

Títol del treball:

SKINNY AND FAT DNA

Estudiant: Cristina Berga Domínguez.

Grau en: Química.

Correu electrònic: u1935842@campus.udg.edu

Tutora: Sílvia Simon Rabasseda.

Empresa/institució: Institut de química computacional i catàlisi.

Vistiplau tutor:

Sílvia Simon
Rabasseda - DNI
40306168W
(TCAT)

Digitally signed by Sílvia Simon Rabasseda
- DNI 40306168W (TCAT)
DN: c=ES, o=Universitat de Girona,
2.5.4.97=IATES-06750002E, ou=Empleat
public de nivell mig, sn=Simon Rabasseda
- DNI 40306168W, givenName=Sílvia,
serialNumber=IDCES-40306168W,
cn=Sílvia Simon Rabasseda - DNI
40306168W (TCAT)
Date: 2021.06.01 16:28:48 +02'00'

Nom de la tutora: Sílvia Simon Rabasseda.

Empresa/institució: Institut de química computacional i catàlisi.

Correu electrònic: silvia.simon@udg.edu

Cotutores (no oficials):

Olga Stasyuk – IQCC

Cèlia Fonseca-Guerra – Universiteit Leiden and VU Amsterdam

Índex

RESUM.....	2
RESUMEN.....	3
SUMMARY.....	4
1. INTRODUCCIÓ.....	5
2. OBJECTIVES.....	8
3. METODOLOGIA.....	9
3.1. DFT.....	9
3.2. TEOREMES DE HOHENBERG I KOHN.....	9
3.3. EQUACIONS DE KOHN-SHAM.....	10
3.4. APROXIMACIONS FUNCIONAL D'INTERCANVI I CORRELACIÓ.....	11
3.5. ADF I MÈTODES COMPUTACIONALS.....	11
3.6. GENERACIÓ DE LES SUPERFÍCIES DE POTENCIALS ROTACIONALS.....	12
3.7. EDA.....	13
3.8. ÈTICA I SOSTENIBILITAT.....	13
4. RESULTATS I DISCUSSIÓ.....	15
4.1. GENERACIÓ DE LES SUPERFÍCIES DE POTENCIALS ROTACIONALS.....	16
4.2. INTERACCIÓ π - π STACKING I ANÀLISI DE LA DESCOMPOSICIÓ DE L'ENERGIA.....	18
4.3. EFECTES DELS SUBSTITUENTS EN LA INTERACCIÓ π - π STACKING.....	21
5. CONCLUSIONS.....	27
6. REFERÈNCIES.....	28

RESUM.

L'ADN és el lloc on es guarda tota la informació genètica necessària pel funcionament, desenvolupament, creixement i reproducció dels éssers vius. És l'encarregat de transmetre les instruccions requerides per crear altres components de les cèl·lules, com són les proteïnes. Està compost per dos polímers llargs col·locats antiparal·lelament que s'entrellacen formant una doble hèlix. Aquestes dues cadenes estan compostes per nucleòtids. Les bases nitrogenades dels nucleòtids poden ser: adenina (A), citosina (C), guanina (G) i timina (T). Sempre s'enllacen, mitjançant enllaços d'hidrogen, les adenines amb les timines i les citosines amb les guanines, és a dir, una purina amb una pirimidina. Aquestes parelles de bases es van alternant una sobre l'altra amb un angle de 35° i s'estabilitzen mitjançant les interaccions π - π *stacking*.

Estudis recents han començat a investigar amb bases artificials, per tal d'introduir-les dins l'ADN i que aquest pugui guardar més informació i construir proteïnes diferents. Això permetria, entre altres coses, un benefici en la medicina. Un d'aquests estudis recents, ha creat dos grups de bases artificials: les *skinny* i les *fat*. Les *skinny* són parelles de pirimidines i les *fat* parelles de purines, per tant, no es segueix la complementarietat de mida de les bases nitrogenades naturals (AT i GC).

En aquest projecte, es fa un estudi computacional dels sistemes formats per les diferents parelles de bases artificials i parelles de bases naturals amb la finalitat de comprovar si les bases artificials són suficientment estables com per introduir-les dins l'ADN. Els paràmetres que s'investiguen són l'angle de gir òptim, les interaccions π - π *stacking* en fase gas i en aigua i l'efecte que provoquen diferents substituents a aquesta interacció.

Els resultats obtinguts mostren que, tot i haver-hi un rang ampli d'angles, aquests difereixen molt poc d'energia amb l'energia a 35° , l'original de l'ADN, el que permetria que les bases artificials s'hi adaptessin correctament. A més, les interaccions π - π *stacking* són semblants o superiors a les produïdes per les bases naturals, el que implica una major estabilitat del sistema. Les parelles de bases CV i BP han sigut les seleccionades per comprovar l'efecte dels substituents, perquè han sigut les que han obtingut una energia d'interacció més favorable. En el cas del CV, s'ha vist una millora amb els substituents electroacceptors, però els electrodonadors no l'afecten. Per contra, en el cas del BP, hi ha hagut molta variabilitat, tot i que en general, també s'han obtingut millors resultats amb un substituent electroacceptor. D'aquesta manera es pot concloure que els resultats obtinguts demostren que les bases artificials són estables i obren un món de possibilitats.

RESUMEN.

El ADN es el lugar donde se guarda toda la información genética necesaria para el funcionamiento, desarrollo, crecimiento y reproducción de los seres vivos. Es el encargado de transmitir las instrucciones requeridas para crear otros componentes de las células como, por ejemplo, las proteínas. Está compuesto por dos polímeros largos colocados antiparalelamente que se entrelazan formando una doble hélix. Estas dos cadenas están compuestas por nucleótidos. Las bases nitrogenadas de los nucleótidos pueden ser: adenina (A), citosina (C), guanina (G) y timina (T). Siempre se enlazan, mediante enlaces de hidrógeno, las adeninas con las timinas y las guaninas con las citosinas, es decir, una purina con una pirimidina. Estas parejas de bases se van alternando una sobre la otra con un ángulo de 35° y se estabilizan mediante las interacciones π - π *stacking*.

Estudios recientes han empezado a investigar con bases artificiales, para poder introducir las dentro del ADN y que este pueda guardar más información y construir proteínas diferentes. Esto permitiría, entre otras cosas, un beneficio en la medicina. Uno de estos estudios recientes, ha creado dos grupos de bases artificiales: las *skinny* y las *fat*. Las *skinny* son parejas de pirimidinas y las *fat* son parejas de purinas, por lo tanto, no se sigue la complementariedad de tamaño de las bases nitrogenadas naturales (AT i GC).

En este proyecto, se hace un estudio computacional de los sistemas formados por las diferentes parejas de bases artificiales y parejas de bases naturales con la finalidad de comprobar si las bases artificiales son suficientemente estables como para introducir las dentro el ADN. Los parámetros que se investigan son el ángulo de giro óptimo, las interacciones π - π *stacking* en fase gas y en agua y el efecto que producen diferentes sustituyentes a esta interacción.

Los resultados obtenidos muestran que, incluso con un rango amplio de ángulos, estos difieren muy poco de energía con la energía a 35° , la original del ADN, lo que permitiría que las bases artificiales se le adaptaran correctamente. Además, las interacciones π - π *stacking* son muy parecidas o superiores a las producidas por las bases naturales, lo que implica una mayor estabilidad del sistema. Las parejas de bases CV y BP han sido las seleccionadas para comprobar el efecto de los sustituyentes, porque han sido las que han obtenido una energía de interacción más favorable. En el caso de CV, se ha visto una mejora con los sustituyentes electroaceptores, pero los electrodonadores no le afectan. Por el contrario, en el caso de BP, ha habido mucha variabilidad, aunque en general, también se han obtenido mejores resultados con un sustituyente electroceptor. De esta manera se puede concluir que los resultados obtenidos demuestran que las bases nitrogenadas artificiales son estables y abren un mundo de posibilidades.

SUMMARY.

DNA is where all the genetic information necessary for the functioning, development, growth and reproduction of living things is stored. It is responsible for transmitting the instructions required to create other components of cells, such as proteins. It is composed of two long polymers placed antiparallely that intertwine to form a double helix. These two chains are composed of nucleotides. The nitrogenous bases of nucleotides can be: adenine (A), cytosine (C), guanine (G), and thymine (T). They always are linked, by hydrogen bonds, adenines with thymines and cytosines with guanines, i.e., a purine with a pyrimidine. These base pairs alternate with each other at an angle of 35° and are stabilized by π - π stacking interactions.

Recent studies start to investigate with artificial bases in order to introduce them into DNA so that it can store more information and build different proteins. This would allow, among other things, a benefit in medicine. One of these recent studies has created two groups of artificial bases: skinny and fat. The skinny are pairs of pyrimidines and the fat pairs of purines; therefore, the size complementarity of natural nitrogenous bases (AT and GC) is not followed.

In this project, a computational study of the systems formed for the different artificial pairs and natural pairs is made in order to check if the artificial bases are stable enough to introduce them into the DNA. The parameters investigated are the optimal twist angle, the π - π stacking interactions in the gas and water phases and the effect caused by different substituents on this interaction.

The results obtained show that, despite having a wide range of angles, they differ very little between optimal angle energy and the energy at 35° , the original DNA, which would allow the artificial bases to be they adapted to it correctly. In addition, π - π stacking interactions are similar or superior to those produced by natural bases, implying greater system stability. The base pairs CV and BP have been selected to check the effect of the substituents, because they have been the ones that have obtained a more favourable interaction energy. In the case of the CV, there has been an improvement with EWG substitutes, but EDG do not affect it. In contrast, in the case of BP, there has been a lot of variability, although in general, better results have also been obtained with an EWG substituent. In this way, it can be concluded that the results obtained show that the artificial bases are stable and open up a world of possibilities.

