Article

Molecular Dynamics Simulations Predict that rSNP Located in the *HNF-1α* Gene Promotor Region Linked with MODY3 and Hepatocellular Carcinoma Promotes Stronger Binding of the HNF-4α Transcription Factor

Eva Španinger [1] **, Uroš Potočnik** [1,2] **and Urban Bren** [1,3,*]

[1] Faculty of Chemistry and Chemical Engineering, University of Maribor, Smetanova ulica 17, SI-2000 Maribor, Slovenia; eva.spaninger@gmail.com (E.Š.); uros.potocnik@um.si (U.P.)

[2] Faculty of Medicine, University of Maribor, Taborska 8, SI-2000 Maribor, Slovenia

[3] Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, SI-6000 Koper, Slovenia

* Correspondence: urban.bren@um.si; Tel.: +386-2-2294-421

Abstract: Our study aims to investigate the impact of the Maturity-onset diabetes of the young 3 disease-linked rSNP rs35126805 located in the *HNF-1α* gene promotor on the binding of the transcription factor HNF-4α and consequently on the regulation of *HNF-1α* gene expression. Our focus is to calculate the change in the binding affinity of the transcription factor HNF-4α to the DNA, caused by the regulatory single nucleotide polymorphism (rSNP) through molecular dynamics simulations and thermodynamic analysis of acquired results. Both root-mean-square difference (RMSD) and the relative binding free energy $\Delta\Delta G_{bind}$ reveal that the HNF-4α binds slightly more strongly to the DNA containing the mutation (rSNP) making the complex more stable/rigid, and thereby influencing the expression of the *HNF-1α* gene. The resulting disruption of the HNF-4α/HNF-1α pathway is also linked to hepatocellular carcinoma metastasis and enhanced apoptosis in pancreatic cancer cells. To the best of our knowledge, this represents the first study where thermodynamic analysis of the results obtained from molecular dynamics simulations is performed to uncover the influence of rSNP on the protein binding to DNA. Therefore, our approach can be generally applied for studying the impact of regulatory single nucleotide polymorphisms on the binding of transcription factors to the DNA.

Keywords: HNF-1α; HNF-4α; MODY3; rSNP; molecular dynamics; binding free energy

1. Introduction

Studying the effect of single nucleotide polymorphisms (SNPs) in the coding part of the genome directly impacting protein structures has been at the center of attention for the vast majority of the scientific community. On the other hand, besides the somatic and germline mutations in oncogenes and tumor-suppressor genes, genome-wide association studies (GWASes) suggest that even germline single-nucleotide polymorphisms located in introns are associated with altered cancer risks [1]. The poorly investigated non-coding part of the genome also contains sequences for the regulation of gene expression [2,3]. Consequentially, only a few available databases on regulatory SNPs (rSNPs) exist, and their biological roles in disease (cancer) development, progression, and response to therapy ought to be investigated further [1,4].

Even though the Protein Data Bank (PDB) contains an ever-increasing number of structures, finding a human transcription factor-DNA complex especially highly homologous to a sequence containing rSNPs represents quite a challenge. Therefore, we had only a few options to choose from. We have successfully found a complex of hepatocyte nuclear factor 4 alpha (HNF-4α) with DNA. A study of an Italian family linked the SNP in the *hepatocyte nuclear factor 1 alpha (HNF-1α)* gene regulatory region with the Maturity-onset diabetes of the young 3 (MODY3) as a result of disruption of the transcription factor HNF-4α binding site [5].

Accounting for up to two percent of all diabetes cases in the United States, MODY represents the prevailing form of monogenic diabetes [6,7]. Various gene mutations demonstrated to cause MODY decrease the ability of the pancreas to produce insulin, resulting in high blood glucose levels and, with time, even in damaged body tissues, particularly the eyes, kidneys, nerves, and blood vessels [7]. Based on the gene mutations, MODY can be divided into six types among, which MODY3 represents the most frequent one. MODY3 is linked with an autosomal dominant mutation in the Transcription Factor 1 gene that affects the HNF-1α protein [5,7]. Even though MODY3 represents a permanent type of diabetes, it can be, with a few exceptions, initially treated using a special diet and/or oral sulfonylureas [5,7].

The hepatocyte nuclear factors (HNFs) are already regarded as promising biomarkers for prognostic predictions and as potential therapeutic targets in cancer since they play an important role in the sustenance of solid tumors, however with numerous associated downstream targets they are not easily druggable [8,9]. The *HNF-1α* gene encodes for homeobox A protein nuclear transcription factor essential for the expression of several hepatic genes participating in detoxification, homeostasis, and metabolism of glucose, lipids, steroids, and amino acids [5,10]. Reported consequences of mutations in this gene include MODY3, type II diabetes, hepatic adenomas, and pancreatic ductal adenocarcinoma [7,10–13]. The expression of the *HNF-1α* gene is regulated by the transcription factor HNF-4α that binds DNA as a homodimer [8,14]. It is presumed that this gene plays a role in the development of the liver, kidneys, and intestines [14]. It has been found that both *HNF-1α* and *HNF-4α* regulate not only the function of differentiated beta-cells, but the growth and function of islet beta-cells as well [11]. Irregular expression or mutations in the *HNF-4α* gene promote or prevent definitive endoderm differentiation, cause monogenic autosomal dominant non-insulin-dependent diabetes mellitus type I, and possibly contribute to the progression of hepatocellular carcinoma [14–16]. Additionally, the disruption of the HNF-4α/HNF-1α pathway may form the important molecular mechanism of renal cell carcinogenesis [8].

Molecular dynamics (MD) simulations have been previously used for exploring the DNA binding sites of various proteins (including transcription factors) exploiting three different approaches. In the first one, the focus is on the DNA and its behavior to reveal the recognition/binding motifs on the DNA [17] and/or to improve the accuracy in the prediction of binding sites [18]. The second one focuses on the protein structure and how various mutations in proteins or their specific domains affect their ability to bind to the DNA [19–22]. The third approach focuses on studying the protein–DNA complexes to uncover types of interactions between them and the binding affinity of proteins towards DNA [23–26].

A majority of studies focus on large variations in DNA or protein structures, however, even the smallest variation like single nucleotide polymorphism (SNP) can have detrimental [27] and even oncogenic effects [1].

The aim of our study is to elucidate the impact of the MODY3 disease-linked rSNP rs35126805 located in the non-coding DNA region, in the *HNF-1α* gene promotor -58 (upstream), on the binding of the transcription factor HNF-4α, and the regulation of *HNF-1α* gene expression. Our focus was to investigate the change in the binding affinity of the transcription factor HNF-4α to the DNA, caused by the studied rSNP through molecular dynamics simulations and thermodynamic analysis of acquired results. To our knowledge, this is the first study where thermodynamic analyses of the obtained results

from molecular dynamics simulations were performed to uncover the influence of rSNP on the protein binding to DNA.

