



Article

# A Computer-Based Methodology to Design Non-Standard Peptides Potentially Able to Prevent HOX-PBX1-Associated Cancer Diseases

Maria Rita Gulotta <sup>1,\*</sup>, Giada De Simone <sup>1</sup>, Justin John <sup>2</sup>, Ugo Perricone <sup>1,†</sup> and Andrea Brancale <sup>3,†</sup>

<sup>1</sup> Molecular Informatics Unit, Fondazione Ri.MED, Via Filippo Marini 14, 90128 Palermo, Italy;

gdesimone@fondazionerimed.com (G.D.S.); uperricone@fondazionerimed.com (U.P.)

<sup>2</sup> NRN Tech LTD, Henstaff Court, Llantrisant Road, Groesfaen CF72 8NG, UK; justin@nrntech.co.uk

<sup>3</sup> School of Pharmacy and Pharmaceutical Sciences, Cardiff University, King Edward VII Avenue, Cardiff CF10 3NB, UK; brancalea@cardiff.ac.uk

\* Correspondence: mrgulotta@fondazionerimed.com

† These authors equally contributed to this work.

**Abstract:** In the last decades, HOX proteins have been extensively studied due to their pivotal role in transcriptional events. HOX proteins execute their activity by exploiting a cooperative binding to PBX proteins and DNA. Therefore, an increase or decrease in HOX activity has been associated with both solid and haematological cancer diseases. Thus, inhibiting HOX-PBX interaction represents a potential strategy to prevent these malignancies, as demonstrated by the patented peptide HTL001 that is being studied in clinical trials. In this work, a computational study is described to identify novel potential peptides designed by employing a database of non-natural amino acids. For this purpose, residue scanning of the HOX minimal active sequence was performed to select the mutations to be further processed. According to these results, the peptides were point-mutated and used for Molecular Dynamics (MD) simulations in complex with PBX1 protein and DNA to evaluate complex binding stability. MM-GBSA calculations of the resulting MD trajectories were exploited to guide the selection of the most promising mutations that were exploited to generate twelve combinatorial peptides. Finally, the latter peptides in complex with PBX1 protein and DNA were exploited to run MD simulations and the  $\Delta G_{\text{binding}}$  average values of the complexes were calculated. Thus, the analysis of the results highlighted eleven combinatorial peptides that will be considered for further assays.

**Keywords:** HOX; PBX; Protein-Protein Interactions; Residue Scanning; Molecular Dynamics; MM-GBSA; Cancer; Non-standard amino acids



**Citation:** Gulotta, M.R.; De Simone, G.; John, J.; Perricone, U.; Brancale, A. A Computer-Based Methodology to Design Non-Standard Peptides Potentially Able to Prevent HOX-PBX1-Associated Cancer Diseases. *Int. J. Mol. Sci.* **2021**, *22*, 5670. <https://doi.org/10.3390/ijms22115670>

Academic Editor:

Alessandra Magistrato

Received: 19 April 2021

Accepted: 24 May 2021

Published: 26 May 2021

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



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Human phenotype development, evolution and physiopathological processes are regulated by several key actors. Among these, *HOX* genes have been associated with the control of the final morphology [1,2]. Their increase or decrease in activity can often result in homeotic transformations provoking the formation of structures or organs in erroneous locations within the organism. Three different levels of *HOX* genes evolutionary conservation have been identified: (1) at a molecular level, they all encode homeodomain transcription factors [3]; (2) at a structural level, *HOX* genes are usually organised in complexes, reflecting their phylogeny and regulatory aspects of their expression [4,5], and (3) at a functional level, they trigger similar effects in most animals and can work in substitution of an orthologue in other species [6].

When becoming highly dysregulated and overexpressed, *HOX* genes have been associated with a wide range of both solid and haematological cancers [7].

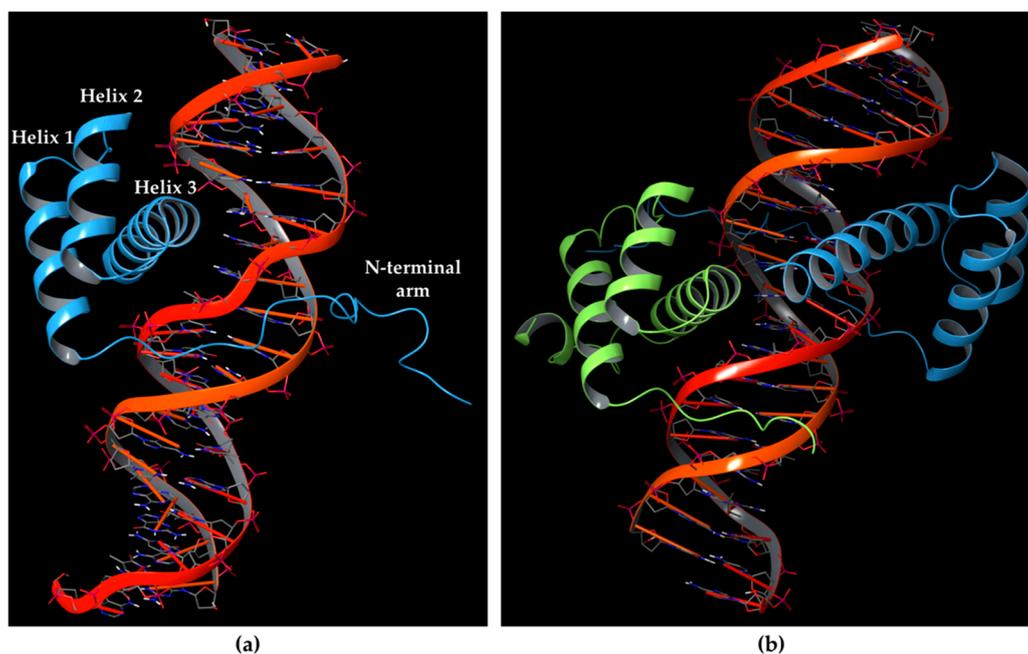
For this reason, the processes modulated by *HOX* genes have been extensively studied providing a substantial, although not exhaustive, analysis of them. Recent studies high-

lighted that *HOX* genes also contribute to organogenesis [8] by influencing a huge number of cellular functions such as differentiation, proliferation, migration or death [9].

*HOX* proteins consist of two highly conserved portions: the hexapeptide (HX) motif and the homeodomain (HD). The HX motif establishes interactions with protein members of the PBC class, such as Pre-B-cell Leukemia Homeobox (PBX) proteins in humans [10]; while the HD motif corresponds to the DNA-binding domain. The HD folds into a triple-helix structure, including the N-terminal arm binding the minor groove of DNA, and helix 3 (also named the recognition helix) contacting the DNA in the major groove, as depicted in Figure 1a. Residues involved in *HOX* HD helices 1 and 3 have shown to be the most conserved, and other amino acids of the N-terminal arm and loops between the helices have been reported well conserved among *HOX* proteins. Furthermore, the conservation of HD sequences is highly shared between *HOX* proteins, raising the issue of how they employ functional specificity [11–13].