1. INTRODUCCIÓ.

L'ADN sempre està en boca de tothom i més ara en mig d'una pandèmia mundial. És ben sabut que és un dels punts claus de la vida, però què és exactament, per a què serveix? Les sigles ADN (o DNA de l'anglès) provenen de l'abreviació d'àcid desoxiribonucleic. L'ADN és com una recepta on hi ha tota la informació genètica necessària pel funcionament, desenvolupament, creixement i reproducció dels éssers vius. Mitjançant aquestes instruccions es poden produir altres components de les cèl·lules, especialment les proteïnes.

Va ser aïllat per primer cop per Friedrich Miescher l'any 1869.^[1] Phoebus Levene, l'any 1919, va identificar els nucleòtids i va suggerir que l'ADN era una cadena d'ells enllaçats pels grups fosfat.^[2] Però no va ser fins l'any 1953 quan James Watson i Francis Crick van suggerir l'actual estructura de la doble hèlix de l'ADN.^[3] Cal destacar que sense la feina prèvia de Wilkins i, sobretot, de Rosalind Franklin sobre la difracció de raigs X no haguessin aconseguit conèixer la seva estructura. Ja que el seu model es va basar en una imatge per difracció de raigs X, anomenada foto 51, que havia realitzat Raymond Gosling, el doctorand de Franklin.^[4]

Dins les cèl·lules, l'ADN s'organitza en estructures anomenades cromosomes, les quals proporcionen una ubicació de l'ADN reduït dins el nucli, una millor protecció i una millor distribució en la divisió cel·lular.

Químicament parlant, l'ADN és el conjunt de dos polímers llargs col·locats antiparal·lelament que s'entrellacen formant una doble hèlix. Aquestes dues cadenes estan compostes per nucleòtids, és a dir, per un grup fosfat, un sucre de cinc carbonis (2-desoxiribosa) i una base nitrogenada. El seu esquelet està format pels sucres i els grups fosfats que estan units per enllaços èster.

Deixant de banda l'uracil que només està en l'ARN, hi ha quatre possibles bases nitrogenades: l'adenina (A), la citosina (C), la guanina (G) i la timina (T). L'adenina i la guanina són purines i la citosina i la timina són pirimidines, per tant, les primeres tenen dos anells de cinc i sis membres i les segones només un anell de sis membres. La base nitrogenada d'una cadena s'uneix, majoritàriament mitjançant enllaços d'hidrogen, amb la base nitrogenada de l'altra cadena com si fos una cremallera, donant així l'estabilitat necessària per formar l'ADN complet. Però només hi ha dues possibilitats d'unió, hi ha una complementarietat de bases: l'adenina sempre s'uneix amb la timina i la guanina amb la citosina (figura 1). Per tant, sempre són una purina i una pirimidina, per mantenir així el diàmetre de la doble hèlix estable.

Les parelles de bases nitrogenades es van repetint alternativament una sobre l'altra, com si fossin els esglaons d'una escala, i s'estabilitzen entre elles per la interacció π - π *stacking* i amb un angle d'aproximadament 36 °, tot i que és una estructura no estàtica.

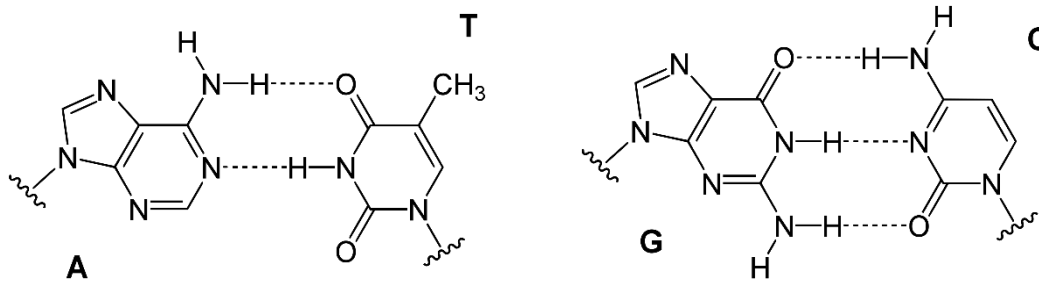


Figura 1. Parelles de bases naturals (AT i GC) que es troben a l'ADN.

Però ara venen les grans preguntes... És veritat que només hi ha aquestes quatre bases? Realment és necessària la complementarietat de bases? Sempre s'ha d'unir una pirimidina amb una purina? Doncs bé, sobre l'existència de només quatre bases i la seva complementarietat en l'ADN, estudis realitzats durant la dècada dels 70 van demostrar que hi ha virus de bacteris que en lloc de tenir l'adenina tenen l'anomenada lletra Z, que és la 2-aminoadenina.^[5,6,7] Per tant, en aquests virus no hi ha cap adenina i és la Z la que s'uneix amb la timina (figura 2). Aquesta unió, a més, és més forta i estable que l'original, ja que forma tres enllaços d'hidrogen en lloc de dos.

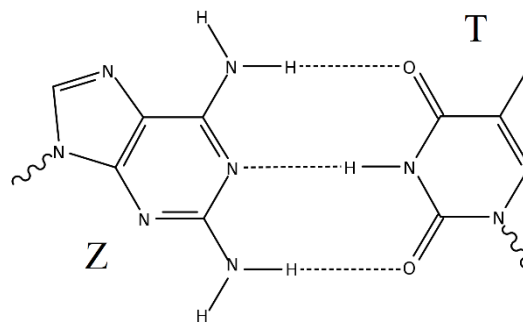


Figura 2. Z, la nova base nitrogenada trobada enllaçada a la timina.

Llavors, poden haver-hi més lletres o només és una excepció? Si fos així el codi genètic seria més extens, es podria guardar més informació i això permetria codificar diferents aminoàcids i construir diferents proteïnes amb altres funcions i propietats a les existents. Les proteïnes realitzen funcions elementals per la vida cel·lular, així que poder crear proteïnes noves mitjançant la modificació de l'ADN suposaria un benefici gegant en l'àmbit, per exemple, de la medicina.

Es poden trobar diferents estudis sobre l'ADN, com és el cas de Poater et al. (2014)^[8], en el qual van estudiar computacionalment l'estructura i l'estabilitat de l'ADN-B. Ho van fer basant-se en el rol dels enllaços d'hidrogen, les interaccions π - π *stacking*, l'angle de gir i la solvatació. Van aplicar els diferents paràmetres a parells de bases naturals (AT i GC) i a altres parells de bases artificials i es va comprovar com en alguns casos les bases artificials milloraven l'estabilitat.

Uns altres estudis molt importants són els que ha dut a terme Hoshika et al. (2018)^[9]. En el més recent van crear dos grups de bases artificials: les *skinny* i les *fat* (figura 3). Els *skinny* són aquells parells de bases on les bases nitrogenades de cadascun té només un anell, és a dir, són pirimidines. Els *fat* són aquells parells de bases on les bases nitrogenades de cadascun tenen dos anells, és a dir, són purines. En aquest article, a part

de demostrar que les bases artificials són més estables termodinàmicament, també s'han pogut sintetitzar tres cristalls i fer-ne el seu posterior estudi per difracció de raigs X. D'aquesta manera sobre completament la porta, és una indicació que s'està anant per bon camí.

En el següent estudi, s'estudien computacionalment els diferents parells artificials *skinny* i *fat* (figura 3), de l'article mencionat anteriorment, amb els parells naturals. Els paràmetres estudiats són l'angle de gir òptim, la interacció π - π *stacking* en fase gas i solvatació sense substituents i, un estudi, previ amb diferents substituents a les bases artificials.

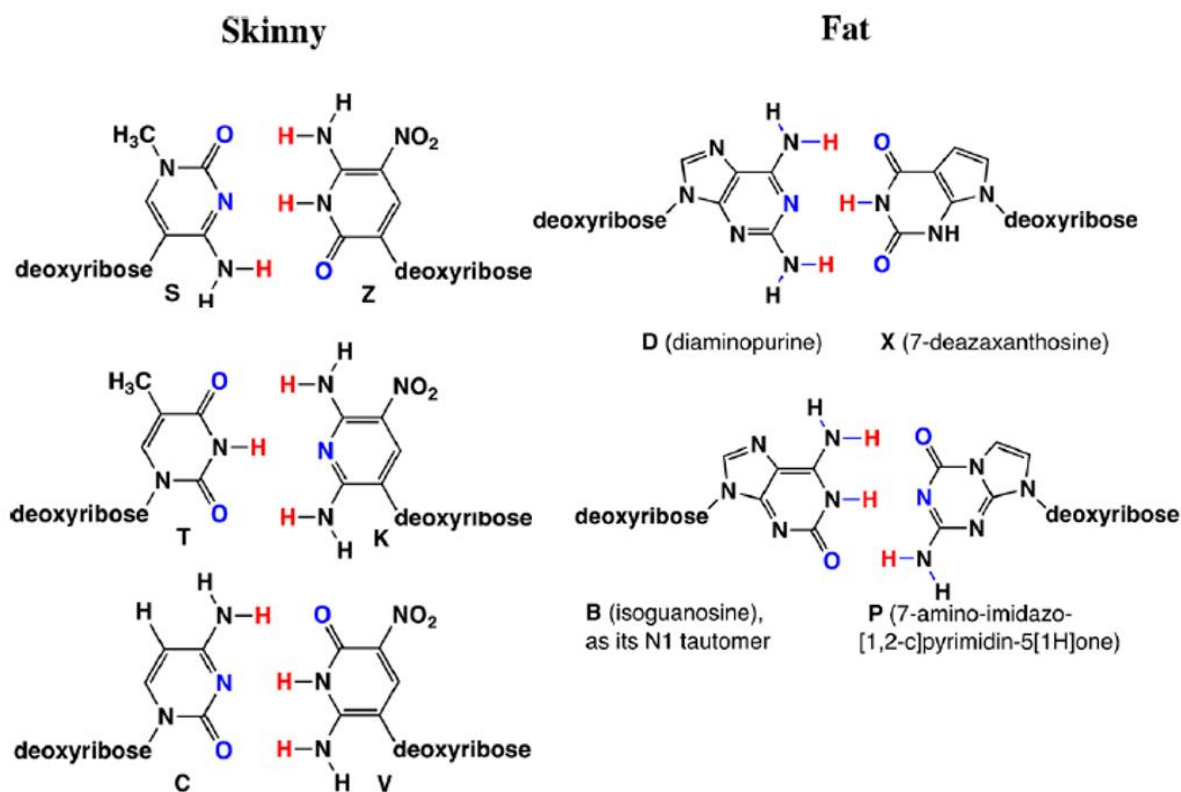


Figura 3. Grups *skinny* i *fat* de parells de bases artificials.