Both atom-positional root-mean-square difference (RMSD) and the binding free energy ΔG_{bind} reveal that there is a significant difference between the structures and behavior of the DNA-HNF-4α complex including the wild-type base pair and the rSNP. Therefore, our approach can be generalized for studying the impact of regulatory single nucleotide polymorphisms on the binding of transcription factors to the DNA.

2. Methods

2.1. Preparation of the Starting Structures for the MD Simulation

The complex of transcription factor HNF-4α with DNA was obtained from the Protein Data Bank (PDB) using the entry code 3CBB. For finding and exploring the DNA sequence and its location on the human genome, the ENSEMBL genome browser [28] and its tool BLAST [28] were employed. The ENSEMBL additionally facilitated the uncovering of the location of rSNP in the *HNF-1α* gene promotor -58 (upstream). Even though in the ENSEMBL and dbSNP [29] our studied rSNP labeled rs35126805 has no reported (severe) clinical significance, it evidently causes MODY3 as a consequence of disruption of the HNF4-α transcription factor binding site, thereby effecting the *HNF-1α* gene transcription [5–7].

A part of the DNA HNF-4α complex with PDB entry code 3CBB represents Zn ions bound to four cysteine amino acid residues, andthus opening a new challenge on how to represent the coordinative bonds, meaning how to distribute the partial charges between the Zn ion and the cysteines (Cys) for subsequent molecular dynamics simulations. The idea of including the parameters for Zn ions in simulations by Marco and coworkers [18] seemed to be a relatively accurate solution for representing the coordinative bonds in our complex. By using the parameters found in the recent versions of AMBER force field parameter files, we complemented the parameter files and created new definitions for Zn ions and their coordinating Cys residues in the libraries of the Molecular Dynamics Package Q version 5.0 [30]. The parameters for Zn ions were obtained from the AMBER force field parameter file amber99.par designed by Hoops, Anderson, and Merz [31]. The AMBER force field parameters for the sulfur ions of cysteine residues coordinatively bound to zinc ions were developed by Peters and coworkers [32].

In the Crystallographic Object-Oriented Toolkit (Coot) program [33], we induced a mutation on the wild-type adenine–thymine base pair and replaced it with the rSNP cytosine–guanine base pair. To prevent protein unfolding, acetyl groups and methionine groups were added to C-terminal and N-terminal amino-acid residues, respectively, in the PyMOL Molecular Graphics System [34]. Certain amino-acid residues and ultimate DNA base pairs (3′ and 5′) were renamed to match the defined ones in the Q libraries.

Using the AMBER force field implemented in the Molecular Dynamics Package Q [30], the configurational ensembles for the evaluation of free energies were generated from molecular dynamics trajectories. In simulations, our solute molecules were immersed in a sphere (30 Å radius) of TIP3P water molecules subjected to the surface-constraint, all-atom solvent (SCAAS)-type boundary conditions [35], as implemented in the Molecular Dynamics Package Q [30]. The imposed constraints are designed to mimic an infinite aqueous solution. Therefore, stable and realistic free energies are produced even in the small spherical models (water droplets) compared to large simulation boxes utilizing periodic boundary conditions [36,37]. Testing the proper water sphere size between 23 and 30 Å, as well as finding the best position for its center using short molecular dynamics simulations with the Qprep program of Molecular Dynamics Package Q [30] on the CROW computer cluster located on The National Institute of Chemistry in Ljubljana, has revealed poor stability of the dangling DNA end. Therefore, we decided to elongate the dangling part of the DNA. Its sequence was obtained from the ENSEMBL genome browser [28], and the x3DNA program v.2.3 [38] was used for its build-up.

Several translations and rotations in the PyMOL Molecular Graphics System [34] were needed to get a proper fitting of the newly built DNA segment to the prepared complex. The newly built DNA and the prepared complex were then joined. To optimize the contacts between the original and additionally built DNA, a short molecular dynamics simulation with the Qprep program (Q-Amber 95 force field) was performed. The charges of ionized amino-acid residues and DNA base pairs close to the solvent-sphere boundary were reduced. [39] Finally, sodium ions were added to neutralize the system. However, just neutralizing the system might lead to certain artefacts, as the dynamics of DNA strongly depend on the ion concentrations [40].

The reference double-stranded DNA systems were obtained by removing the HNF-4α portion of the complex and by adjusting the number of sodium counterions.

2.2. Molecular Dynamics Simulations

For a proper analysis of the influence of rSNP in DNA on the HNF-4α binding, we prepared four separate systems for subsequent molecular dynamic simulations. The first system was the elongated wild-type DNA, the second was the elongated mutated DNA with rSNP. The third and fourth systems were DNA-HNF-4α complex with wild-type and rSNP base pair, respectively, and of course with all the preparations described in the previous paragraph. Preparatory simulations in the Qprep program with 30 Å radius water solvation sphere for obtaining topology files were performed on all four systems and bond lengths, bond angles, and torsions as well as improper angles were checked. As sodium counterions were placed randomly in the water sphere, replacing some of the solvent (water) molecules, we had to change some of their positions as they were too close to the sphere's outside boundary or too close to the DNA or the protein part of the complex. Next, we performed four series of 2 ns long equilibrating MD simulations with the Qdyn program on all four systems to obtain the initial velocities and to slowly heat-up the system to the production simulation temperature of 298.15 K as well as to loosen heavy-atom restraints that prevent the explosion of the system during equilibration. On the equilibrated systems, sixteen 5 ns long unrestrained production runs were performed using the Qdyn program. A 2 fs step size was used along with the shake algorithm for bonds involving hydrogen atoms [41].

DNA and protein atoms protruding beyond the sphere boundaries were restrained to their initial coordinates using harmonic restraints. Nonbonding interactions of these atoms were turned off. Nonbonding interactions between the atoms inside the simulation sphere were subjected to a 10 Å cut-off. The local-reaction field method was used to treat long-range electrostatic interactions for distances beyond this cut-off. [36,42] Energies were sampled every 10 steps.

2.3. Analysis

We aimed to uncover the influence of rSNP on the binding of transcription factor HNF-4α to the DNA in the *HNF-1α* gene promotor. The analysis was performed in the Visual Molecular Dynamics (VMD) program [43]. The atom-positional root-mean-square difference (RMSD) was calculated for all systems yielding structural differences between conformations.