Indeed, the homeodomain of *HOX* proteins does not exhibit high specificity for DNA, by taking part in the molecular recognition through five amino acids [14]. In this context, functional studies in the field of cancer and developmental biology highlighted the role of PBX proteins as *HOX* co-factors [15–17], whereas PBX family members bind to *HOX* proteins 1–11 [18–20]. These proteins may establish a cooperative binding to DNA [21,22] as depicted in Figure 1b, indicating that the interaction of *HOX* proteins to PBX influences the DNA-binding of *HOX* by fostering a greater specificity [23]. Furthermore, the *HOX* co-factors play other key roles influencing transcriptional events, by recruiting the RNA polymerase II and III or transcriptional inhibitors like HDAC, and post-translational events, by fostering the entry of *HOX* proteins into the nucleus.

Four different types of *PBX* genes (*PBX1–4*) are encoded in the human genome. As for *HOX* proteins, *PBX* genes also encode evolutionarily conserved homeodomains and other highly conserved regions [16]. *PBX* proteins present two nuclear localization signals (NLSs) in the homeodomain and a nuclear export sequence (NES) [24–26].



**Figure 1.** (a) Homeodomain of HOXA9 protein (light blue chain) showing a triple-helix structure where helix 3 inserts into the DNA major groove, while the N-terminal arm establishes contacts with DNA minor groove (PDB ID: 1PUF [27]); (b) HOXA9-PBX1 cooperative binding in presence of DNA retrieved from PDB 1PUF [27], where the red double-helix structure is DNA bound to the Homeobox protein HOXA9 (light blue chain) and Pre-B-cell leukaemia transcription factor-1 PBX1 (green chain).

PBX proteins may participate in a DNA binding consensus through the formation of strong complexes with HOX1-11 proteins [21,28,29]. The involved interactions have been shown to exhibit a highly conserved interaction mode between the HX motif of HOX and the three amino acid loop extension (TALE) or three-amino acid insertion peptide of PBX, which is located between helices 1 and 2 of the homeodomain [18,27,29–33].

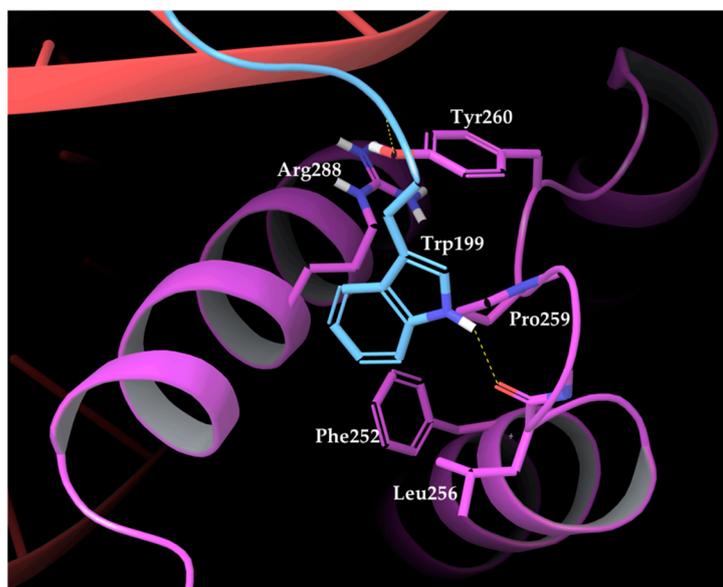
In 1995, Knoepfler and Kamps [14] identified the minimal sequences that enable HOXB8 and HOXA5 proteins to bind cooperatively PBX1 protein, by performing deletion mutagenesis on the above-mentioned HOX proteins. This minimal sequence was the conserved pentapeptide motif *Y/F-P-W-M-R/K*. In particular, mutations at tryptophan residue did not produce binding abrogation. Mutations of Trp135 to phenylalanine (W135F) or alanine (W135A) on HOXB8 did not alter the DNA binding but completely abolished the cooperativity of HOXB8 with PBX1. Met136 was also shown to be important but not essential for the DNA binding of PBX1, whereas its mutation to isoleucine (M136I) or alanine (M136A) strongly disrupted the cooperativity of HOXB8 with PBX1. Therefore, both residues, Trp135 and Met136, were considered crucial for the protein/DNA interaction, with particular attention for Trp135 [32], since it was the only conserved amino acid among all HOX proteins pentapeptide sequences. Knoepfler and Kamps also assumed that the pentapeptide HOX motif stabilizes the trimeric HOX-PBX1-DNA complex by bringing a portion of the HOX protein surface into contact with PBX1 and enhancing DNA binding in the presence of PBX1 [14]. However, the X-ray crystallographic structures of HOXB1-PBX1 and HOXA9-PBX1 in the presence of DNA revealed that the protein-protein-DNA contacts are stabilized by the interaction between HOX and PBX mediated not by a simple pentapeptide sequence, but by a conserved hexapeptide sequence in HOX proteins [3,18,27,31].

In 1999 Piper et al. [18] found that a minimal portion of HOX containing hexapeptide and homeodomain was able to cooperatively stabilize DNA binding with PBX1. Hence, the identified consensus hexapeptide motif from HOX proteins was  $\varphi$ -*Y/F-P-W-M-K/R*, where  $\varphi$  stands for a hydrophobic residue. As reported above, tryptophan and methionine were conserved.

X-ray crystallographic structures of the ternary complex, HOX-PBX1-DNA (e.g., HOXA9 in PDB 1PUF, Resolution: 1.90 Å; and HOXB1 in PDB 1B72, Resolution: 2.35 Å) [27] revealed that HOX protein and PBX1 establish contacts with opposite DNA faces, burying 2400 Å<sup>2</sup> of protein and DNA surface. The HOX hexapeptide mediates contacts with PBX1 within a hydrophobic pocket located between the TALE region and helix 3 of the PBX1 homeodomain.

PCR site-selection experiments performed by Piper et al. [18] allowed us to identify the optimal HOXB1-PBX1 binding site on the 20 bp duplex DNA oligonucleotide, i.e., 5'-ATGATTGATCG-3' [34].

The PDB structure resolved by La Ronde-Le Blanc and Wolberger [27] revealed that the interactions between HOXA9 and PBX1 are mediated by HOXA9 hexapeptide, consisting of residues 196 to 201 with the AANWLH sequence linked to the PBX1 homeodomain. As observable in the PDB structures, the hexapeptide residues mediate mainly hydrophobic contacts, whereas HOXA9 Trp199 side-chain inserts into a hydrophobic pocket of PBX1 formed by the C terminus of helix 3, a handle between helices 3 and 4, and the three-amino acid insertion. The key interactions involve HOXA9 Trp199 with its indole ring that forms van der Waals contacts with several PBX1 residues, such as the Phe252 side chain in helix 1, Leu256 within the TALE region, Pro259 and Tyr260 following the TALE peptide, and Arg288 in helix 3. Furthermore, Trp199 is highly buried into the PBX1 binding pocket by forming a hydrogen bond between the indole nitrogen and the backbone carbonyl of PBX1 Leu256 (Figure 2).