2. OBJECTIVES.

The main goal of this work is to do a computational study to check if pairs of artificial bases, which do not follow the size complementarity, can be part of the DNA with similar stability.

To carry out the study, two groups of artificial bases will be taken: skinny and fat. Skinny ones are pairs of pyrimidines and fat ones are pairs of purines. We will design parallel systems of both natural and artificial bases pair in order to model the π - π interaction between them.

The more specific objectives for each system are:

- Calculate the optimal rotational angle between pairs of bases.
- Calculate the π - π stacking interaction energy in gas phase and with solvent during the rotation.
- Analyse the interaction energy using the decomposition of energy to find out which term most influences stability.
- Study the effect of the substituents on the π - π stacking interaction.

3. METODOLOGIA.

3.1. DFT.

La Teoria del Funcional de la Densitat (DFT) és un mètode mecànic-quàntic computacional que serveix per resoldre l'equació d'Schrödinger independent del temps per tal d'obtenir l'energia electrònica d'un sistema.^[10,11]

Els mètodes *ab initio* convencionals calculen l'energia electrònica mitjançant la funció d'ona, per contra, la DFT utilitza la densitat electrònica. Segons aquesta teoria, les propietats d'un sistema polieletrònic es poden determinar utilitzant funcionals, és a dir, funcions d'una altra funció. Per tant, els funcionals utilitzats són funcions de la densitat electrònica.

Malgrat no es coneix l'equació exacta que relaciona la densitat electrònica amb l'energia, s'utilitzen aproximacions molt bones. A més, la funció d'ona depèn de cada una de les coordenades dels electrons, és a dir, 3N variables (o 4N si es té en compte l'espín), en canvi, la densitat electrònica només depèn d'*x*, *y* i *z*, per tant, simplifica bastant els càlculs.

Tot i que les primeres nocions sobre una possible teoria del funcional de la densitat van ser desenvolupades per Llewellyn Thomas i Enrico Fermi als anys 1920, no va ser fins l'any 1964 quan Pierre Hohenberg i Walter Kohn van trobar la base teòrica de la DFT.^[12]

3.2. TEOREMES DE HOHENBERG I KOHN.

El primer teorema diu que: “Qualsevol observable d'un estat estacionari fonamental no degenerat pot ser calculat, en principi de forma exacte, a partir de la densitat electrònica d'aquest estat fonamental, és a dir, qualsevol observable pot escriure's com un funcional de la densitat”.

Realitzant les demostracions pertinents s'arriba a la conclusió que hi ha una relació directa entre la densitat i la funció d'ona a través del potencial extern, tal i com es mostra a l'equació 1. Això sempre i quan el sistema sigui *v*-representable i *N*-representable, és a dir, que existeixi un potencial extern a partir del qual es pugui derivar la densitat i aquesta sigui una funció definida positiva en tot l'espai donant el nombre total d'electrons del sistema en integrar-se.

$$\rho(\vec{r}) \rightarrow v(\vec{r}) \rightarrow \hat{H} \rightarrow \psi \Leftrightarrow E = E[\Psi] = E[\rho] \quad \text{Eq. 1}$$

D'aquesta manera l'energia queda com a l'equació 2.

$$E[\rho] = T[\rho] + V_{Ne}[\rho] + V_{ee}[\rho] \quad (+V_{NN}) \quad \text{Eq. 2}$$

On T és l'energia cinètica, V_{Ne} el potencial d'interacció electró-nucli, V_{ee} el potencial electró-electró i V_{NN} el potencial nucli-nucli el qual està separat perquè és una constant.

Com que $T[\rho]$ i $V_{ee}[\rho]$ són funcionals universals, perquè no depenen del potencial extern, s'engloben en l'anomenat funcional de Hohenberg i Kohn, $F_{HK}[\rho]$, quedant l'equació 3.

$$E_v[\rho] = \int \rho(\vec{r})v(\vec{r})d\vec{r} + F_{HK}[\rho] \quad \text{Eq. 3}$$

On $E_v[\rho]$ indica que per a un potencial extern $v(r)$ determinat l'energia és un funcional de la densitat.

El segon teorema introdueix el principi variacional a $E[\rho]$ i diu tal que així: “La densitat electrònica d'un estat fonamental no degenerat pot ser calculada, en principi de forma exacta, determinant aquella densitat que minimitza l'energia de l'estat fonamental”.

Aquest principi assegura que qualsevulla densitat de prova proporciona una energia superior o igual a l'energia exacta de l'estat fonamental. Per tant, s'ha de trobar la densitat que minimitza l'energia per obtenir la densitat exacta de l'estat fonamental.

El problema és que el funcional $F_{HK}[\rho]$ no es coneix de forma exacte. Així que un any més tard, l'any 1965, Walter Kohn i Lu Jeu Sham van presentar una forma d'aproximar aquest funcional.^[13]

3.3. EQUACIONS DE KOHN-SHAM.

Kohn i Sham van crear un sistema fictici constituït per un sistema d' N electrons sense interaccionar que, mitjançant un potencial efectiu, v_{eff} , generaven la mateixa densitat que qualsevol sistema amb electrons que interaccionen amb un potencial extern diferent.

Al no interactuar, la seva funció d'ona és un determinant d'Slater únic. Per tant, l'energia cinètica és la suma d'energies cinètiques individuals i la densitat electrònica és la suma de densitats orbitalàries. La interacció electró-electró, V_{ee} , l'atribueixen principalment a la interacció coulòmbica o Hartree, J , i l'aproximen a aquesta. Així F_{HK} queda com a l'equació 4.

$$F[\rho] = T[\rho] + V_{ee}[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] \quad \text{Eq. 4}$$

On E_{xc} és el funcional d'intercanvi i correlació que a partir d'ara és el terme que es desconeix i s'ha d'aproximar.

Per tant, l'energia total d'un sistema s'expressa com (equació 5):

$$E_v[\rho] = \int \rho(\vec{r})v(\vec{r})d\vec{r} + T_s[\rho] + J[\rho] + E_{xc}[\rho] \quad \text{Eq. 5}$$

D'aquesta manera l'equació de Kohn-Sham (equació 6) final és la següent:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + v_{\text{eff}}(\mathbf{r})\right)\varphi_i(\mathbf{r}) = \varepsilon_i\varphi_i(\mathbf{r}) \quad \text{Eq. 6}$$

On v_{eff} correspon al potencial efectiu de Kohn-Sham que incorpora els efectes entre els electrons i els d'un potencial extern.

3.4. APROXIMACIONS FUNCIONAL D'INTERCANVI I CORRELACIÓ.

Les funcions exactes del funcional d'intercanvi i correlació només es coneixen pels gasos de Fermi, és a dir, per un conjunt de fermions que no interactuen. Així que hi ha una sèrie d'aproximacions del funcional.

La més bàsica és l'aproximació de densitat local (LDA). Aquesta depèn només del valor de la densitat electrònica a cada punt de l'espai (equació 7). Hi ha una versió d'aquesta anomenada aproximació de densitat d'espín local (LSDA) on s'inclou l'espín dels electrons.

$$E_{xc}^{\text{LDA}}[\rho] = \int \rho(\mathbf{r})\epsilon_{xc}(\rho(\mathbf{r})) d\mathbf{r} \quad \text{Eq. 7}$$

L'energia d'intercanvi i correlació en LDA es separa en l'energia d'intercanvi i en l'energia de correlació. Per corregir els errors produïts per l'aproximació d'aquests termes, es poden utilitzar els gradients de densitat. D'aquesta manera es té en compte la falta d'homogeneïtat de la densitat d'electrons. Quan s'utilitzen aquestes correccions s'anomena aproximació de gradient generalitzat (GGA).^[14,15]

Tot i que la GGA proporciona resultats força bons, han anat sortint més versions per tal d'obtenir resultats més exactes. Com per exemple, la Meta-GGA en la qual s'inclou la segona derivada de la densitat electrònica a la GGA, ja que aquesta consta només de la densitat i la primera derivada. A més, també hi ha funcionals híbrids amb la teoria de Hartree-Fock.

3.5. ADF I MÈTODES COMPUTACIONALS.

Tots els càlculs del treball s'han realitzat amb el programa Amsterdam Density Functional (ADF).^[16,17] Aquest programa resol l'equació d'Schrödinger basant-se en la DFT i utilitzant orbitals de tipus Slater (STO). Aquest tipus de funcions de base tenen l'avantatge que presenten una convergència molt ràpida, però per contra no són ortogonals i les integrals bielectròniques són difícils de resoldre.

El mètode utilitzat ha sigut el BLYP-D3(BJ)/TZ2P, és a dir, el funcional BLYP amb correcció Grimmer 3 i Becke-Johnson damping i la base TZ2P.^[18,19] El funcional BLYP

és de tipus GGA, el qual combina la funció d'intercanvi de Becke (B) i la funció de correlació de Lee, Yang i Parr (LYP).^[20,21]

La funció de base TZ2P és una funció de qualitat triple zeta amb dues funcions de polarització. Una base és un conjunt de vectors linealment independent que formen el sistema generador d'un espai vectorial. Els vectors que s'utilitzen en mecànica quàntica són funcions. Sempre és possible escriure un orbital com una combinació d'un conjunt de funcions de base, per tant, les funcions de base tenen el paper d'orbitals atòmics. Les funcions de polarització són aquelles que tenen una simetria superior a la precisa per la base mínima d'un àtom. Quan es treballa amb una base mínima els resultats són més dolents que quan es treballa amb una base que contingui funcions de polarització. Això és degut a que les bases amb funcions de polarització utilitzen més funcions de base de les que són necessàries i, per tant, més orbitals. Al tenir més orbitals tenen més possibilitat de donar valors més aproximats.