2.3.1. Analysis of Intermolecular Interactions

Hydrogen bonds exert a great influence on the binding, especially when dealing with nucleic acids and proteins solvated in the water where they represent intra- and intermolecular interactions. Therefore, we examined their quantity, frequency of occurrence, and duration. The water-mediated hydrogen bonds between the protein and the DNA in the 10 Å radius around the rSNP were also considered. The hydrogen-bond existence criteria were strictly geometrical: The donor–acceptor distance was shorter than 3.0 Å and hydrogen–donor–acceptor angle was smaller than 20°.

2.3.2. Thermodynamic Analysis

For the analysis of the binding affinity, the Linear Interaction Energy (LIE) method was used, where the binding free energy is calculated based on the average interaction energies. The LIE method involves molecular dynamics simulations of only two physical states—the bonded and the free ligand. The main idea is to separately treat the contributions of the electrostatic and van der Waals interactions to the total binding affinity. According to the LIE method, the binding free energy ΔG_{bind} is calculated using the following equation:

$$\Delta G_{bind} = \alpha \times \left(\left\langle E_{vdW}^{l-s} \right\rangle_{bound} - \left\langle E_{vdW}^{l-s} \right\rangle_{unbound} \right) + \beta \times \left(\left\langle E_{el}^{l-s} \right\rangle_{bound} - \left\langle E_{el}^{l-s} \right\rangle_{unbound} \right) \tag{1}$$

ΔG_{bind} the binding free energy (kcal/mol);

$\left\langle E_{vdW}^{l-s} \right\rangle$ average vdW interaction energies between the ligand and it's surrounding (kcal/mol);

$\left\langle E_{el}^{l-s} \right\rangle$ average electrostatic interaction energies between the ligand and it's surrounding (kcal/mol);

l ligand;

s surrounding environment;

α, β empirical parameters of the LIE method.

The two above-mentioned physical states encompass the free ligand (in our case the DNA in the water solvation sphere) and the bonded ligand (in our case the complex of HNF-4α bound to the DNA in the water solvation sphere).

Series of both the van der Waals and the electrostatic interaction energies were obtained from the energy trajectories with the Qfep program. From the average interaction energies, the binding free energy ΔG_{bind} was calculated using Equation (1). The following preoptimized values were applied for the LIE empirical parameters $\alpha = 0.18$ and $\beta = 0.43$ [44,45].

To obtain the difference between the binding free energies of the wild-type complex and the mutated rSNP rs35126805 containing complex, Equation (2) was derived based on Scheme 1 and Equation (1).

$$\Delta G_{bind}^{AT} + \Delta G_{Complex}^{AT \rightarrow GC} - \Delta G_{bind}^{GC} - \Delta G_{DNA}^{AT \rightarrow GC} = 0$$
$$\Delta\Delta G_{bind} = \Delta G_{bind}^{GC} - \Delta G_{bind}^{AT} = \Delta G_{Complex}^{AT \rightarrow GC} - \Delta G_{DNA}^{AT \rightarrow GC}$$
$$\Delta\Delta G_{bind} = \alpha \times \left(\left\langle E_{vdW}^{GC-s} \right\rangle_{Complex} - \left\langle E_{vdW}^{GC-s} \right\rangle_{DNA} \right) + \beta \times \left(\left\langle E_{el}^{GC-s} \right\rangle_{Complex} - \left\langle E_{el}^{GC-s} \right\rangle_{DNA} \right)$$
$$- \alpha \times \left(\left\langle E_{vdW}^{AT-s} \right\rangle_{Complex} - \left\langle E_{vdW}^{AT-s} \right\rangle_{DNA} \right)$$
$$- \beta \times \left(\left\langle E_{el}^{AT-s} \right\rangle_{Complex} - \left\langle E_{el}^{AT-s} \right\rangle_{DNA} \right) \tag{2}$$

ΔG_{bind} the binding free energy (kcal/mol);

$\Delta\Delta G_{bind}$ the difference between the binding free energy of the wild-type complex and the mutated rSNP rs35126805 containing complex (kcal/mol);

$\left\langle E_{vdW}^{l-s} \right\rangle$ average van der Waals interaction energies between the ligand and it's surrounding (kcal/mol);

$\left\langle E_{el}^{l-s} \right\rangle$ average electrostatic interaction energies between the ligand and it's surrounding (kcal/mol);

l ligand: AT—wild-type base pair, GC—mutated rSNP base pair; Complex—complex of HNF-4α transcription factor bound to the DNA, DNA—solely the DNA double helix;

s surrounding environment included in the water solvent sphere;

α, β empirical parameters of the LIE method.

The Linear Interaction Energy (LIE) method addresses the entropic terms implicitly by parametrizing the values of its empirical parameters α and β on a large set of experimental binding free energies and by performing molecular dynamics simulations on both end-states. It was observed before that the absolute binding free energies of LIE predictions are not particularly reliable, but that