**Figure 2.** Amino acids composing PBX1 binding pocket (purple residues and chain) surrounding HOXA9 Trp199 (light blue chain) [27].

The nitrogen atom at HOXA9 Leu200 backbone establishes van der Waals contacts with Ly292 and a hydrogen bond with Tyr260 hydroxyl group within the binding pocket of PBX1. Finally, His201 of HOXA9 hexapeptide forms a hydrogen bond with Lys292 of PBX1.

Moreover, Piper et al. [18] on PBX1 conducted some mutational studies focused on the hexapeptide-contacting residues Leu252 and Pro259, which were substituted for alanine, triggering the disruption of the interactions with the hexapeptide *in vitro* and in a yeast two-hybrid assay [35]. Furthermore, deletion assays at the three-amino acid insertion abolished the cooperative binding of PBX1 with HOX proteins [36]. On the other hand, the deletion of the HOX hexapeptide caused the disappearance of cooperative interactions between PBX1 and HOX proteins [14,29,37].

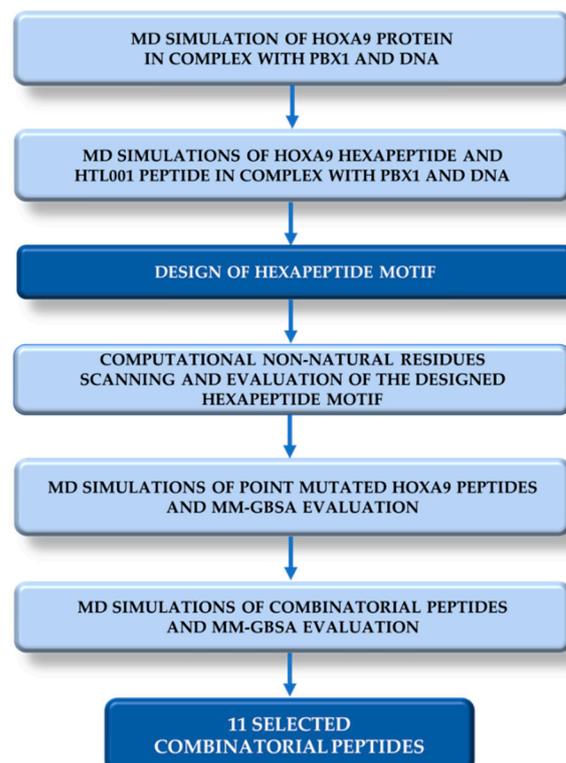
Although 3D structures of the ternary complex HOX-PBX-DNA have been experimentally solved, the drug discovery process for this protein-protein interaction (PPI) has met some issues typical of targeting a PPI through designing potential effective small molecule inhibitors [38,39]. However, an accepted strategy is to target the HOX-PBX binding interface at the highly conserved residues involving HOX hexapeptide and exploiting the hydrophobic nature of the PBX protein binding pocket. In the last decades, a small molecule inhibitor of this interaction was identified. However, its  $K_D$  was in the micromolar range (65  $\mu\text{M}$ ) and it was neglected for further experimental assays or clinical trials [40]. On the other hand, in the last years, several peptides have been designed based on the hexapeptide consensus motif of HOX proteins, to work as a competitive antagonist of HOX-PBX binding [41]. The most promising peptide among these was HXR9, an 18-amino acid peptide containing the hexapeptide sequence together with a polyarginine portion.

HXR9 was first shown to be cytotoxic to melanoma cell lines and primary melanoma cells and registered a reduction of B16F10 murine melanoma tumours growth in an orthotopic model [42]. Other experimental studies reported that HXR9 was able to inhibit the growth of several tumour types in mouse xenograft models, including non-small cell lung [43], breast [44], ovarian [44], and prostate cancer [45], and mesothelioma [46], melanoma [47], and meningioma [48].

Recently, modifications performed on the HXR9 sequence shed light on another peptide, i.e., the HTL001 peptide [49] with the sequence WYPWMKKHHRRRRRRRRR, that was tested in cancer cells representative of 14 human and animal malignancies. HTL001 registered selective toxicity for cancer cells and safety for normal cells. To date, this peptide has reached the human clinical trials that are ongoing to assay the efficacy and safety. However, the mechanism associated with HOX-PBX inhibition and the resulting cell death

through employing the HTL001 peptide is still to be fully elucidated, although generally in most solid tumours cell death is mediated by apoptosis [42,44–46,50].

In this article, we describe a computer-based strategy to design novel peptides including non-standard amino acids to potentially bind PBX1 and inhibit HOX-PBX1 interaction (Figure 3). For this purpose, a Molecular Dynamics (MD) simulation of 200 ns was performed on the ternary complex HOXA9-PBX1-DNA retrieved from the Protein Data Bank [51] to identify key residues and interactions [52]. Two other MD were run, one on HOXA9 hexapeptide (196-AANWLH-201) from PDB 1PUF and the other on the patented core peptide HTL001 without polyarginine coil in complex with PBX1-DNA. The resulting MD trajectories were then used to compute MM-GBSA (Molecular Mechanics–Generalised Born Surface Area) calculations to obtain  $\Delta G_{\text{binding}}$  average values as references for the design of the new peptides. Subsequently, the HOXA9 hexapeptide sequence was submitted to a point mutational scanning exploiting a non-natural amino acid database populated by the Swiss Institute of Bioinformatics [53,54]. The mutants were selected according to  $\Delta\Delta G_{\text{affinity}}$  and  $\Delta\Delta G_{\text{stability}}$  values and, in complex with PBX1-DNA, were further explored by MD simulations and MM-GBSA calculations. All those complexes point-mutated peptide-PBX1-DNA reporting  $\Delta G_{\text{binding}}$  average values lower compared to the reference  $\Delta G_{\text{binding}}$  average values (involving HOXA9 hexapeptide and HTL001 core peptide) were chosen for the next steps of the work. Thus, the most promising mutated peptides were used as starting point to generate twelve combinatorial peptides. The newly generated peptides were reprocessed in MD and MM-GBSA calculations as peptide-PBX1-DNA complex. Finally, eleven peptides showed promising  $\Delta G_{\text{binding}}$  values compared to the HOXA9 hexapeptide and HTL001 peptide.



**Figure 3.** Overview of the computational workflow performed to identify the eleven combinatorial peptides potentially inhibiting HOX-PBX1 cooperative binding.

## 2. Results and Discussion

### 2.1. Molecular Dynamics Simulation of HOXA9-PBX1-DNA Complex

The first step of this work was the analysis and selection of a high-quality PDB structure of the trimeric complex HOXA9-PBX1-DNA currently available in the Protein Data Bank [51] to perform an MD simulation. For this purpose, PDB 1PUF (resolution 1.90 Å) [27] was chosen for a 200 ns MD simulation, to analyse and identify the most stable interactions and key residues for both proteins. The complex stability was also investigated by reporting the RMSD plot illustrated in Figure S1 in Supplementary materials, which showed a stable behaviour of the system.