Un dels problemes que té la DFT és que no té en compte les interaccions no covalents i aquestes són molt importants a l'ADN. Per tant, s'ha utilitzat la dispersió de Grimme combinada amb Becke-Johnson damping per descriure aquestes interaccions.

Els efectes del dissolvent, concretament aigua, s'han descrit amb el model COSMO, el qual té en compte els efectes de cavitació, energia interna i entropia del dissolvent.

3.6. GENERACIÓ DE LES SUPERFÍCIES DE POTENCIALS ROTACIONALS.

El primer pas ha sigut obtenir les estructures òptimes de les parelles de bases, mitjançant una optimització i assegurant les freqüències positives per tal de confirmar el mínim en la superfície d'energia potencial (PES). Un cop optimitzades, s'han muntat els sistemes amb dues parelles de bases cada un i una distància entre elles de 3.4 Å (figura 4), sempre una parella natural amb una *skinny* o *fat*.

Mitjançant un script creat per Marcel Swart i càlculs puntuals de la PES, és a dir, sense tornar a optimitzar, s'ha fet un escombrat entre 0 i 60 ° en passos de 5 °, per tal de trobar l'angle òptim (figura 5).

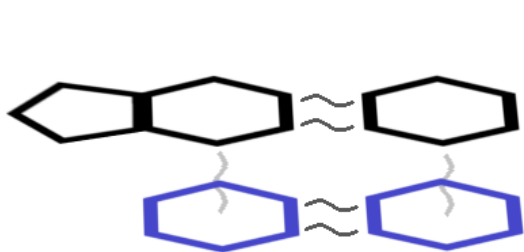


Figura 4. Esquema dels sistemes creats, en negre el parell de bases naturals i en blau el parell de bases *skinny* o *fat*.

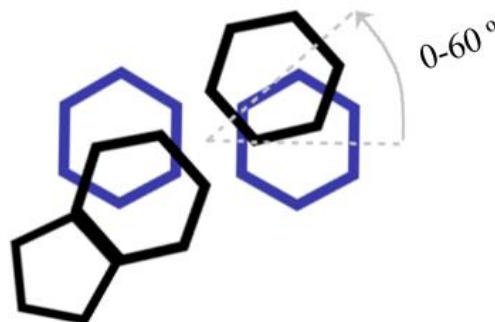


Figura 5. Esquema de la rotació dels sistemes creats.

L'energia que s'obté al realitzar els càlculs puntuals durant l'escombrat és l'energia d'interacció π - π *stacking* en fase gas. Aquesta energia correspon a les interaccions atractives no covalents (dispersió) entre anells aromàtics, ja que aquests tenen enllaços π . Són forces dèbils si es comparen amb els enllaços covalents, però el conjunt de totes les interaccions *stacking* provoquen una gran estabilització a l'ADN.

Com que l'ADN en el seu estat natural no està en fase gas, sinó que està en solució aquosa, també s'han calculat les energies d'interacció π - π *stacking* en aigua.

3.7. EDA.

A la fase gas, per poder analitzar les interaccions entre dos fragments, s'ha fet una anàlisi de la descomposició de l'energia (EDA),^[22] el qual et descompon l'energia en: repulsió de Pauli (ΔE_{Pauli}), interacció orbitalària (ΔE_{oi}), interacció electrostàtica (ΔV_{elstat}) i energia de dispersió (ΔE_{disp}). La suma de tots aquests termes dóna l'energia d'interacció (ΔE_{int}), en aquest cas, l'*stacking* (equació 8).

$$\Delta E_{\text{int}} = \Delta V_{\text{elstat}} + \Delta E_{\text{Pauli}} + \Delta E_{\text{oi}} + \Delta E_{\text{disp}} \quad \text{Eq. 8}$$

La repulsió de Pauli és deguda a la repulsió entre els parells d'electrons. La interacció orbitalària és deguda a la interacció electrònica i a la seva deslocalització. La interacció electrostàtica és deguda a les càrregues i quant més polaritzada més negativa serà aquesta energia. Per últim, l'energia de dispersió ja s'ha comentat anteriorment i és la produïda per les interaccions no covalents, en aquest cas, les interaccions π - π i, en menor mesura, els enllaços d'hidrogen.^[23] Tots els termes són negatius perquè són forces atractives, menys la repulsió de Pauli que és positiva perquè és una repulsió, tal i com indica el seu nom.

Aquesta descomposició és molt útil per conèixer què és el que contribueix més al sistema i comprovar quin efecte hi tenen els diferents substituents.

3.8. ÈTICA I SOSTENIBILITAT.

La química computacional és la branca de la química que utilitza ordinadors per resoldre problemes químics mitjançant programari basat en la química teòrica. Que tot es faci amb ordinador implica un benefici molt gran en forma de temps, diners i salut pròpia.

Realitzar experiments al laboratori, els qual s'han d'anar repetint perquè no es coneixen les condicions idònies, provoca que es gastin quantitats molt grans de reactius i dissolvents, per tant, una pèrdua de diners i, si no es reciclen correctament, una pèrdua pel medi ambient. Aquest fet implica també una pèrdua de temps, en ocasions molt gran, fins que no s'arriba a conèixer la solució del problema químic.

A més, durant tot aquest temps el personal encarregat d'estar al laboratori s'exposa a reactius i dissolvents nocius per la salut, provocant així un augment de la probabilitat a contraure diferents tipus de malalties i, en conseqüència, patir una mort prematura.

Un altre benefici important és que es pot treballar des de casa sense cap tipus d'inconvenient. Aquest fet s'ha vist reflectit notablement aquest últim any degut a la pandèmia mundial.

Però, òbviament, no tot són beneficis. Cada càlcul que es fa requereix de cert temps, no són instantanis i depenent del sistema utilitzat pot tardar més o menys. Això implica que quant més temps es tardi més electricitat serà necessària. Per tant, un desavantatge en el qual se li hauria de trobar solució immediatament, és la gran contaminació de CO₂ que provoca.

Per tot això, els càlculs realitzats en aquest treball han sigut fets amb DFT, un dels mètodes més ràpids. A més, s'ha intentat fer-ho de manera rigorosa per tal de no equivocar-se amb els càlculs i haver-los de tornar a realitzar.

4. RESULTATS I DISCUSSIÓ.

Per tal de comprovar si bases artificials, és a dir, bases amb característiques diferents a les naturals de l'ADN, també s'adaptarien correctament a l'ADN, s'han escollit dos grups d'aquestes: les *skinny* i les *fat* (figura 3). Les *skinny* són parelles de bases de pirimidines (un anell de sis) i les *fat* parelles de bases de purines (dos anells, de cinc i de sis). D'aquesta manera, a més de comprovar si diferents bases són possibles, també es demostra si és estrictament necessària la complementarietat de mida de les bases naturals, ja que sempre són el conjunt d'una purina i una pirimidina.

Els sistemes estudiats són conjunts entre parelles de bases naturals i parelles de bases *skinny* o *fat*. Per tal de realitzar un estudi més precís, s'ha diferenciat quan les parelles de bases naturals estan a sobre de les bases artificials i quan estan a sota, ja que al no ser molècules planes podria ser que el resultat fos diferent. Però com que els resultats obtinguts han sigut molt semblants i per tal de simplificar-los, només es mostren els resultats quan les bases naturals estan a sobre de les artificials (figura 4).

Els resultats es poden dividir en tres parts. La primera és la referent a l'angle òptim dels sistemes, per saber si s'adaptarien correctament dintre l'ADN. La segona part correspon al càlcul de les interaccions π - π *stacking* en fase gas i solvent, per saber quin sistema és més estable, i a l'anàlisi de la descomposició de l'energia en fase gas, per conèixer quin terme dels implicats és el que contribueix més. Per últim, a la tercera part, tot seguint un estudi previ dut a terme per Franc Blanco on es va analitzar l'efecte dels substituents en l'enllaç d'hidrogen,^[24] s'estudia l'efecte dels substituents en les interaccions π - π *stacking*. Com a grup electrodonador (EDG) s'ha utilitzat el grup amino, NH₂, i com a grup electroacceptor (EWG) el grup nitro, NO₂. No s'ha utilitzat cap substituent neutre com seria el clor, perquè segons els resultats obtinguts per Fran Blanco no hi té cap efecte.

Per poder comparar els resultats, s'han calculat els sistemes completament naturals, els que es troben a l'ADN, per tal d'utilitzar-los de referència. Tot i així, els resultats són aproximats, ja que en cap cas s'ha tingut en compte la cadena principal de sucres i fosfats.

4.1. GENERACIÓ DE LES SUPERFÍCIES DE POTENCIALS ROTACIONALS.

Com ja s'ha comentat, el primer objectiu ha sigut trobar l'angle òptim de cada sistema. Per aconseguir-ho s'ha anat girant un dels parells de bases de 0 fins a 60 ° per etapes de 5 °. S'ha escollit aquest rang d'angles, perquè les bases de l'ADN estan entre elles al voltant dels 35 °, així que si s'allunya gaire d'aquest valor la base artificial no resultaria gaire viable.

En els resultats que es mostren, el parell que s'ha girat és el de les bases naturals (AT i GC) mentre que les bases artificials s'han quedat immòbils. El gir ha sigut dextrogir i sobre l'eix z, tal i com ho fa l'ADN. Per entendre-ho millor es pot utilitzar l'exemple del sistema AT-BP que es mostra al codi QR de la figura 6.



Figura 6. Gif rotació sistema AT-BP.

Taula 1. Angles òptims i diferències entre l'energia d'aquest amb l'energia de 35 ° dels sistemes amb la parella natural AT.