the relative ($\Delta\Delta G_{bind}$) ranking of binding free energies can still be useful, as it usually correlates with the relative rankings in experiments [46,47].
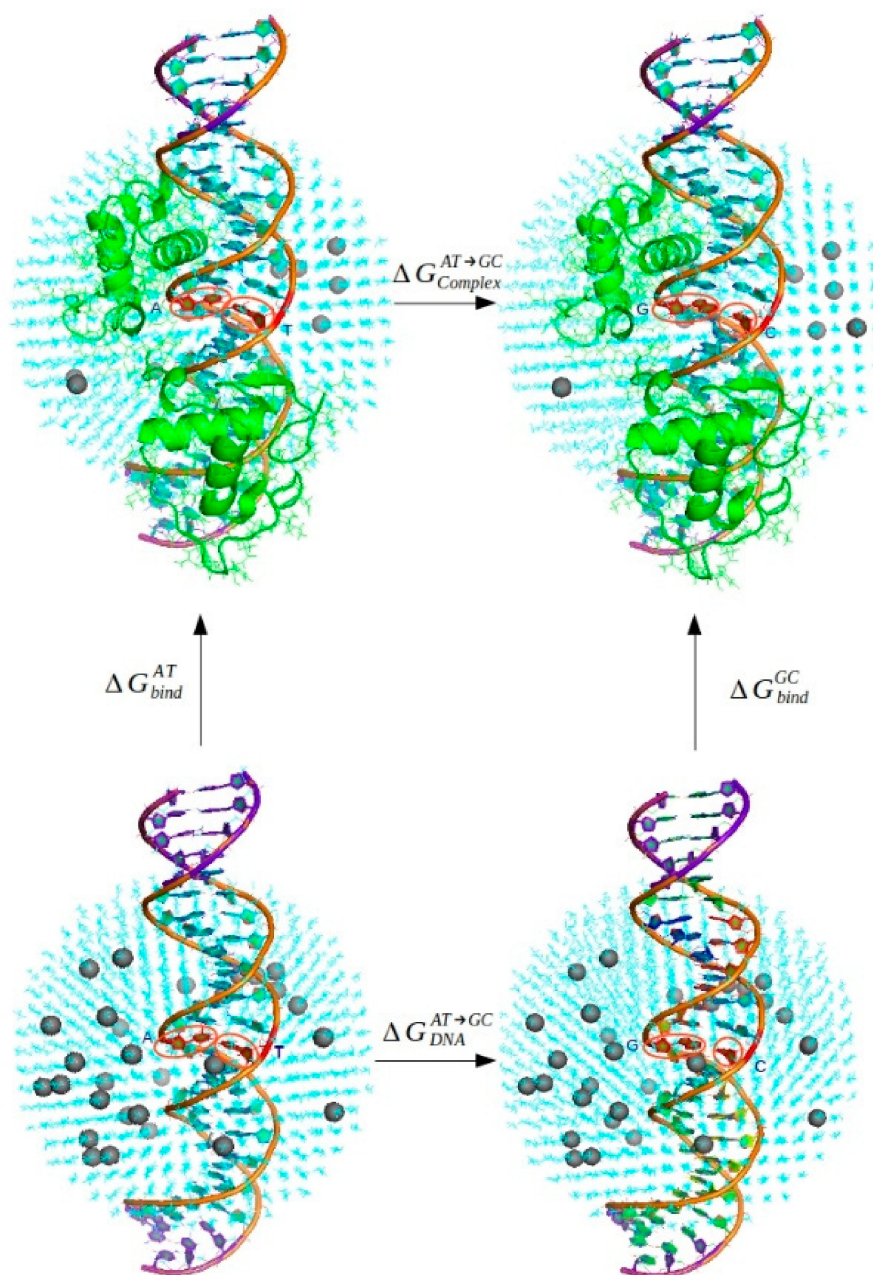
Scheme 1. The four investigated systems in solvated water spheres prepared for the subsequent molecular dynamics simulations: Upper left—wild-type complex; upper right—mutated rSNP rs35126805 containing complex; lower left—wild-type DNA, mutated rSNP rs35126805 containing DNA.

3. Results and Discussion

As our goal was to uncover the influence of rSNP on the binding of transcription factor (TF) HNF-4α to the *HNF-1α* gene promotor DNA sequence, we focused on elucidating the differences between HNF-4α bound to the DNA containing SNP and to the DNA containing the wild-type base pair.

First, we visually followed the structures of both complexes throughout the molecular dynamics simulation and tried to locate the differences in their conformations. Comparison between the structures

of both complexes is shown in Figure 1, where the (mutated) nucleobases represent the center of structures belonging to the ultimate frames of all four molecular dynamics simulation production runs.
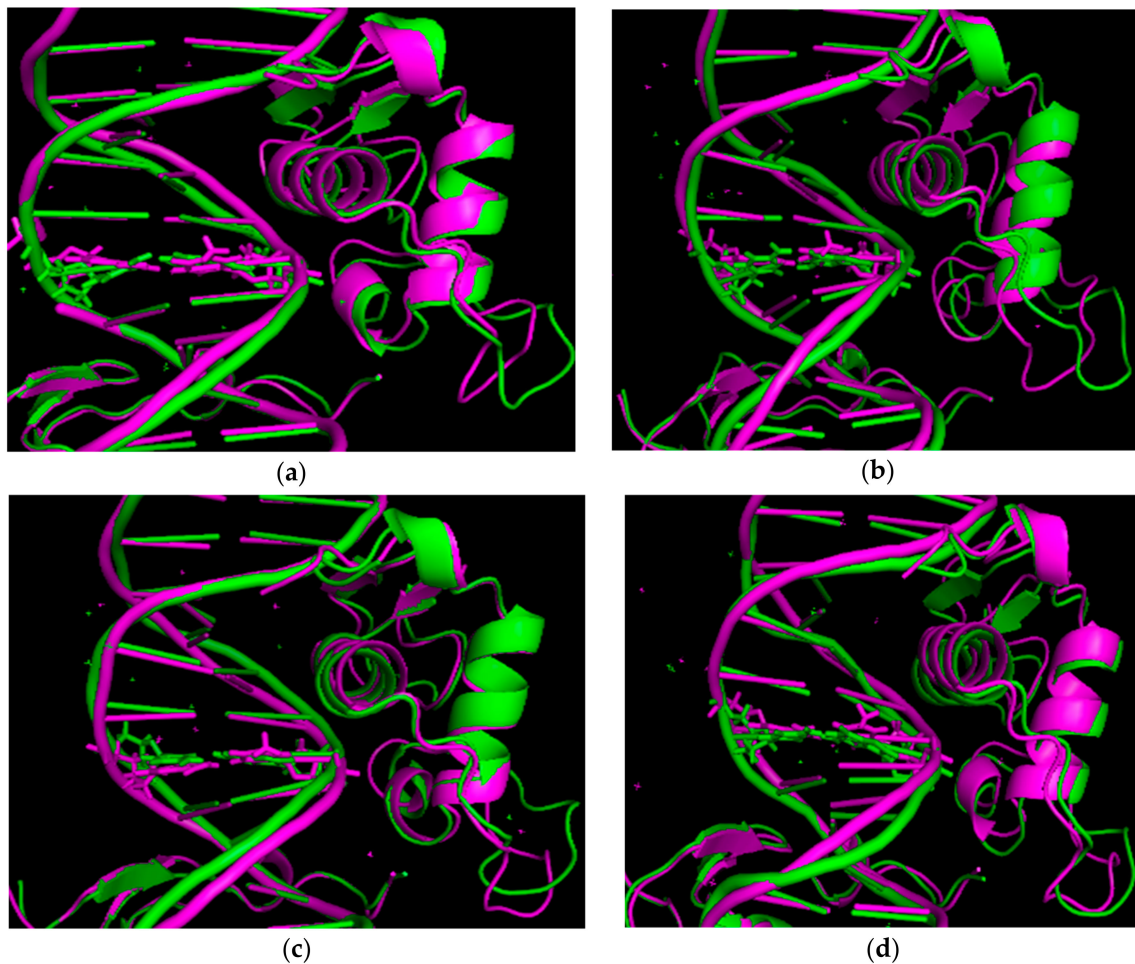

Figure 1. The ultimate frames of the molecular dynamics simulation of HNF-4α DNA complexes: (**a**) Production run 1; (**b**) production run 2; (**c**) production run 3; and (**d**) production run 4. The wild-type complex is depicted in violet, and the complex containing the rSNP is depicted in green. The base pair that gets mutated is shown in sticks, while the remaining base pairs are represented as a ladder.