The trajectory was further analysed to retrieve the most frequent interactions established between the HOX and PBX proteins. MD frames were clustered in 10 groups based on the RMSD matrix by using both protein backbone and sidechains and setting a frequency of 10 steps at which the frames were analysed. The centroid frames for the most abundant clusters were: frame 880 (representative for 63 frames), frame 70 (representative for 34 frames), frame 540 (representative for 28 frames), frame 360 (representative for 22 frames), and frame 270 (representative for 15 frames) [55].

The observed interactions in these frames were analysed and considered as the most stable and frequent during the trajectory. Table 1 lists the key residues involved in contacts between the two proteins, whereas for HOXA9 only residues involved in the PBX-contacting hexapeptide region 196-AANWLH-201 were included in Table 1.

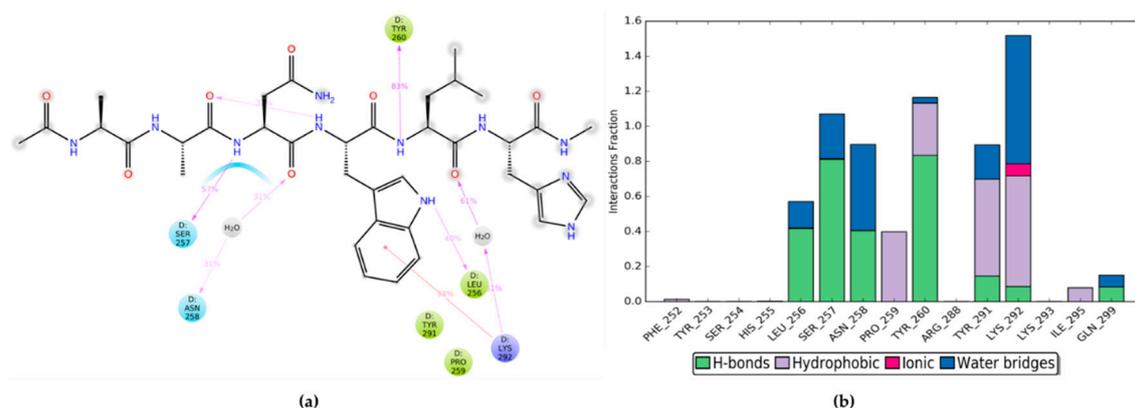
**Table 1.** The most frequent interactions and the related involved residues for HOXA9 hexapeptide 196-AANWLH-201 and PBX1 homeodomain proteins retrieved from MD frames clustering.

| HOXA9 Residue | PBX1 Residue | Interaction Type |
|---------------|--------------|------------------|
| Trp199        | Ser257       | 1 H-bond         |
| Trp199        | Leu256       | 1 H-bond         |
| Trp199        | Tyr291       | $\pi$ -stacking  |
| Trp199        | Tyr260       | $\pi$ -stacking  |
| Leu200        | Tyr260       | 1 H-bond         |
| Ala197        | Asn258       | 1 H-bond         |

As it can be observed, Trp199 was detected as the key residue showing the majority of the interactions with PBX1 residues. This result was in agreement with information reported in the literature highlighting this tryptophan amino acid [14] as the fundamental one and was considered for the design of novel peptides described in the next sections.

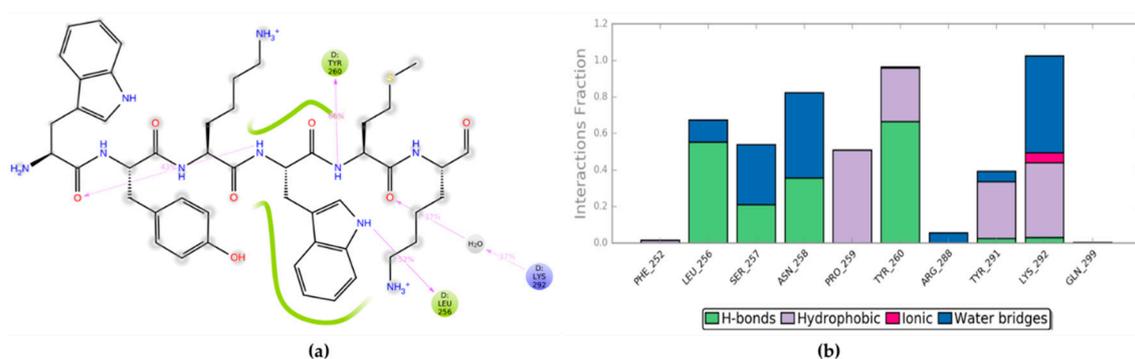
### 2.2. MD Simulations of HOXA9 Hexapeptide and HTL001 Core Peptide in Complex with PBX1 Protein

Before proceeding with the design of novel peptides, other investigations were considered necessary to explore more in details the binding mode of PBX1 with the minimal active HOXA9 sequence (hexapeptide). For this reason, an MD simulation of 200 ns was performed by using the PDB 1PUF including HOXA9, PBX1 and DNA, where HOX protein was modified by deleting all those amino acids not included into the PBX-contacting hexapeptide 196-AANWLH-201. The protein and ligand RMSD plot of the trajectory was analysed by reporting the trend illustrated in Figure S2 in Supplementary materials. The HOXA9-PBX1 interaction diagram and bar chart for this MD simulation are depicted in Figure 4.



**Figure 4.** (a) Diagram of HOXA9 hexapeptide interactions with PBX1 residues during MD simulation; (b) bar chart of protein-ligand interaction occurrences during MD simulation.

As mentioned above, to date the patented peptide HTL001 is being clinically employed, showing promising results by preventing HOX-PBX cooperative binding [49]. In detail, the HTL001 peptide sequence incorporates the hexapeptide WYKWMK responsible for the binding affinity with PBX proteins and a polyarginine coil that functions as a cell-penetrating fragment. Therefore, PDB 1PUF was used for another MD simulation of 200 ns, where HOXA9 hexapeptide was modified and energy minimised to reproduce the HTL001 hexapeptide sequence (WYKWMK), without the polyarginine segment, in complex with PBX1 and DNA. Then, the entire trajectory was monitored and the RMSD plot of PBX1 protein and HTL001 hexapeptide is depicted in Figure S3 in Supplementary materials. Finally, the ligand interaction diagram and the bar chart of protein-ligand interaction occurrences are shown in Figure 5.



**Figure 5.** (a) Diagram of HTL001 hexapeptide (WYKWMK) interactions with PBX1 residues during MD simulation; (b) bar chart of PBX1-HTL001 hexapeptide interaction occurrences during MD simulation.

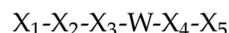
The two above-described MD trajectories were processed to compute MM-GBSA calculations, to analyse the  $\Delta G_{\text{binding}}$  average values, that are reported in Table 2. These results were used as references for the next steps of this work to compare MM-GBSA outputs of the designed peptides below described in complex with PBX1.