	Sistema	Angle òptim (°)	$\Delta E_{\text{opt-35}}$ (kcal/mol)
<i>Natural</i>	<i>AT-AT</i>	39	0.09
	<i>AT-GC</i>	32	0.02
<i>Fat</i>	<i>AT-BP</i>	24	0.48
	<i>AT-DX</i>	29	0.27
<i>Skinny</i>	<i>AT-CV</i>	39	0.19
	<i>AT-KT</i>	36	0.00
	<i>AT-SZ</i>	21	0.53

A la taula 1 es mostren els resultats obtinguts quan els sistemes estan formats per AT. Pels sistemes totalment naturals (AT-AT i AT-GC), el valor esperat era al voltant de 35°, però no s'ha obtingut aquest resultat. Aquest fet pot ser degut a diversos motius, un d'ells pot ser per l'aproximació que es fa al no tenir en compte les desoxiriboses ni els fosfats. L'altre motiu pot ser que hi hagi molt poca diferència d'energia entre un angle i l'altre, com és el cas, per això també es mostra la diferència d'energies entre l'angle òptim i el natural de l'ADN de 35 °.

Pel cas dels sistemes formats per una parella d'AT i una de *fat* (AT-BP i AT-DX), els angles difereixen relativament poc dels sistemes naturals. La diferència d'energia en aquests casos és major, però tot i així, suficientment petita com perquè puguin adoptar l'angle de l'ADN.

Amb els sistemes amb parelles *skinny* (AT-CV, AT-KT i AT-SZ), dos dels tres casos s'ajusten quasi perfectament a l'angle ideal, arribant fins i tot a tenir una diferència de zero. Tot i que el sistema AT-SZ és en el qual s'han obtingut pitjors resultats.

Taula 2. Angles òptims i diferències entre l'energia d'aquest amb l'energia de 35 ° dels sistemes amb la parella natural GC.

	Sistema	Angle òptim (°)	$\Delta E_{\text{opt-35}}$ (kcal/mol)
<i>Natural</i>	<i>GC-GC</i>	47	0.45
	<i>GC-AT</i>	32	0.03
<i>Fat</i>	<i>GC-BP</i>	27	0.34
	<i>GC-DX</i>	34	0.00
<i>Skinny</i>	<i>GC-CV</i>	26	0.33
	<i>GC-KT</i>	20	2.49
	<i>GC-SZ</i>	22	1.10

A la taula 2 es poden observar els resultats obtinguts quan els sistemes estan formats per GC. El resultat obtingut pel sistema GC-GC ha sigut el menys esperat, ja que al ser un dels sistemes de referència hauria de donar al voltant dels 35 °, encara que la diferència d'energies sigui prou bona. Aquest fet deu ser degut a que el sistema s'ha creat amb la complementarietat GC-GC i probablement sigui més estable el sistema GC-CG.

En els sistemes formats per GC, per contra als formats per AT, els resultats obtinguts pels *fat* són millors que pels *skinny*, ja que només un de tres ha donat resultats bons, és a dir, només un de tres té l'angle de rotació amb energia mínima molt semblant al de les bases naturals. Això pot voler dir que seria més estable un ADN format amb bases artificials, si aquest fos ric en AT i no tant en GC.

Taula 3. Angles òptims i diferències entre l'energia d'aquest amb l'energia de 35 ° dels sistemes amb les mateixes parelles.

	Sistema	Angle òptim (°)	$\Delta E_{\text{opt-35}}$ (kcal/mol)
<i>Natural</i>	<i>AT-AT</i>	39	-0.09
	<i>CG-CG</i>	47	-0.45
<i>Fat</i>	<i>BP-BP</i>	37	-0.02
	<i>DX-DX</i>	30	-0.24
<i>Skinny</i>	<i>CV-CV</i>	160	-7.68
	<i>KT-KT</i>	180	-8.88
	<i>SZ-SZ</i>	180	-8.67

També es van calcular els angles òptims dels sistemes formats per les mateixes parelles (taula 3). Tan els sistemes naturals com els *fat*, han donat resultats ajustables a un sistema real. El problema ha vingut amb els sistemes *skinny*, en cap d'ells s'ha trobat un mínim en el rang de 0 a 60 °, així que s'ha fet un escombrat de 0 a 360 ° i s'ha trobat que l'angle òptim és al voltant dels 180 °, és a dir, que els sistemes són més estables amb les bases girades, fet que no podria donar-se a l'ADN degut a l'impediment de la cadena de sucres i fosfats. Així que aquests sistemes ja no s'utilitzen més a partir d'aquí i els resultats queden a l'espera d'un estudi més exhaustiu per tal de determinar el perquè d'aquests resultats.

Per tant, segons els resultats obtinguts es pot veure com hi ha un rang ampli d'angles, gairebé tots estan entre 20 i 40 °. A priori pot semblar que no és possible utilitzar aquests sistemes, però l'ADN no és estàtic, no sempre està a 35 ° i les diferències d'energia entre els angles òptims i 35 ° és, en quasi tots els casos, insignificant. Una diferència que es pot pagar perfectament.

4.2. INTERACCIÓ π - π STACKING I ANÀLISI DE LA DESCOMPOSICIÓ DE L'ENERGIA.

La interacció π - π *stacking* és la principal responsable de l'estabilitat entre les parelles de bases a l'ADN. Al definir les dues parelles de bases com a fragments dins un sistema, l'energia que s'obté és l'*stacking*. Quan més negativa sigui aquesta, més favorable serà el sistema. Per defecte, els càlculs es fan en fase gas, però com que el medi real de l'ADN és en solució aquosa, també s'ha calculat l'energia d'interacció en aigua.

A les taules que venen a continuació, agafant com a referència els resultats dels sistemes naturals, s'han pintat de color verd els resultats que s'han vist estabilitzats, de color vermell si s'han desestabilitzats i, si quasi no s'han vist modificats, s'han deixat de color negre.

Taula 4. Energies d'interacció π - π *stacking* en fase gas i en aigua dels sistemes amb la parella natural AT.

	Sistema	Gas (kcal/mol)	Aigua (kcal/mol)
<i>Natural</i>	<i>AT-AT</i>	-14.6	-11.0
	<i>AT-GC</i>	-13.4	-11.4
<i>Fat</i>	<i>AT-BP</i>	-17.4	-13.2
	<i>AT-DX</i>	-13.5	-11.9
<i>Skinny</i>	<i>AT-CV</i>	-16.8	-11.5
	<i>AT-KT</i>	-14.8	-11.7
	<i>AT-SZ</i>	-14.7	-12.6

A la taula 4 es mostra que els resultats dels sistemes naturals són semblants als previstos, segons l'article de Poater et al. (2014),^[8] on per la parella AT-AT van obtenir -12.2 kcal/mol en fase gas i -9.8 kcal/mol en aigua. En general, totes les interaccions són força semblants, però cal destacar dos casos: l'AT-BP i l'AT-CV. Tan un com l'altre, en fase gas, milloren els resultats naturals, provocant una estabilitat major a l'original. A més, l'AT-BP millora inclús en aigua, quan amb els altres casos pràcticament no varia.

Per tal de conèixer millor a què és deguda aquesta energia d'interacció, s'ha dut a terme una anàlisi de la descomposició de l'energia (EDA) en fase gas. Aquesta descomposició conté la repulsió de Pauli (ΔE_{Pauli}), la interacció electrostàtica (ΔV_{elstat}), la interacció orbitalària (ΔE_{oi}) i l'energia de dispersió (ΔE_{disp}). La suma de totes elles proporciona l'energia d'interacció (ΔE_{int}), en aquest cas, la interacció π - π *stacking*.

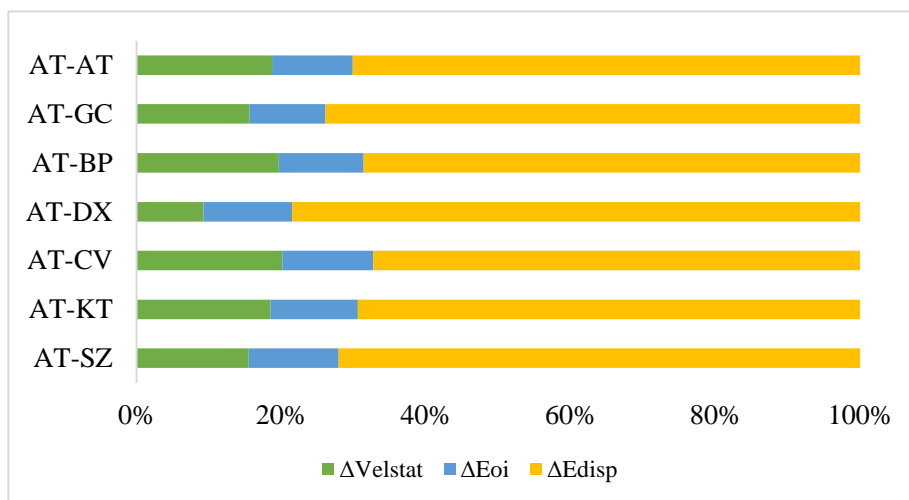


Figura 7. Representació gràfica de la contribució de les forces atractives dels sistemes formats per AT.

A la figura 7 hi ha representat el tant per cent de contribució de les forces atractives per tal de veure-ho més clarament. En verd hi ha la interacció electrostàtica, en blau la orbitalària i en groc la de dispersió. Gairebé en tots els sistemes el tant per cent de les contribucions és semblants tot i que, inconfusiblement, en tots els casos el terme que contribueix més a l'estabilitat del sistema és l'energia de dispersió.

Taula 5. Anàlisi de la descomposició de l'energia en kcal/mol pels sistemes formats per AT.

	Sistema	$\Delta E_{\text{paulí}}$	ΔV_{estat}	ΔE_{oi}	ΔE_{disp}	ΔE_{int}
<i>Natural</i>	AT-AT	16.4	-5.8	-3.5	-21.7	-14.6
	AT-GC	17.6	-4.9	-3.2	-22.9	-13.4
<i>Fat</i>	AT-BP	20.2	-7.4	-4.4	-25.8	-17.4
	AT-DX	16.3	-2.8	-3.7	-23.5	-13.5
<i>Skinny</i>	AT-CV	17.1	-6.8	-4.3	-22.8	-16.8
	AT-KT	18.8	-6.2	-4.1	-23.3	-14.8
	AT-SZ	20.1	-5.4	-4.3	-25.1	-14.7

Per tal de veure-ho amb més detall, a la taula 5 també s'hi mostra la repulsió de Pauli i l'energia d'interacció. La repulsió de Pauli és en la majoria de casos quasi tan gran, en valor absolut, com l'energia de dispersió. Que l'energia de dispersió sigui el terme que més contribueix entra dintre els resultats esperats, ja que les interaccions π - π *stacking* són interaccions no covalents.