Visual inspection of the conformations of both complexes did not reveal any significant differences, as the complexes differ so little and the mutated nucleobases are mostly too far away from the protein part of the complex (the transcription factor HNF-4α) to form any stable interactions.

For the subsequent thermodynamic analysis four systems were needed: Two systems of sole DNA chains (*HNF-1α* gene promotor DNA sequences), one containing the rSNP and the other with the wild-type base pair, and two systems of complexes where the transcription factor HNF-4α was bound to these DNA chains.

Having these four separate systems additionally enabled us to follow the impact of the rSNP on the behavior of sole DNA chains, not only of complexes, which was analyzed through atom-positional root-mean-square deviation (RMSD).

3.1. The Results of RMSD Analysis

Figure 2 shows histograms of RMSD of atomic positions throughout 16 independent 5 *ns* molecular dynamics simulation production runs. Under a) the wild-type DNA HNF-1α gene promotor sequence and the DNA containing the rSNP rs35126805 are presented. Under b) wild-type DNA *HNF-1α*

gene promotor sequence in complex with transcription factor HNF-4α and its mutated counterpart containing the rSNP rs35126805 are presented.
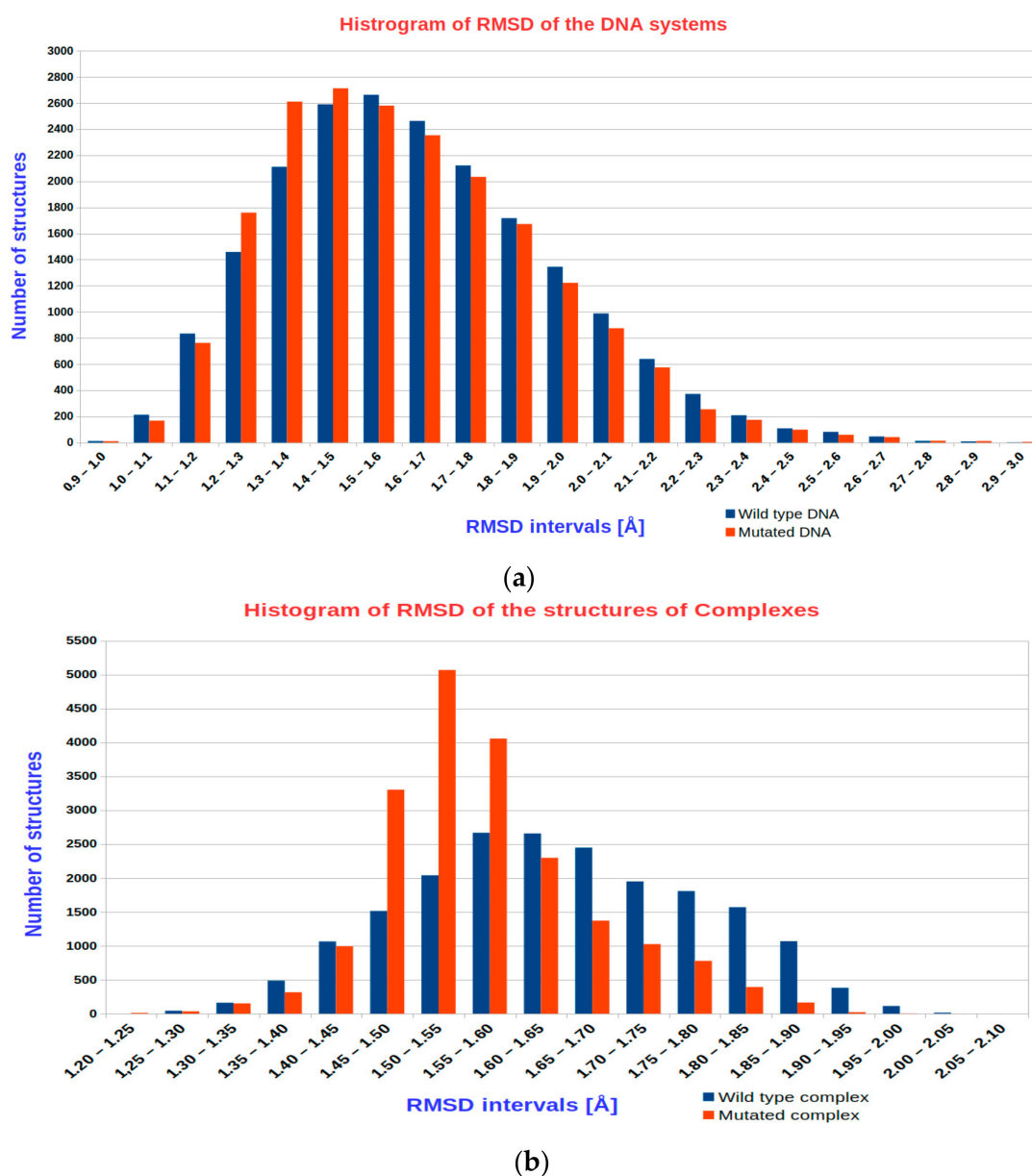

(**a**)

(**b**)

Figure 2. Histograms of root-mean-square difference (RMSD) of atomic positions throughout 16 independent 5 *ns* molecular dynamics simulation production runs of the four investigated systems: (**a**) The wild-type DNA chain containing the *HNF-1α* gene promotor sequence in blue, and the mutated DNA chain containing the *HNF-1α* gene promotor sequence with the rSNP rs35126805 in red color; (**b**) the complex of transcription factor HNF-4α bound to the wild-type DNA *HNF-1α* gene promotor sequence in blue color, and the complex of transcription factor HNF-4α bound to the mutated DNA *HNF-1α* gene promotor sequence with the rSNP rs35126805 in red color.

From Figure 2 and from Supplementary Materials Figure S1, we can observe that in both the mutated DNA containing rSNP rs35126805 and in its complex with transcription factor HNF-4α, the RMSD reaches lower values and fluctuates less, meaning that the structures are more rigid, and consequentially more stable, than their wild-type counterparts. Therefore, the rSNP rs35126805 provides better structural stability to both the DNA and its complex with HNF-4α.

3.2. Analysis of Intermolecular Interactions—Hydrogen Bonds

Hydrogen bonds exert a great influence on the binding, especially when dealing with nucleic acids and proteins solvated in water, where they represent vital intra- and intermolecular interactions. Therefore, we examined their quantity, frequency of occurrence, and duration.