**Table 2.** MM-GBSA calculation results of MD simulations performed on HOXA9 and HTL001 hexapeptides in complex with PBX1 protein and DNA.

| Peptide Involved                               | HOXA9 Hexapeptide    | HTL001 Hexapeptide   |
|--|----------------------|----------------------|
| $\Delta G_{\text{binding}}$ average (kcal/mol) | −58.1922             | −53.6882             |
| $\Delta G_{\text{binding}}$ Std. Dev.          | 8.99                 | 8.53                 |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   | −84.6286 to −34.1107 | −78.0904 to −28.9169 |

### 2.3. Residue Scanning of Point-mutated Peptides and Related MD Simulations and MM-GBSA Calculations

HOX proteins have been extensively studied, thus experimental evidence [14,18] highlighted the consensus HOX hexapeptide sequence  $\varphi$ -Y/F-P-W-M-R/K (where  $\varphi$  is a hydrophobic residue) [18]. This consensus motif has been shown responsible for the cooperative binding to PBX proteins and to increase specificity for DNA. Therefore, based on this information from the literature and considering the above-described computational data that shed light on Trp199 as a key amino acid for the interactions, a peptide motif was designed to guide the next steps of this work, as follows:



The amino acid tryptophan was maintained from the consensus HOX motif and the other positions,  $X_1$  to  $X_5$  amino acids, were substituted with non-standard (or non-natural) residues, to generate peptides with different sequence from HTL001.

For this purpose, the SwissSidechain database of non-natural residues populated by the Swiss Institute of Bioinformatics [53] was used. SwissSidechain is a structural and molecular mechanics database of 200 amino acid with non-standard side chains (both D and L conformations), that was developed to study in silico their insertion into natural peptides or proteins. In this work, HOXA9 hexapeptide retrieved from PDB 1PUF was used as a reference to design and identify peptides incorporating non-standard amino acids. The aim was to find peptides able to increase the HOX-PBX inhibitory activity of HTL001 peptide. Indeed, Gfeller et al. [54] demonstrated very good reliability of the

SwissSidechain database based on a comparison between predicted and experimental binding free energies for a BCL9 peptide targeting beta-catenin. These results indicated that such non-natural residues can be used to design novel protein-protein inhibitors. The non-standard side chains of this database were designed based on structural information collected from the Protein Data Bank (PDB) [51] and also commercially available non-natural amino acids. To avoid potential perturbations to peptides or proteins conformation, all those residues that might induce modifications of the backbone (such as  $\beta$ -homo, cyclic or aromatic backbones, or proline derivatives) were not included in the database. Therefore, a total of 200 non-natural side chains populated the SwissSidechain database, where 141 residues were extracted from the Protein Data Bank [51]. All these non-standard residues were collected and used in this work. This database was merged with the non-natural residue library already available in the Schrödinger suite, achieving overall 220 non-standard amino acids. Then point mutations were performed on the HOXA9-PBX1-DNA complex (PDB 1PUF [27]) by running the residue scanning for each of the five  $X_{1-5}$  amino acid positions present in the designed peptide motif and corresponding to Ala196, Ala197, Asn198, Leu200 and His201 of HOXA9 protein. Only Trp199 was maintained unchanged due to its relevance for PBX1 binding [32].

After running the residue scanning calculations, the first four most promising non-standard amino acids were selected for each  $X_{1-5}$  residue of the designed peptide. For this purpose, the residues were chosen according to the following three criteria:

(1) A cut-off of  $-3.0$  kcal/mol for the affinity free-energy difference ( $\Delta\Delta G_{\text{affinity}}$ ) between mutated and wild-type complexes was tuned according to a study performed by Beard et al. [56]. The authors demonstrated a scale factor of 3 to relate the predicted  $\Delta\Delta G_{\text{affinity}}$  values of a mutation through the Schrödinger suite and the experimental energies. According to their study, a computed cutoff of  $-3$  kcal/mol was applied for  $\Delta\Delta G_{\text{affinity}}$  between the mutant and the wild-type form of the protein to predict the key pointed mutations [57];

(2) Negative values for stability free-energy difference ( $\Delta\Delta G_{\text{stability}}$ ) between mutated and wild-type complexes were considered affordable. Due to the lack of a defined tertiary structure of HOX hexapeptide, it was assumed that mutations should not significantly affect the  $\Delta\Delta G_{\text{stability}}$  of the peptide. Therefore, the  $\Delta\Delta G_{\text{stability}}$  values were considered acceptable if negative;

## (3) Commercial availability of non-standard amino acids.

In Table 3 the selected non-standard amino acids are listed, whereas  $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$  and  $X_5$  amino acids provided respectively 119, 10, 52, 35 and 5 acceptable mutations. However, only the best four non-natural amino acids were chosen to proceed with the studies according to the three above-listed criteria.

**Table 3.** The best four non-standard amino acids selected through residue scanning calculations according to  $\Delta\Delta G_{\text{affinity}}$  and  $\Delta\Delta G_{\text{stability}}$ .

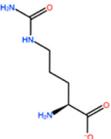
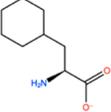
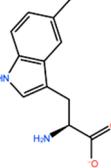
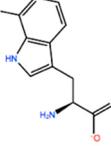
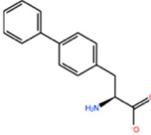
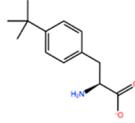
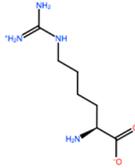
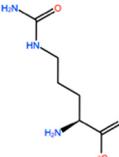
| Corresponding HOXA9 aa | Non-natural Amino Acid | Non-natural Amino Acid Structure  | $\Delta\Delta G_{\text{affinity}}$ | $\Delta\Delta G_{\text{stability}}$ |
|------------------------|------------------------|---|------------------------------------|-------------------------------------|
| ALA196                 | CIR                    |    | −45.128 kcal/mol                   | −0.816 kcal/mol                     |
|                        | ALC                    |    | −18.704 kcal/mol                   | −4.095 kcal/mol                     |
|                        | MTR                    |   | −17.088 kcal/mol                   | −3.208 kcal/mol                     |
|                        | CTE                    |  | −15.008 kcal/mol                   | −5.778 kcal/mol                     |
| ALA197                 | BIF                    |  | −12.892 kcal/mol                   | −3.615 kcal/mol                     |
|                        | TBP                    |  | −8.133 kcal/mol                    | −3.556 kcal/mol                     |
|                        | HRG                    |  | −6.681 kcal/mol                    | −9.341 kcal/mol                     |
|                        | CIR                    |  | −5.075 kcal/mol                    | −2.606 kcal/mol                     |

Table 3. Cont.

