands, resulting in the transfer of the ligand from the TBP transcription factor to the TATATTTT site at two docking sites, with the resultant residues of the TBP transcription factor positioned in correspondence with the experimentally observed configuration.

Bibliography

- [1] P. Cramer, "Organization and regulation of gene transcription," *Nature*, vol. 573, no.7772, pp. 45–54, 2019.
- [2] O. Fornes, J. A. Castro-Mondragon, A. Khan, R. Van der Lee, X. Zhang, P. A. Richmond, B. P. Modi, S. Correard, M. Gheorghe, D. Baranasic et al., "Jaspar 2020: update of the open-access database of transcription factor binding profiles," *Nucleic acids research*, vol. 48, no. D1, pp. D87–D92, 2020.
- [3] C. Chronis, P. Fiziev, B. Papp, S. Butz, G. Bonora, S. Sabri, J. Ernst, and K. Plath, "Cooperative binding of transcription factors orchestrates reprogramming," *Cell*, vol. 168, no. 3, pp. 442–459, 2017.
- [4] M. P. Stemmler, R. L. Eccles, S. Brabletz, and T. Brabletz, "Non-redundant functions of emt transcription factors," *Nature cell biology*, vol. 21, no. 1, pp. 102–112, 2019.
- [5] S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, and M. T. Weirauch, "The human transcription factors," *Cell*, vol. 172, no. 4, pp. 650–665, 2018.
- [6] M. Nadal, S. Prekovic, N. Gallastegui, C. Helsen, M. Abella, K. Zielinska, M. Gay, M. Vilaseca, M. Taul'es, A. B. Houtsmuller et al., "Structure of the homodimeric androgen receptor ligand-binding domain," *Nature communications*, vol. 8, no. 1, pp. 1–14, 2017.
- [7] J. F. Kribelbauer, C. Rastogi, H. J. Bussemaker, and R. S. Mann, "Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes," *Annual review of cell and developmental biology*, vol. 35, pp. 357–379, 2019.
- [8] T. Bhuiyan and H. T. M. Timmers, "Promoter recognition: putting tfiid on the spot," *Trends in cell biology*, vol. 29, no. 9, pp. 752–763, 2019.
- [9] C. Schaper, "Endogenous Binding of Steroid Molecules to DNA Nucleotides by a Ca²⁺/PO₄ Process to Enable Gene Transcription," *ChemRxiv*, 2020, <https://doi.org/10.26434/chemrxiv.11868261.v2>.
- [10] C. Schaper, "Structural Symmetry of DNA Nucleotides and Steroid Hormones," *ChemRxiv*, 2020, <https://doi.org/10.26434/chemrxiv.11991567.v1>.
- [11] C. Schaper, *Design of DNA, Genetic Codes, and Life Function*. ISBN 978-1-7353721-0-5, Molecular Sciences Publishing House, 2020.
- [12] C. Schaper, "Intermolecular Binding of Functional Cyclic Compounds to RNA Nucleotides by an Ionic Process," *Preprints*, 2020, <https://doi.org/10.20944/preprints202004.0522.v1>.

- [13] B. Feng, R. P. Sosa, A. K. Martensson, K. Jiang, A. Tong, K. D. Dorfman, M. Takahashi, P. Lincoln, C. J. Bustamante, F. Westerlund et al., “Hydrophobic catalysis and a potential biological role of dna unstacking induced by environment effects,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 35, pp. 17 169–17 174, 2019.
- [14] C. Schaper, “Competitive Inhibition of Cortisol by Prostaglandins at the Ligand Binding Domain of Glucocorticoid Receptors,” *bioRxiv*, 2019, <https://doi.org/10.1101/851501>.
- [15] A. Ou, J. W. Schmidberger, K. A. Wilson, C. W. Evans, J. A. Hargreaves, M. Grigg, M. L. O’Mara, K. S. Iyer, C. S. Bond, and N. M. Smith, “High resolution crystal structure of a kras promoter g-quadruplex reveals a dimer with extensive poly-a stacking interactions for small-molecule recognition,” *Nucleic acids research*, vol. 48, no. 10, pp. 5766–5776, 2020.
- [16] G. Papai, A. Frechard, O. Kolesnikova, C. Crucifix, P. Schultz, and A. Ben-Shem, “Structure of saga and mechanism of tbp deposition on gene promoters,” *Nature*, vol. 577, no. 7792, pp. 711–716, 2020.
- [17] H. Wang, C. Dienemann, A. Stutzer, H. Urlaub, A. C. Cheung, and P. Cramer, “Structure of the transcription coactivator saga,” *Nature*, vol. 577, no. 7792, pp. 717–720, 2020.
- [18] J. Vierstra, J. Lazar, R. Sandstrom, J. Halow, K. Lee, D. Bates, M. Diegel, D. Dunn, F. Neri, E. Haugen et al., “Global reference mapping of human transcription factor footprints,” *Nature*, vol. 583, no. 7818, pp. 729–736, 2020.
- [19] S. A. Lambert, A. W. Yang, A. Sasse, G. Cowley, M. Albu, M. X. Caddick, Q. D. Morris, M. T. Weirauch, and T. R. Hughes, “Similarity regression predicts evolution of transcription factor sequence specificity,” *Nature genetics*, vol. 51, no. 6, pp. 981–989, 2019.
- [20] C. Tan and S. Takada, “Nucleosome allostery in pioneer transcription factor binding,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 34, pp. 20 586–20 596, 2020.
- [21] X. He, D. Ni, H. Zhang, X. Li, J. Zhang, Q. Fu, Y. Liu, and S. Lu, “Zinc-mediated conformational preselection mechanism in the allosteric control of dna binding to the zinc transcriptional regulator (zitr),” *Scientific RepoRtS*, vol. 10, no. 1, pp. 1–12, 2020.
- [22] M. M. Makowski, G. Gaullier, and K. Luger, “Picking a nucleosome lock: Sequence and structure-specific recognition of the nucleosome,” *Journal of Biosciences*, vol. 45, no. 1, p. 13, 2020

- [23] C. Schaper, “New Code for DNA Nucleotide Sequences,” ChemRxiv, 2020, <https://doi.org/10.26434/chemrxiv.12003786.v1>.
- [24] K. Kamada, F. Shu, H. Chen, S. Malik, G. Stelzer, R. G. Roeder, M. Meisterernst, and S. K. Burley, “Crystal structure of negative cofactor 2 recognizing the tbp-dna transcription complex,” *Cell*, vol. 106, no. 1, pp. 71–81, 2001.
- [25] Y. Liu, S. Gonen, T. Gonen, and T. O. Yeates, “Near-atomic cryo-em imaging of a small protein displayed on a designed scaffolding system,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 13, pp. 3362–3367, 2018.
- [26] T. A. Halgren, “Merck molecular force field. i. basis, form, scope, parameterization, and performance of MMFF94,” *Journal of computational chemistry*, vol. 17, no. 5-6, pp. 490–519, 1996.
- [27] X. Liu, Y. Wang, and E. A. Ortlund, “First high-resolution crystal structures of the glucocorticoid receptor ligand-binding domain–peroxisome proliferator-activated coactivator 1- α complex with endogenous and synthetic glucocorticoids,” *Molecular Pharmacology*, vol. 96, no. 4, pp. 408–417, 2019.