At first, the DNA HNF-4α complex with the wild-type base pair adenine–thymine naturally contains fewer hydrogen bonds than the one with mutated rSNP cytosine–guanine base pair as the later forms three hydrogen bonds and the former only two. Moreover, the analysis confirmed that throughout the molecular dynamics simulation the complex with the mutated DNA contains more hydrogen bonds than the wild-type complex thereby explaining its superior structural stability to at least a certain extent. However, as the water molecules in the surrounding solvation sphere are constantly moving and changing their identity, the resulting transient water bridges must exert a certain structure-stabilizing influence as well. [48]

3.3. The Results of the Thermodynamical Analysis

The difference in the binding affinity between the wild-type HNF-4α DNA complex and the one with the rSNP rs35126805 mutation was obtained using thermodynamic analysis of the energy trajectory gained during the molecular dynamics simulations of the four investigated systems. As the Linear Interaction Energy (LIE) method was applied the corresponding free energy difference was calculated based on the average electrostatic and van der Waals interaction energies. See Section 2.3.2. of the Methods section and particularly Equations (1) and (2) for more information. In Tables S1–S4 (see Supplementary Materials), the ΔG_{bind} for all four investigated systems in all the production runs are presented.

In Table 1 the $\Delta\Delta G_{bind}$ along with its electrostatic and van der Waals contribution is calculated for all the production runs as a difference between the complex containing rSNP rs35126805 and the complex with the wild-type DNA sequence. However, as pointed out by Smith and van Gunsteren, contributions to free energies are only strictly valid if there is no coupling between them [49].

Table 1. The free energy differences $\Delta\Delta G_{bind}$ between the complex containing rSNP rs35126805 and the wild-type DNA sequence for all production runs along with their electrostatic and van der Waals components.

Production Run	$\Delta\Delta G_{bind}^{vdW}$ [1][kcal/mol]	$\Delta\Delta G_{bind}^{es}$ [2] [kcal/mol]	$\Delta\Delta G_{bind}$ [kcal/mol]
Production run 1	−0.20	−0.44	−0.64
Production run 2	0.14	−0.92	−0.78
Production run 3	−0.31	−0.95	−1.30
Production run 4	−0.09	−0.35	−0.44
Average	−0.1 ± 0.1	−0.7 ± 0.3	−0.8 ± 0.3

[1] The van der Waals component of $\Delta\Delta G_{bind}$ between the complex containing rSNP rs35126805 and the wild-type DNA sequence. [2] The electrostatic component of $\Delta\Delta G_{bind}$ between the complex containing rSNP rs35126805 and the wild-type DNA sequence.

As we can see from Table 1, the complex containing the rSNP rs35126805 exhibits a lower binding free energy than the wild-type complex, and the difference between them is approximately—0.8 kcal/mol. The negative value means that the complex containing the rSNP is more stable and therefore HNF-4α binds more strongly to the mutated DNA. This difference stems primarily from electrostatic interactions with shape complementarity through van der Waals interactions playing only a minor role. Consequently, the investigated mutation disrupts the metabolic pathway HNF-4α/HNF-1α, causing MODY3 and potentially hepatocellular carcinoma metastasis as well as enhanced apoptosis in pancreatic cancer cells.

The HNF-4α transcription factor forms a homodimer with its plane of symmetry lying one base pair away from the studied rSNP. The studied rSNP rests in the HNF-4α binding site and probably

forms the base pair closest to the protein itself, therefore, a measurable effect on the binding affinity can be anticipated. If some other A-T base pair of the HNF-1α regulatory domain would be replaced by a C-G base pair, a smaller binding free energy difference $\Delta\Delta G_{bind}$ could be expected.

4. Conclusions

To the best of our knowledge, this represents the first study where thermodynamic analyses of the results obtained from molecular dynamics simulations were performed to uncover the influence of rSNP on the protein binding to DNA.

Even though the non-coding part of the genome contains sequences for the regulation of gene expression and according to GWAS germline SNPs located in introns are also associated with altered cancer risks, this research field remains poorly addressed [1–3]. Consequentially, only a few available databases on regulatory SNPs (rSNPs) exist. Moreover, structures of complexes between human transcription factors and DNA containing rSNPs are exceedingly rare.

The aim of our study is to elucidate the impact of the MODY3 disease-linked rSNP rs35126805 located in the *HNF-1α* gene promotor on the binding of the transcription factor HNF-4α and consequently on the regulation of *HNF-1α* gene expression. Our focus was to investigate the change in the binding affinity of the transcription factor HNF-4α to the DNA, caused by the rSNP through molecular dynamics simulations and thermodynamic analyses of the acquired results. Both RMSD of atomic positions and the relative difference in the binding free energy $\Delta\Delta G_{bind}$ revealed that the HNF-4α binds slightly more strongly to the DNA containing the mutation (rSNP) making the complex more stable/rigid, thereby influencing the expression of the *HNF-1α* gene. This difference mainly originates from electrostatic interactions.

The investigated mutation rSNP rs35126805 disrupts the metabolic pathway HNF-4α/HNF-1α by changing the electrostatic interactions between the HNF-4α transcription factor and the DNA causing MODY3 and potentially also hepatocellular carcinoma metastasis, as well as enhanced apoptosis in pancreatic cancer cells.

As so many questions on the location and biological roles of regulatory SNPs (rSNPs) in disease (cancer) development, progression, and response to therapy remain unanswered, they ought to be investigated in future studies, where novel bsc1 or OL corrections for nucleic acid force fields could be tested out [50–52]. Our computational approach was proven generally useful for studying the impact of regulatory single nucleotide polymorphisms on the binding of transcription factors to the DNA.