Concretament, el sistema AT-BP augmenta la repulsió de Pauli, però tots els altres termes també augmenten bastant, provocant una energia d'interacció més negativa que l'original. En el cas d'AT-CV, tant la repulsió de Pauli com l'energia de dispersió es queden pràcticament igual, però les interaccions electrostàtica i orbitalària augmenten donant una energia d'interacció més estable. Els altres casos gairebé no veuen modificada l'energia d'interacció, perquè quan augmenta o disminueix un terme, hi ha un altre que el contraresta.

Taula 6. Energies d'interacció π - π *stacking* en fase gas i en aigua dels sistemes amb la parella natural GC.

	Sistema	Gas (kcal/mol)	Aigua (kcal/mol)
<i>Natural</i>	<i>GC-GC</i>	-13.0	-9.9
	<i>GC-AT</i>	-13.6	-11.6
<i>Fat</i>	<i>GC-BP</i>	-18.5	-12.9
	<i>GC-DX</i>	-15.2	-10.5
<i>Skinny</i>	<i>GC-CV</i>	-21.0	-12.2
	<i>GC-KT</i>	-16.5	-12.7
	<i>GC-SZ</i>	-17.6	-12.7

En el cas de la taula 6, també s'han obtingut resultats comparables a l'estudi de Poater et al. (2014),^[8] on per la parella GC-GC van obtenir -9.8 kcal/mol en fase gas i -9.7 kcal/mol en aigua. Al contrari que amb AT, on quasi tots els sistemes donaven números semblants, aquí en tots els casos en fase gas s'estabilitza i en fase aigua també. A l'igual que passava amb l'AT, els sistemes més estabilitzats són els quals estan formats per BP i CV

Un altre cop, per tal de conèixer quin terme és el que contribueix més a aquesta energia, s'ha fet una anàlisi de la descomposició de l'energia que es mostra a la taula 7.

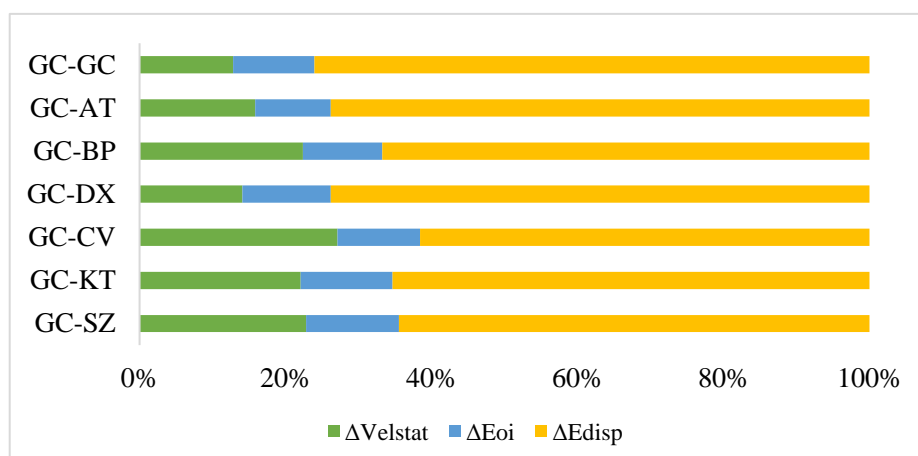


Figura 8. Representació gràfica de la contribució de les forces atractives dels sistemes formats per GC.

A la figura 8, es pot veure com clarament altra vegada el terme que influeix més és l'energia de dispersió, tot i que aquesta vegada la interacció electrostàtica contribueix una mica més.

En els sistemes GC-BP, GC-KT i GC-SZ augmenta la repulsió de Pauli, però totes les forces atractives també ho fan notablement, això provoca que els tres casos siguin més favorables. En el cas GC-CV, igual que passava amb l'AT, la repulsió i la dispersió no es veuen afectades, però sí que ho fan les interaccions electrostàtica i orbitalària, conduint al sistema a una menor energia d'interacció.

Taula 7. Anàlisi de la descomposició de l'energia en kcal/mol pels sistemes formats per GC.

	Sistema	ΔE_{pauli}	ΔV_{estat}	ΔE_{oi}	ΔE_{disp}	ΔE_{int}
<i>Natural</i>	<i>GC-GC</i>	14.2	-3.5	-3.0	-20.7	-13.0
	<i>GC-AT</i>	17.6	-4.9	-3.2	-23.0	-13.6
<i>Fat</i>	<i>GC-BP</i>	18.8	-8.3	-4.1	-24.9	-18.5
	<i>GC-DX</i>	14.7	-4.2	-3.6	-22.0	-15.2
<i>Skinny</i>	<i>GC-CV</i>	16.4	-10.1	-4.2	-23.0	-21.0
	<i>GC-KT</i>	23.4	-8.8	-5.0	-26.1	-16.5
	<i>GC-SZ</i>	22.2	-9.1	-5.1	-25.7	-17.6

Dels resultats obtinguts d'aquesta part, es pot deduir que no hi ha cap sistema desfavorable, que els que tenen l'energia d'interacció més negativa i, per tant, són més estables, són els sistemes formats per BP i CV i que el terme que contribueix més a aquesta energia d'interacció, tal i com era d'esperar, és l'energia de dispersió.

4.3. EFECTES DELS SUBSTITUENTS EN LA INTERACCIÓ π - π STACKING.

Mitjançant una anàlisi de la descomposició de l'energia (EDA), s'ha realitzat un breu estudi per conèixer la influència que tenen substituents de característiques diferents sobre els sistemes formats per CV, pel grup *skinny*, i BP, pel grup *fat*. S'han escollit aquestes parelles de bases perquè, en els càlculs anteriors, són les que han donat millors resultats. Els substituents sempre s'han posat a la parella artificial i els triats han sigut com a grup electrodonador (EDG) el grup amino, NH_2 , i com a grup electroacceptor (EWG) el grup nitro, NO_2 .

4.3.1. SISTEMES FORMATS PER CV.

La posició escollida ha sigut la marcada com a P1 a la figura 9, és a dir, a l'àtom del carboni 4. Les posicions 5 i 19 no són possibles degudes a un impediment estèric i la posició 3 i 18 perquè és on es col·loca la deoxyribose.

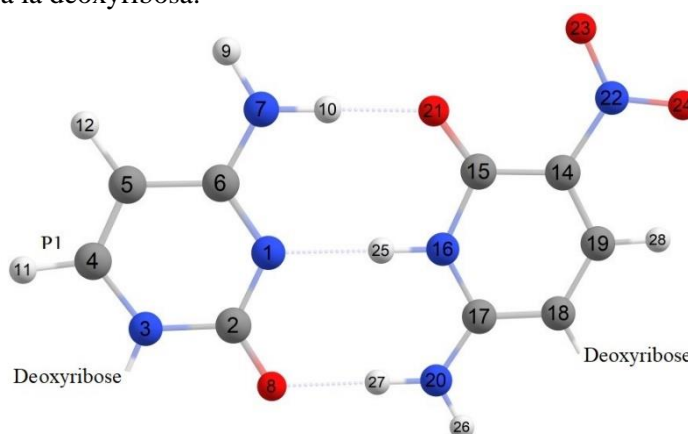


Figura 9. Parella de bases CV amb la posició substituída marcada.

- AT-CV.

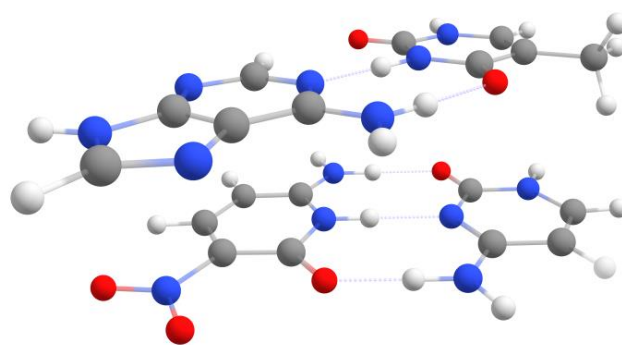


Figura 10. Sistema AT-CV.

Taula 8. Anàlisi de la descomposició de l'energia en kcal/mol del sistema AT-CV amb diferents substituents i sense.

Substituent	<i>H</i>	<i>EDG</i>	<i>EWG</i>
$\Delta E_{\text{paulí}}$	17.1	17.3	18.2
ΔV_{estat}	-6.8	-6.2	-7.6
ΔE_{oi}	-4.3	-4.3	-4.7
ΔE_{disp}	-22.8	-23.4	-24.0
ΔE_{int}	-16.8	-16.6	-18.1

El primer sistema que s'ha substituït ha sigut l'AT-CV, representat a la figura 10. A la taula 8, es poden veure els resultats obtinguts de l'EDA. El substituent electrodonador no té cap efecte respecte el sistema sense substituents. Per contra, el grup electroacceptor provoca un augment de la repulsió, però també de la dispersió i, en menor mesura, de la interacció electrostàtica, així que aconseguix una menor energia d'interacció.

- GC-CV.

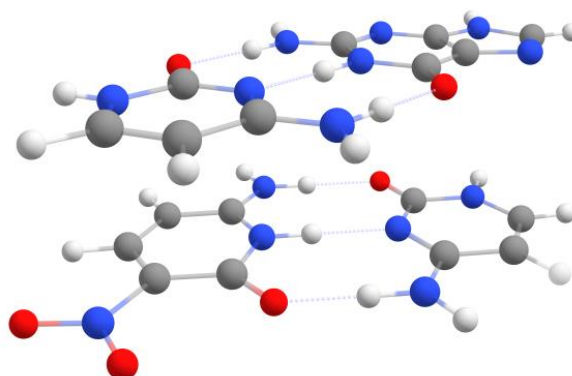


Figura 11. Sistema GC-CV.