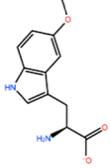
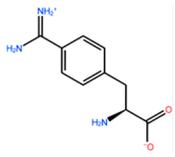
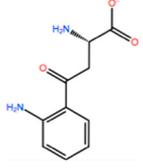
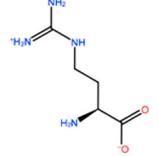
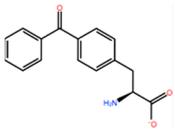
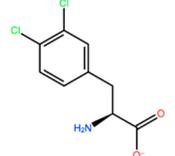
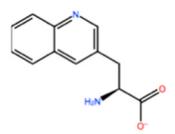
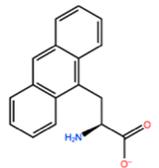
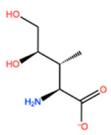
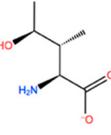
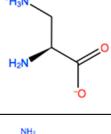
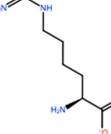
| Corresponding HOXA9 aa | Non-natural Amino Acid | Non-natural Amino Acid Structure  | $\Delta\Delta G_{\text{affinity}}$ | $\Delta\Delta G_{\text{stability}}$ |
|------------------------|------------------------|---|------------------------------------|-------------------------------------|
| ASN198                 | MOT                    |    | −7.505 kcal/mol                    | −13.303 kcal/mol                    |
|                        | 0BN                    |    | −6.051 kcal/mol                    | −10.151 kcal/mol                    |
|                        | KYN                    |    | −6.041 kcal/mol                    | −5.511 kcal/mol                     |
|                        | GBU                    |   | −5.867 kcal/mol                    | −0.688 kcal/mol                     |
| LEU200                 | PBF                    |  | −51.368 kcal/mol                   | −3.302 kcal/mol                     |
|                        | CP3                    |  | −11.929 kcal/mol                   | −1.106 kcal/mol                     |
|                        | QU4                    |  | −11.415 kcal/mol                   | −4.372 kcal/mol                     |
|                        | ANT                    |  | −11.134 kcal/mol                   | −1.353 kcal/mol                     |

Table 3. Cont.

| Corresponding HOXA9 aa | Non-natural Amino Acid | Non-natural Amino Acid Structure  | $\Delta\Delta G_{\text{affinity}}$ | $\Delta\Delta G_{\text{stability}}$ |
|------------------------|------------------------|---|------------------------------------|-------------------------------------|
| HIS201                 | ILX                    |  | −11.562 kcal/mol                   | −1.371 kcal/mol                     |
|                        | HIL                    |  | −10.330 kcal/mol                   | −2.198 kcal/mol                     |
|                        | DPP                    |  | −4.195 kcal/mol                    | −1.029 kcal/mol                     |
|                        | HRG                    |  | −4.018 kcal/mol                    | −5.686 kcal/mol                     |

Each point-mutated HOXA9 peptide was further processed to run an MD simulation of 200 ns in complex with PBX1-DNA to explore the binding stability of each complex for a total of 20 MD simulations. The analysis of the MD outputs showed that the peptides mainly established the crucial interactions with Trp199. The MD trajectories were further processed to compute MM-GBSA calculations, to compare the resulting  $\Delta G_{\text{binding}}$  average values of the complexes to the  $\Delta G_{\text{binding}}$  of the wild-type system, whereas the complex including HOXA9 hexapeptide reported  $\Delta G_{\text{binding-HOXA9}} = -58.1922$  kcal/mol and the one incorporating HTL001 core peptide showed  $\Delta G_{\text{binding-HTL001}} = -53.6882$  kcal/mol. All those mutations reporting  $\Delta G_{\text{binding}}$  average values lower than the two above-mentioned ones were considered for the next steps of the work. In Table 3 MM-GBSA values are reported for each point-mutated peptide.

In detail, for position  $X_1$  of the designed hexapeptide motif three non-natural amino acids (CIR, MTR and ALC) showed low  $\Delta G_{\text{binding}}$  average values, for position  $X_2$  only residue TBP reported good  $\Delta G_{\text{binding}}$  average value; for position  $X_3$  only the second non-standard amino acid 0BN was considered for further analysis; for  $X_4$  all the four amino acids (PBF, CP3, QU4 and ANT) showed good  $\Delta G_{\text{binding}}$  average values; and finally, for  $X_5$  none of the four non-natural amino acids was suitable to be used for the next steps (see Table 4 for MM-GBSA values).

**Table 4.** Data results from MM-GBSA calculations of MD trajectories performed on point mutated HOXA9 peptides in complex with PBX1 and DNA

|  | $X_1 = \text{ALA196 Mutation}$ | $X_2 = \text{ALA197 Mutation}$ | $X_3 = \text{ASN198 Mutation}$ |
|--|--------------------------------|--------------------------------|--------------------------------|
| $\Delta G_{\text{binding}}$ average (kcal/mol) | −68.1064                       | −52.5114                       | −51.3925                       |
| $\Delta G_{\text{binding}}$ Std. Dev.          | 9.19                           | 8.23                           | 9.77                           |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   | −91.4424 to −33.6024           | −77.3444 to −25.9087           | −79.7187 to −21.3466           |
| $\Delta G_{\text{binding}}$ average (kcal/mol) | −58.3419                       | −59.0603                       | −59.1051                       |
| $\Delta G_{\text{binding}}$ Std. Dev.          | 7.37                           | 9.34                           | 8.71                           |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   | −85.6739 to −35.5824           | −80.7314 to −31.1314           | −82.4310 to −27.3523           |

Table 4. Cont.

|  | <b>X<sub>1</sub> = ALA196 Mutation</b> | <b>X<sub>2</sub> = ALA197 Mutation</b> | <b>X<sub>3</sub> = ASN198 Mutation</b> |
|--|--|--|--|
| $\Delta G_{\text{binding}}$ average (kcal/mol) |  | −59.6952                               | −54.3707                               |
| $\Delta G_{\text{binding}}$ Std. Dev.          | <b>ALC</b>                             | 7.39                                   | <b>HRG</b>                             |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   |  | −78.3738 to −26.9511                   | −85.1600 to −24.8071                   |
| $\Delta G_{\text{binding}}$ average (kcal/mol) |  | −56.3011                               | −57.6592                               |
| $\Delta G_{\text{binding}}$ Std. Dev.          | <b>CTE</b>                             | 7.55                                   | <b>CIR</b>                             |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   |  | −80.4156 to −29.7472                   | −81.7311 to −26.2628                   |
|  |  |  | <b>GBU</b>                             |
|  |  |  | 7.97                                   |
|  |  |  | −77.5916 to −30.5304                   |
|  | <b>X<sub>4</sub> = LEU200 Mutation</b> | <b>X<sub>5</sub> = HIS201 Mutation</b> |  |
| $\Delta G_{\text{binding}}$ average (kcal/mol) |  | −68.1857                               | −48.9087                               |
| $\Delta G_{\text{binding}}$ Std. Dev.          | <b>PBF</b>                             | 8.44                                   | <b>ILX</b>                             |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   |  | −95.1687 to −39.7321                   | −74.3783 to −20.2857                   |
| $\Delta G_{\text{binding}}$ average (kcal/mol) |  | −64.6802                               | −50.3148                               |
| $\Delta G_{\text{binding}}$ Std. Dev.          | <b>CP3</b>                             | 9.49                                   | <b>HIL</b>                             |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   |  | −89.3192 to −34.0239                   | −79.2197 to −22.2265                   |
| $\Delta G_{\text{binding}}$ average (kcal/mol) |  | −61.8016                               | −55.6169                               |
| $\Delta G_{\text{binding}}$ Std. Dev.          | <b>QU4</b>                             | 10.53                                  | <b>DPP</b>                             |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   |  | −88.4144 to −33.9895                   | −91.1993 to −22.9813                   |
| $\Delta G_{\text{binding}}$ average (kcal/mol) |  | −63.3043                               | −57.0861                               |
| $\Delta G_{\text{binding}}$ Std. Dev.          | <b>ANT</b>                             | 8.50                                   | <b>HRG</b>                             |
| $\Delta G_{\text{binding}}$ range (kcal/mol)   |  | −87.5841 to −35.9672                   | −86.8681 to −27.0434                   |