Abbreviations

Coot	Crystallographic Object-Oriented Toolkit
Cys	cysteine
GWAS	genome-wide association studies
HNF	hepatocyte nuclear factor
HNF-1α	hepatocyte nuclear factor 1 alpha
HNF-4α	hepatocyte nuclear factor 4 alpha
LIE	Linear Interaction Energy method
MODY3	Maturity-onset diabetes of the young 3
MD	molecular dynamics
PDB	Protein Data Bank
RMSD	root-mean-square deviation
rSNP	regulatory single nucleotide polymorphism
SNP	single nucleotide polymorphism

References

1. Fagny, M.; Platig, J.; Kuijjer, M.L.; Xihong Lin, X.; Quackenbush, J. Nongenic cancer-risk SNPs affect oncogenes, tumour-suppressor genes, and immune function. *Br. J. Cancer* **2020**, *122*, 569–577. [CrossRef] [PubMed]

2. Kalia, N.; Sharma, A.; Kaur, M.; Kamboj, S.S.; Singh, J. A comprehensive in silico analysis of non-synonymous and regulatory SNPs of human MBL2 gene. *Springerplus* **2016**, *5*, 811. [CrossRef] [PubMed]

3. Zhang, Z.; Liu, M.; Li, B.; Wang, Y.; Yue, J.; Liang, L.; Sun, J. Exploring the mechanism of regulatory SNP of KLK3 by molecular dynamics simulation. *J. Biomol. Struct. Dyn.* **2013**, *31*, 426–440. [CrossRef] [PubMed]

4. Sud, A.; Kinnersley, B.; Houlston, R.S. Genome-wide association studies of cancer: Current insights and future perspectives. *Nat. Rev. Cancer* **2017**, *17*, 692–704. [CrossRef]

5. Gragnoli, C.; Lindner, T.; Cockburn, B.N.; Kaisaki, P.J.; Gragnoli, F.; Marozzi, G.; Bell, G.I. Maturity-Onset Diabetes of the Young Due to a Mutation in the Hepatocyte Nuclear Factor-4α Binding Site in the Promoter of the Hepatocyte Nuclear Factor-1α Gene. *Diabetes* **1997**, *46*, 1648–1651. [CrossRef]

6. Pihoker, C.; Gilliam, L.K.; Ellard, S.; Dabelea, D.; Davis, C.; Dolan, L.M.; Greenbaum, C.J.; Imperatore, G.; Lawrence, J.M.; Marcovina, S.M. Prevalence, characteristics and clinical diagnosis of maturity onset diabetes of the young due to mutations in HNF1A, HNF4A, and glucokinase: Results from the SEARCH for Diabetes in Youth. *J. Clin. Endocrinol. Metab.* **2013**, *98*, 4055–4062. Available online: https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/monogenic-neonatal-mellitus-mody (accessed on 13 March 2019).

7. Hattersley, A.T.; Pearson, E.R. Minireview: Pharmacogenetics and beyond: The interaction of therapeutic response, beta-cell physiology, and genetics in diabetes. *Endocrinology* **2006**, *147*, 2657–2663. [CrossRef]

8. Sel, S.; Ebert, T.; Ryffel, G.U.; Drewes, T. Human renal cell carcinogenesis is accompanied by a coordinate loss of the tissue specific transcription factors HNF4 alpha and HNF1 alpha. *Cancer Lett.* **1996**, *101*, 205–210. [CrossRef]

9. Azmi, A.S.; Bao, G.F.; Gao, J.; Mohammad, R.M.; Sarkar, F.H. Network Insights into the Genes Regulated by Hepatocyte Nuclear Factor 4 in Response to Drug induced Perturbations. *Curr. Drug. Discov. Technol.* **2013**, *10*, 147–154. [CrossRef]

10. National Center for Biotechnology Information, U.S. National Library of Medicine. HNF1A. Available online: https://www.ncbi.nlm.nih.gov/gene/6927 (accessed on 14 March 2019).

11. Maestro, M.A.; Cardalda, C.; Boj, S.F.; Luco, R.F.; Servitja, J.M.; Ferrer, J. Distinct roles of HNF1beta, HNF1alpha, and HNF4alpha in regulating pancreas development, beta-cell function and growth. *Endocr. Dev.* **2007**, *12*, 33–45. [CrossRef]

12. Abel, E.V.; Goto, M.; Magnuson, B.; Abraham, S.; Ramanathan, N.; Hotaling, E.; Alaniz, A.A.; Kumar-Sinha, C.; Dziubinski, M.L.; Urs, S.; et al. HNF1A is a novel oncogene that regulates human pancreatic cancer stem cell properties. *Elife* **2018**, *7*, e33947. [CrossRef]

13. Luo, Z.; Li, Y.; Wang, H.; Fleming, J.; Li, M.; Kang, Y.; Zhang, R.; Li, D. Hepatocyte Nuclear Factor 1A (HNF1A) as a Possible Tumor Suppressor in Pancreatic Cancer. *PLoS ONE* **2015**, *10*, e0121082. [CrossRef] [PubMed]

14. National Center for Biotechnology Information, U.S. National Library of Medicine. HNF4A. Available online: https://www.ncbi.nlm.nih.gov/gene/3172 (accessed on 14 March 2019).

15. Hanawa, M.; Takayama, K.; Sakurai, F.; Tachibana, M.; Mizuguchi, H. Hepatocyte Nuclear Factor 4 Alpha Promotes Definitive Endoderm Differentiation from Human Induced Pluripotent Stem Cells. *Stem. Cell. Rev. Rep.* **2017**, *13*, 542–551. [CrossRef]

16. Yan, H.; Wang, Q.; Shen, Q.; Li, Z.; Tian, J.; Jiang, Q.; Gao, L. Identification of potential transcription factors, long noncoding RNAs, and microRNAs associated with hepatocellular carcinoma. *J. Cancer Res. Ther.* **2018**, *14*, S622–S627. [CrossRef]

17. Mura, C.; McCammon, J.A. Molecular dynamics of a κB DNA element: Base flipping via cross-strand intercalative stacking in a microsecond-scale simulation. *Nucleic Acids Res.* **2008**, *36*, 4941–4955. [CrossRef] [PubMed]

18. Marco, E.; García-Nieto, R.; Gago, F. Assessment by Molecular Dynamics Simulations of the Structural Determinants of DNA-binding Specificity for Transcription Factor Sp1. *J. Mol. Biol.* **2003**, *328*, 9–32. [CrossRef]

19. Zhao, L.; Sun, T.; Pei, J.; Ouyang, Q. Mutation-induced protein interaction kinetics changes affect apoptotic network dynamic properties and facilitate oncogenesis. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E4046–E4054. [CrossRef] [PubMed]

20. Ma, S.; Tan, S.; Fang, D.; Zhang, R.; Zhou, S.; Wu, W.; Zheng, K. Probing the binding mechanism of novel dual NF-κB/AP-1 inhibitors by 3D_QSAR, docking and molecular dynamics simulations. *RSC Adv.* **2015**, *5*, 81523–81532. [CrossRef]

21. Klvaňa, M.; Bren, U.; Florián, J. Uniform Free-Energy Profiles of the P–O Bond Formation and Cleavage Reactions Catalyzed by DNA Polymerases β and λ. *J. Phys. Chem. B* **2016**, *120*, 13017–13030. [CrossRef]