Taula 9. Anàlisi de la descomposició de l'energia en kcal/mol del sistema GC-CV amb diferents substituents i sense.

Substituent	<i>H</i>	<i>EDG</i>	<i>EWG</i>
$\Delta E_{\text{paulí}}$	16.4	15.0	15.9
ΔV_{estat}	-10.1	-9.6	-10.9
ΔE_{oi}	-4.2	-4.2	-4.3
ΔE_{disp}	-23.0	-22.2	-23.3
ΔE_{int}	-21.0	-21.0	-22.6

El següent sistema ha sigut el GC-CV de la figura 11. A la taula 9 s'observen els resultats de l'anàlisi de la descomposició de l'energia. En aquest cas, igual que en l'anterior, l'únic substituent que afecta és l'electroacceptor i ho fa disminuint la repulsió de Pauli.

- CV-CV.

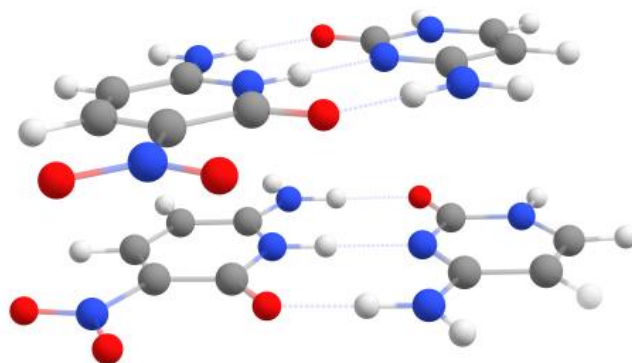


Figura 12. Sistema CV-CV.

Taula 10. Anàlisi de la descomposició de l'energia en kcal/mol del sistema CV-CV amb diferents substituents i sense.

Substituent	<i>H</i>	<i>EDG</i>	<i>EWG</i>
$\Delta E_{\text{paulí}}$	17.4	16.7	17.6
ΔV_{estat}	-2.5	-2.0	-3.9
ΔE_{oi}	-3.8	-3.7	-4.1
ΔE_{disp}	-23.1	-22.9	-23.5
ΔE_{int}	-12.0	-11.8	-13.9

El tercer i últim sistema per la parella CV, és el representat a la figura 12. Tal i com es veu a la taula 10, igual que els altres dos casos, l'únic substituent que afecta és l'electroacceptor, però aquesta vegada perquè augmenta la interacció electrostàtica.

Per tant, amb un sistema format per la parella de bases CV, si es vol estabilitzar s'haurà d'afegir un substituent EWG, ja que els EDG no afecten gaire.

4.3.2. SISTEMES FORMATS PER BP.

Les posicions escollides són les marcades a la figura 13 com a P1, per la base B, i com a P2 i P3, per la base P. La posició 1 correspon a l'àtom del carboni 8, la posició 2 a l'àtom del carboni 23 i la posició 3 a l'àtom del carboni 24. A les posicions 4 i 25 és on s'enllaça la deoxyribosa.

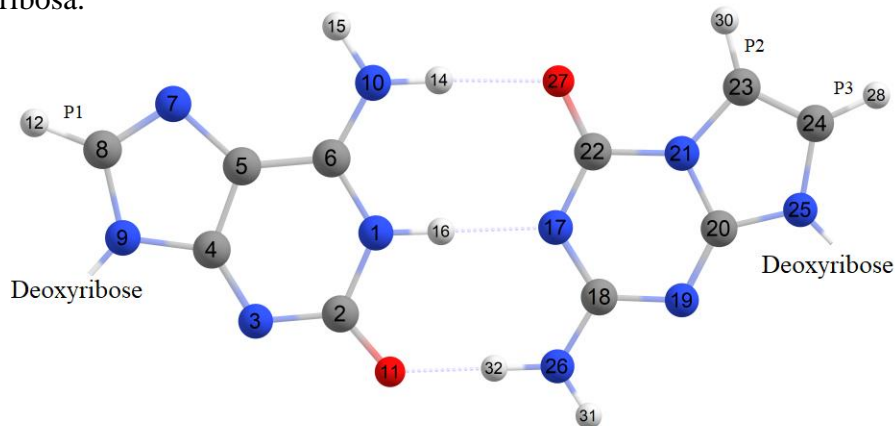


Figura 13. Parella de bases BP amb les posicions substituïdes marcades com a P1, P2 i P3.

- AT-BP.

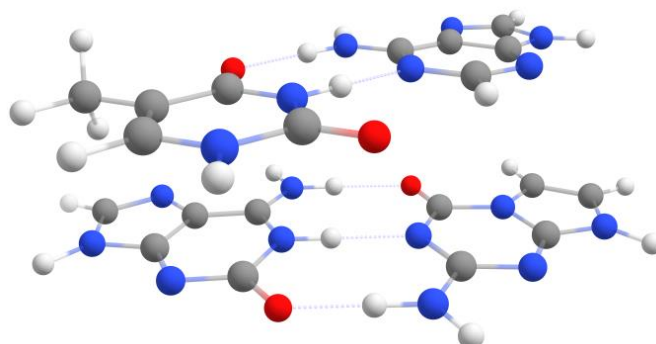


Figura 14. Sistema AT-BP.

Taula 11. Anàlisi de la descomposició de l'energia en kcal/mol del sistema AT-BP amb diferents substituents i posicions i sense.

Substituent	<i>H</i>			<i>EDG</i>			<i>EWG</i>		
	-	1	2	3	1	2	3		
$\Delta E_{\text{paulí}}$	20.2	15.8	16.6	14.4	15.0	22.1	14.6		
ΔV_{estat}	-7.4	-3.9	-4.8	-2.8	-4.9	-6.0	-4.0		
ΔE_{oi}	-4.4	-3.5	-3.5	-3.4	-3.3	-4.1	-3.6		
ΔE_{disp}	-25.8	-23.1	-24.0	-22.3	-22.7	-24.9	-22.6		
ΔE_{int}	-17.4	-14.7	-15.7	-14.1	-15.9	-12.9	-15.6		

El primer sistema estudiat ha sigut el que es mostra a la figura 14. Tal i com es pot observar a la taula 11, no hi ha hagut cap posició ni substituent que millorés l'estabilitat original. Tot i que en gairebé tots els casos la repulsió disminueix, no és suficient per

contrarestar la disminució de les forces atractives. El pitjor dels casos ha sigut a la posició 2 amb el substituent electroacceptor, tots els seus termes han empitjorat, provocant un augment de l'energia d'interacció de fins a 4.5 kcal/mol.

- GC-BP.

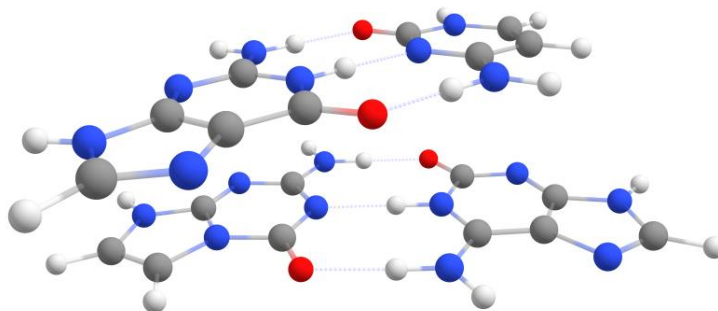


Figura 15. Sistema GC-BP.

Taula 12. Anàlisi de la descomposició de l'energia en kcal/mol del sistema GC-BP amb diferents substituents i posicions i sense.

Substituent	<i>H</i>			<i>EDG</i>			<i>EWG</i>		
	Posició	-	1	2	3	1	2	3	
$\Delta E_{\text{paulí}}$		18.8	17.4	19.5	16.3	16.3	19.5	16.4	
ΔV_{estat}		-8.3	-8.4	-9.5	-7.0	-8.5	-9.7	-8.7	
ΔE_{oi}		-4.1	-3.8	-3.9	-3.7	-3.6	-4.5	-4.1	
ΔE_{disp}		-24.9	-23.6	-25.1	-23.3	-23.0	-24.7	-23.8	
ΔE_{int}		-18.5	-18.4	-18.9	-17.8	-18.8	-19.4	-20.2	

Pel sistema GC-BP (figura 15) els resultats no tenen res a veure amb el cas anterior (taula 12). Per la posició 1, tan si el substituent és electrodonador o electroacceptor, la repulsió disminueix, però la dispersió també, fet que provoca que les energies d'interacció quedin pràcticament iguals a l'original. A la posició 2 quan el substituent és electrodonador, el resultat final és el mateix, però degut a què la repulsió augmenta i la interacció electrostàtica també. Per contra, quan el substituent és electroacceptor els efectes són els mateixos, però l'energia d'interacció augmenta més. A la posició 3 hi ha els dos extrems, quan el substituent és EDG, tot i la disminució de la repulsió, les forces atractives disminueixen i provoquen una desestabilització comparat a l'inicial. Quan el substituent és EWG, la repulsió disminueix fins a 2.4 kcal/mol, provocant una estabilització del sistema.

- BP-BP.

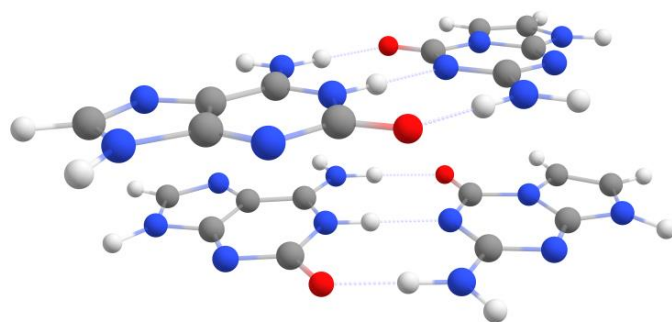


Figura 16. Sistema BP-BP.

Taula 13. Anàlisi de la descomposició de l'energia en kcal/mol del sistema BP-BP amb diferents substituents i posicions i sense.