#### 2.4. Combinatorial Peptides Generation and Related MD Simulations and MM-GBSA Calculations

The above-described MM-GBSA calculations allowed to select overall nine mutations that were combined based on the designed peptide motif (X<sub>1</sub>-X<sub>2</sub>-X<sub>3</sub>-W-X<sub>4</sub>-X<sub>5</sub>), by employing HOXA9 hexapeptide scaffold. Thus, twelve combinatorial peptides were generated, as listed below, where only two amino acids of HOXA9 were maintained: tryptophan due to its crucial role in PBX-binding and histidine because mutations did not report good  $\Delta G_{\text{binding}}$  average values.

- 1 CIR-TBP-0BN-Trp-PBF-His
- 2 CIR-TBP-0BN-Trp-CP3-His
- 3 CIR-TBP-0BN-Trp-QU4-His
- 4 CIR-TBP-0BN-Trp-ANT-His
- 5 ALC-TBP-0BN-Trp-PBF-His
- 6 ALC-TBP-0BN-Trp-CP3-His
- 7 ALC-TBP-0BN-Trp-QU4-His
- 8 ALC-TBP-0BN-Trp-ANT-His
- 9 MTR-TBP-0BN-Trp-ANT-His
- 10 MTR-TBP-0BN-Trp-CP3-His
- 11 MTR-TBP-0BN-Trp-QU4-His
- 12 MTR-TBP-0BN-Trp-ANT-His

These combinatorial peptides in complex with PBX1 protein and DNA were used to run MD simulations of 200 ns per complex. Thus, twelve MD were run and the RMSD values were plotted for each ternary complex DNA-protein-peptide, as depicted in Table S1 in Supplementary materials. On the other hand, Table S2 in Supplementary materials depicts the bar charts of protein-ligand interactions and the plots illustrating the frequency

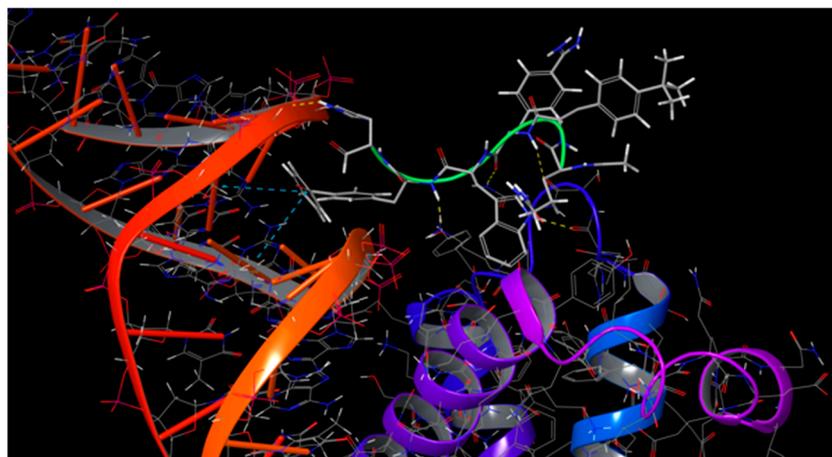
of the interactions during the MD trajectories. The analysis of these RMSD plots confirmed the stability of the simulated systems. Furthermore, the bar charts and the interaction frequency plots highlighted that the twelve combinatorial peptides met most of the key interactions previously identified from the analysis of PDB 1PUF structure and MD simulations of HOXA9 protein, HOXA9 hexapeptide and HTL001 peptide with PBX1 and DNA. More details about statistics of H-bonds and  $\pi$ -stacking established by the twelve peptides have been reported in Table S3 in Supplementary materials.

Finally, MM-GBSA calculations were performed and the results are reported in Table 5. Even for these peptides, the resulting  $\Delta G_{\text{binding}}$  average values were compared to those retrieved from MD simulations of HOXA9 and HTL001 hexapeptides in complex with PBX1 and DNA. Only the sixth peptide showed a higher  $\Delta G_{\text{binding}}$  average value, hence it was not considered for further studies.

**Table 5.** MM-GBSA calculation results of MD simulations performed on the combinatorial peptides in complex with PBX1 protein and DNA.

| PEPTIDE INVOLVED                      | First Peptide                 | Second Peptide                 | Third Peptide                  |
|---------------------------------------|-------------------------------|--------------------------------|--------------------------------|
| $\Delta G_{\text{binding}}$ average   | −79.6771 kcal/mol             | −61.8602 kcal/mol              | −68.0795 kcal/mol              |
| $\Delta G_{\text{binding}}$ Std. Dev. | 10.18                         | 12.72                          | 10.63                          |
| $\Delta G_{\text{binding}}$ range     | −104.585 to −38.2615 kcal/mol | −99.0013 to −30.7190 kcal/mol  | −98.3690 to −29.5313 kcal/mol  |
| PEPTIDE INVOLVED                      | Fourth Peptide                | Fifth Peptide                  | Sixth Peptide                  |
| $\Delta G_{\text{binding}}$ average   | −64.6664 kcal/mol             | −81.8766 kcal/mol              | −55.1927 kcal/mol              |
| $\Delta G_{\text{binding}}$ Std. Dev. | 7.53                          | 7.44                           | 10.09                          |
| $\Delta G_{\text{binding}}$ range     | −87.5689 to −30.1013 kcal/mol | −101.5164 to −45.3623 kcal/mol | −85.5158 to −22.5652 kcal/mol  |
| PEPTIDE INVOLVED                      | Seventh Peptide               | Eighth Peptide                 | Ninth Peptide                  |
| $\Delta G_{\text{binding}}$ average   | −62.8885 kcal/mol             | −71.9163 kcal/mol              | −74.0909 kcal/mol              |
| $\Delta G_{\text{binding}}$ Std. Dev. | 9.19                          | 9.19                           | 11.42                          |
| $\Delta G_{\text{binding}}$ range     | −89.1247 to −19.4438 kcal/mol | −101.6790 to −44.4808 kcal/mol | −105.5444 to −32.2303 kcal/mol |
| PEPTIDE INVOLVED                      | Tenth Peptide                 | Eleventh Peptide               | Twelfth Peptide                |
| $\Delta G_{\text{binding}}$ average   | −60.2167 kcal/mol             | −65.0198 kcal/mol              | −68.3222 kcal/mol              |
| $\Delta G_{\text{binding}}$ Std. Dev. | 9.56                          | 8.24                           | 8.13                           |
| $\Delta G_{\text{binding}}$ range     | −89.0633 to −28.6783 kcal/mol | −89.3709 to −36.2812 kcal/mol  | −95.3406 to −42.1593 kcal/mol  |