22. Martínek, V.; Bren, U.; Goodman, M.F.; Warshel, A.; Florián, J. DNA polymerase β catalytic efficiency mirrors the Asn279–dCTP H-bonding strength. *Febs. Lett.* **2007**, *581*, 775–780. [CrossRef]

23. Fenollar-Ferrer, C.; Anselmi, C.; Carnevale, V.; Raugei, S.; Carloni, P. Insights on the acetylated NF-kB transcription factor complex with DNA from molecular dynamics simulations. *Proteins* **2012**, *80*, 1560–1568. [CrossRef]

24. Dutta, S.; Agrawal, Y.; Mishra, A.; Kaur Dhanjal, J.; Sundar, D. A theoretical investigation of DNA dynamics and desolvation kinetics for zinc finger proteinZif268. *BMC Genom.* **2015**, *16* (Suppl. 12), S5. [CrossRef] [PubMed]

25. Bouard, C.; Terreux, R.; Honorat, M.; Manship, B.; Ansieau, S.; Vigneron, A.M.; Puisieux, A.; Payen, L. Deciphering the molecular mechanisms underlying the binding of the TWIST1/E12 complex to regulatory E-box sequences. *Nucleic. Acids. Res.* **2016**, *44*, 5470–5489. [CrossRef] [PubMed]

26. MacKerell, A.D.; Nilsson, L. Molecular dynamics simulations of nucleic acid–protein complexes. *Curr. Opin. Struct. Biol.* **2008**, *18*, 194–199. [CrossRef] [PubMed]

27. George Priya Doss, C.; Nagasundaram, N.; Chakraborty, C.; Chen, L.; Zhu, H. Extrapolating the effect of deleterious nsSNPs in the binding adaptability of flavopiridol with CDK7 protein: A molecular dynamics approach. *Hum. Genom.* **2013**, *7*, 10. [CrossRef] [PubMed]

28. Aken, B.L.; Ayling, S.; Barrell, D.; Clarke, L.; Curwen, V.; Fairley, S.; Banet, J.F.; Billis, K.; Girón, C.G.; Hourlier, T.; et al. The Ensembl gene annotation system. *Database* **2016**, *2016*, baw093. Available online: https://www.ensembl.org/index.html (accessed on 15 July 2016). [CrossRef]

29. National Center for Biotechnology Information, U.S. National Library of Medicine. dbSNP. Available online: https://www.ncbi.nlm.nih.gov/SNP/ (accessed on 13 March 2019).

30. Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J.Q. A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graph. Model.* **1998**, *16*, 213–225. Available online: http://xray.bmc.uu.se/aqwww/Q (accessed on 1 January 2004). [CrossRef]

31. Hoops, S.C.; Anderson, K.W.; Merz, K.M., Jr. Force Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262–8270. [CrossRef]

32.   Peters, M.B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M.N.; Merz, K.M., Jr. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6*, 2935–2947. [CrossRef]

33.   Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D* **2010**, *66*, 486–501. Available online: https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/ (accessed on 18 December 2020). [CrossRef]

34.   The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Available online: https://pymol.org/2/ (accessed on 18 December 2020).

35.   King, G.; Warshel, A. A surface constrained all-atom solvent model for effective simulations of polar solutions. *J. Chem. Phys.* **1989**, *91*, 3647. [CrossRef]

36.   Sham, Y.Y.; Warshel, A. The surface constraint all atom model provides size independent results in calculations of hydration free energies. *J. Chem. Phys.* **1998**, *109*, 7940. [CrossRef]

37.   Bren, U.; Lah, J.; Bren, M.; Martínek, V.; Jan Florián, J. DNA Duplex Stability: The Role of Preorganized Electrostatics. *J. Phys. Chem. B* **2010**, *114*, 2876–2885. [CrossRef] [PubMed]

38.   Lu, X.-J.; Olson, W.K. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003**, *31*, 5108–5121, The 3DNA. Available online: https://x3dna.org/ (accessed on 18 December 2020). [CrossRef] [PubMed]

39.   Florián, J.; Goodman, M.F.; Warshel, A. Free-Energy Perturbation Calculations of DNA Destabilization by Base Substitutions: The Effect of Neutral Guanine·Thymine, Adenine·Cytosine and Adenine·Difluorotoluene Mismatches. *J. Phys. Chem. B* **2000**, *104*, 10092–10099. [CrossRef]

40.   Cheatham, T.E., III; Kollman, P.A. Molecular Dynamics Simulation of Nucleic Acids. *Annu. Rev. Physchem.* **2000**, *51*, 435–471. [CrossRef]

41.   Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comp. Phys.* **1977**, *23*, 327–341. [CrossRef]

42.   Lee, F.S.; Warshel, A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.* **1992**, *97*, 3100. [CrossRef]

43.   Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. Available online: http://www.ks.uiuc.edu/Research/vmd/ (accessed on 18 December 2020). [CrossRef]

44.   Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des.* **1998**, *12*, 27–35. [CrossRef]

45.   Klvaňa, M.; Bren, U. Aflatoxin B1-Formamidopyrimidine DNA adducts: Relationships between structures, free energies, and melting temperatures. *Molecules* **2019**, *24*, 150. [CrossRef]

46.   Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding affinities in the SAMPL3 trypsin and host–guest blind tests estimated with the MM/PBSA and LIE methods. *J. Comput. Aided Mol. Des.* **2012**, *26*, 527–541. [CrossRef] [PubMed]

47.   König, G.; Brooks, B.R. Predicting binding affinities of host-guest systems in the SAMPL3 blind challenge: The performance of relative free energy calculations. *J. Comput. Aided Mol. Des.* **2012**, *26*, 543–550. [CrossRef] [PubMed]

48.   Chan-Yao-Chong, M.; Marsin, S.; Quevillon-Cheruel, S.; Durand, D.; Ha-Duong, T. Structural ensemble and biological activity of DciA intrinsically disordered region. *J. Struct. Biol.* **2020**, *212*, 107573. [CrossRef] [PubMed]

49.   Smith, P.E.; van Gunsteren, W.F. When Are Free Energy Components Meaningful? *J. Phys. Chem.* **1994**, *98*, 13735–13740. [CrossRef]

50.   Galindo-Murillo, R.; Robertson, J.C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T.E., III. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127. [CrossRef] [PubMed]

51.   Zgarbová, M.; Luque, F.J.; Šponer, J.; Cheatham III, T.E.; Otyepka, M.; Jurečka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339–2354. [CrossRef] [PubMed]

52. Ivani, I.; Dans, P.D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. PARMBSC1: A Refined Force-Field for DNA Simulations. *Nat. Methods* **2016**, *13*, 55–58. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.