Substituent	<i>H</i>			<i>EDG</i>			<i>EWG</i>		
	Posició -	1	2	3	1	2	3		
$\Delta E_{\text{paulí}}$	16.3	17.4	19.1	15.8	16.5	20.9	16.0		
ΔV_{estat}	-3.8	-4.0	-5.3	-2.6	-4.9	-5.7	-5.1		
ΔE_{oi}	-3.6	-3.8	-3.9	-3.7	-3.6	-4.3	-4.0		
ΔE_{disp}	-24.2	-25.0	-26.2	-24.2	-24.6	-25.9	-24.7		
ΔE_{int}	-15.3	-15.3	-16.4	-14.7	-16.6	-15.0	-17.8		

Per últim, el sistema BP-BP (figura 16), com es mostra a la taula 13, presenta un empitjorament del sistema quan a la posició 3 es posa un EDG i un millorament quan es posa un EWG, tal i com passava en l'últim cas. Tan la posició 1 quan es posa un EDG, com la posició 2 quan es posa un EWG es queden pràcticament igual, ja que en ambdós casos augmenta la repulsió, però augmenta la dispersió. A la posició 2 amb substituent EDG i a la posició 1 amb substituent EWG, hi ha una estabilització final degut a l'augment de la interacció electrostàtica.

Per tant, cada sistema amb la parella de bases BP té un afecte diferent degut al tipus de substituent. Tot i que, en general, proporciona millors resultats un substituent electroacceptor d'electrons.

5. CONCLUSIONS.

The main objective of this work was to check if pairs of artificial bases form systems stable enough to be placed in DNA. Two groups of artificial bases were taken, skinny and fat, which apart from being unnatural do not meet the size complementarity either.

To achieve the goal, each pair of artificial bases was joined with a pair of natural bases. The optimal twist angle was found, the π - π stacking interactions in gas and water phase, an analysis of the energy decomposition was made and, finally, the effect of the substituents on the π - π stacking interaction was studied in pairs formed by CV and BP. The selected positions are shown in figures 9 and 13, respectively.

Regarding twist angle, a fairly wide range has been obtained, almost all of which are between 20 and 40 ° when the ideal would have been about 35 °. A priori, these may seem not very good results, but DNA is not static and therefore it not always form a 35 °. In addition, the difference between the energy of the optimal angle and that of 35 ° is, in almost all cases, insignificant, i. e., the bases are easily adaptable.

About the π - π stacking interaction and the analysis of energy decomposition, all systems have been found to be equally or more favourable than natural ones. For the fat ones the most favorable ones turned out to be the AT-BP and the GC-BP and for the skinny ones the AT-CV and GC-CV. In addition, according to the EDA, the term that contributes most to this interaction energy, as expected, has been the dispersion energy.

The systems formed by the base pair CV, have not been affected by the incorporation of an electrodonor substituent. On the contrary, all the cases have been stabilized by putting an electroacceptor substituent. For systems formed by the base pair BP, there has been more variability, albeit generally placing an electroacceptor substituent yields better results than with an electrodonator.

Therefore, it can be clearly said that the results obtained have been very good, as they show that the pairs of artificial bases are easily adaptable to the structure of the original DNA, especially the fat pair BP and the skinny pair CV. In addition, it is also shown that size complementarity is not strictly necessary. These results are an open door for further research.

6. REFERÈNCIES.

- [1] Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. In *Developmental Biology* (Vol. 278, Issue 2, pp. 274–288). Academic Press Inc.
<https://doi.org/10.1016/j.ydbio.2004.11.028>
- [2] Levene, P., Biochem, Z., Levene, P. A., Jacobs, W. A., Ber, T., Ges ; Levene, P. A., Forge, L., Levene, F. B., Jacobs, P. A., Thannhauser, W. A., Dorfmiiller, S. J., Jones, G., & Germann, W. (1919). THE STRUCTURE OF YEAST NUCLEIC ACID. IV. AMMONIA HYDROLYSIS. In *Ber. them. Ges* (Vol. 608, Issue 2).
- [3] Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738.
<https://doi.org/10.1038/171737a0>
- [4] Franklin, R. E., & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, *171*(4356), 740–741. <https://doi.org/10.1038/171740a0>
- [5] Grome, M. W., & Isaacs, F. J. (2021). ZTCG: Viruses expand the genetic alphabet. *Science*, *372*(6541).
- [6] Kirnos, M. D., Khudyakov, I. Y., Alexandrushkina, N. I., & Vanyushin, B. F. (1977a). 2-Amino adenine is an adenine substituting for a base in S-2L cyanophage DNA. *Nature*, *270*(5635), 369–370. <https://doi.org/10.1038/270369a0>
- [7] Khudyakov, I. Y., Kirnos, M. D., Alexandrushkina, N. I., & Vanyushin, B. F. (1978). Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine. *Virology*, *88*(1), 8–18. [https://doi.org/10.1016/0042-6822\(78\)90104-6](https://doi.org/10.1016/0042-6822(78)90104-6)
- [8] Poater, J., Swart, M., Bickelhaupt, F. M., & Fonseca Guerra, C. (2014). B-DNA structure and stability: The role of hydrogen bonding, π - π Stacking interactions, twist-angle, and solvation. *Organic and Biomolecular Chemistry*, *12*(26), 4691–4700. <https://doi.org/10.1039/c4ob00427b>
- [9] Hoshika, S., Singh, I., Switzer, C., Molt, R. W., Leal, N. A., Kim, M. J., Kim, M. S., Kim, H. J., Georgiadis, M. M., & Benner, S. A. (2018). “Skinny” and “fat” DNA: Two New Double Helices. *Journal of the American Chemical Society*, *140*(37), 11655–11660. <https://doi.org/10.1021/jacs.8b05042>
- [10] *DFT*. (n.d.). Retrieved May 27, 2021, from <https://www.cup.uni-muenchen.de/ch/compchem/energy/dft1.html>
- [11] Becke, Axel D. (2014). Perspective: Fifty years of density-functional theory in chemical physics. *Journal of Chemical Physics*, *140*(18), 18–301.
<https://doi.org/10.1063/1.4869598>
- [12] Hohenberg, P., & Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review*, *136*(3B), B864. <https://doi.org/10.1103/PhysRev.136.B864>
- [13] Kohn, W., & Sham, L. J. (1965). Self-consistent equations including exchange and

- correlation effects. *Physical Review*, 140(4A), A1133.
<https://doi.org/10.1103/PhysRev.140.A1133>
- [14] Perdew, J. P., Burke, K., & Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Physical Review Letters*, 77(18), 3865–3868.
<https://doi.org/10.1103/PhysRevLett.77.3865>
- [15] Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., & Fiolhais, C. (1992). Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B*, 46(11), 6671–6687. <https://doi.org/10.1103/PhysRevB.46.667>
- [16] G. te Velde, F.M. Bickelhaupt, E.J. Baerends, C. Fonseca Guerra, S.J.A. van Gisbergen, J.G. Snijders and T. Ziegler, *Chemistry with ADF*, Journal of Computational Chemistry 22, 931 (2001). <https://doi.org/10.1002/jcc.1056>
- [17] ADF 2021.1, E.J. Baerends, T. Ziegler, A.J. Atkins, J. Autschbach, O. Baseggio, D. Bashford, A. Bérces, F.M. Bickelhaupt, C. Bo, P.M. Boerrigter, C. Cappelli, L. Cavallo, C. Daul, D.P. Chong, D.V. Chulhai, L. Deng, R.M. Dickson, J.M. Dieterich, F. Egidi, D.E. Ellis, M. van Faassen, L. Fan, T.H. Fischer, A. Förster, C. Fonseca Guerra, M. Franchini, A. Ghysels, A. Giammona, S.J.A. van Gisbergen, A. Goetz, A.W. Götz, J.A. Groeneveld, O.V. Gritsenko, M. Grüning, S. Gusarov, F.E. Harris, P. van den Hoek, Z. Hu, C.R. Jacob, H. Jacobsen, L. Jensen, L. Joubert, J.W. Kaminski, G. van Kessel, C. König, F. Kootstra, A. Kovalenko, M.V. Krykunov, P. Lafiosca, E. van Lenthe, D.A. McCormack, M. Medves, A. Michalak, M. Mitoraj, S.M. Morton, J. Neugebauer, V.P. Nicu, L. Noodleman, V.P. Osinga, S. Patchkovskii, M. Pavanello, C.A. Peeples, P.H.T. Philipsen, D. Post, C.C. Pye, H. Ramanantoanina, P. Ramos, W. Ravenek, M. Reimann, J.I. Rodríguez, P. Ros, R. Rüger, P.R.T. Schipper, D. Schlüns, H. van Schoot, G. Schreckenbach, J.S. Seldenthuis, M. Seth, J.G. Snijders, M. Solà, M. Stener, M. Swart, D. Swerhone, V. Tognetti, G. te Velde, P. Vernooijs, L. Versluis, L. Visscher, O. Visser, F. Wang, T.A. Wesolowski, E.M. van Wezenbeek, G. Wiesenekker, S.K. Wolff, T.K. Woo, A.L. Yakovlev, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, <http://www.scm.com>.
- [18] Grimme, S. (2004). Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12), 1463–1473. <https://doi.org/10.1002/jcc.20078>
- [19] Grimme, S., Ehrlich, S., & Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 32(7), 1456–1465. <https://doi.org/10.1002/jcc.21759>
- [20] Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6), 3098–3100.
<https://doi.org/10.1103/PhysRevA.38.3098>
- [21] Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti

correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2), 785–789. <https://doi.org/10.1103/PhysRevB.37.785>

- [22] *Análisis de enlaces químicos: software para química y materiales Software para química y materiales*. (n.d.). Retrieved May 6, 2021, from <https://www.scm.com/applications/chemical-bonding-analysis/>
- [23] Riley, K. E., & Hobza, P. (2013). On the importance and origin of aromatic interactions in chemistry and biodisciplines. *Accounts of Chemical Research*, 46(4), 927–936. <https://doi.org/10.1021/ar300083h>
- [24] Blanco Martín, F. (2020). *Skinny and fat DNA: Substituent effect* [Trellat fi de grau]. Universitat de Girona.