The analysis of the MD trajectories shed light on an interesting observed binding behaviour of the combinatorial peptides. Indeed, the tryptophan amino acid was kept stuck within the hydrophobic pocket of PBX1 during all the twelve MD simulations. Moreover, the selected non-standard amino acids in position  $X_1$  of the designed peptides presented hydrophobic side chains, as required from the consensus HOX hexapeptide [18]. These amino acids in position  $X_1$  showed also to stack their side chain into the DNA minor groove by establishing  $\pi$ -stacking contacts with the nitrogenous bases (Figure 6). Moreover, the interactions between DNA and combinatorial peptides were also explored and Table S4 depicts the interaction histograms for each complex DNA-PBX1-peptide in Supplementary materials. The analysis of the histogram plots highlighted three combinatorial peptides with the lowest  $\Delta G_{\text{binding}}$  values (i.e.,  $\Delta G_{\text{first\_peptide}} = -79.6771$  kcal/mol,  $\Delta G_{\text{fifth\_peptide}} = -81.8766$  kcal/mol and  $\Delta G_{\text{ninth\_peptide}} = -74.0909$  kcal/mol, respectively) as the ones establishing  $\pi$ -stacking interaction with the DNA minor groove, especially with the deoxyguanosine 28 (DG28). Furthermore, hydrophobic interactions and H-bonds between peptides and DNA also seemed to be important for the ternary complex stabilisation. In this case, DG28 had again an important role for the stabilisation, being, together with deoxyadenosine 16 (DA16), the main residue involved in H-bonds interactions. Indeed, the sixth peptide, which reported a  $\Delta G_{\text{binding}}$  value of  $-55.1927$  kcal/mol, exhibited no  $\pi$ -stacking and H-bonds and very few hydrophobic contacts with DNA. These findings suggested that the contacts between combinatorial peptides and DNA minor groove might contribute to stabilising the complex DNA-PBX1-peptide. Table S5 in Supplementary materials illustrates the combinatorial peptide binding modes with PBX1 protein and DNA.



**Figure 6.** Binding mode of the first combinatorial peptide to PBX1 and DNA.

Finally, considering that both proteins HOX and PBX are involved in transcriptional events, to check the potential ability of the combinatorial peptides to permeate cell membranes [58], the polar surface area (PSA) and  $\log P_{o/w}$  [59] were computed and Table 6 lists the values of the twelve designed peptides.

**Table 6.** PSA and  $\log P_{o/w}$  values of the twelve combinatorial peptides.

|    | PEPTIDES                | PSA    | $\log P_{o/w}$ |
|----|-------------------------|--------|----------------|
| 1  | CIR-TBP-0BN-TRP-PBF-HIS | 302.15 | 3.2            |
| 2  | CIR-TBP-0BN-TRP-CP3-HIS | 310.38 | 2.6            |
| 3  | CIR-TBP-0BN-TRP-QU4-HIS | 322.80 | 2.0            |
| 4  | CIR-TBP-0BN-TRP-ANT-HIS | 307.47 | 3.3            |
| 5  | ALC-TBP-0BN-TRP-PBF-HIS | 300.04 | 6.1            |
| 6  | ALC-TBP-0BN-TRP-CP3-HIS | 284.52 | 6.0            |
| 7  | ALC-TBP-0BN-TRP-QU4-HIS | 297.08 | 5.4            |
| 8  | ALC-TBP-0BN-TRP-ANT-HIS | 250.48 | 5.9            |
| 9  | MTR-TBP-0BN-TRP-PBF-HIS | 276.41 | 4.3            |
| 10 | MTR-TBP-0BN-TRP-CP3-HIS | 268.43 | 5.4            |
| 11 | MTR-TBP-0BN-TRP-QU4-HIS | 280.80 | 4.8            |
| 12 | MTR-TBP-0BN-TRP-ANT-HIS | 260.87 | 6.0            |

### 3. Methods

#### 3.1. Preparation of HOXA9-PBX1-DNA Complex

The 3D trimeric complex of HOXA9-PBX1-DNA was downloaded from the Protein Data Bank [51] (PDB ID: 1PUF) and imported in Schrödinger suite (Schrödinger Inc., New

York, NY, USA, software release v2018-4,) to optimise the structure by using the "Protein preparation" tool [60]. The bond orders for untemplated residues were assigned by using known HET groups based on their SMILES strings in Chemical Component Dictionary. Hydrogen atoms were added to the structure, eventual bonds to metals were broken, zero-order bonds between metals and nearby atoms were added and formal charges to metals and neighbouring atoms were corrected. Disulphide bonds between two sulphur atoms, if they were close to each other, were created and water molecules beyond 5.0 Å from any of the HET groups, including ions, were deleted. Then, protonation and metal charge states for the ligands, cofactors and metals were generated [61,62]. Finally, PROPKA [62] was run under pH 7.0 to optimise hydroxyl, Asn, Gln and His states using ProtAssign.

### 3.2. HOXA9 Hexapeptide Residues Scanning Using Non-standard "SwissSidechain" Amino Acids

The SwissSidechain database of non-natural amino acids was downloaded from the Swiss Institute of Bioinformatics website [53] as a .nsr file including parameter and topology data [54]. It was imported in Schrödinger tool "Manage non-standard amino acids" and joined with the non-natural residue library available in the Schrödinger suite, achieving overall 220 non-standard amino acids. The joined database was used to perform point mutations through the "Residue Scanning" tool of Schrödinger suite on HOXA9 residues (Ala196, Ala197, Asn198, Trp199, Leu200 and His201) by using PDB 1PUF [27].

The predicted changes in binding affinity and stability were calculated according to Equation (1) [56]. The resulting structures were refined by selecting side-chain prediction with backbone minimization.

$$\Delta\Delta G_{Affinity} = \left( E_{A\Delta B}^{MUT} - E_A^{MUT} - E_B^{MUT} \right) - \left( E_{A\Delta B}^{WT} - E_A^{WT} - E_B^{WT} \right) \quad (1)$$

Where  $E$  is the calculated energy of each protein ( $A$  and  $B$ ) or complex ( $A \cdot B$ ) after refinement considering the mutant form ( $MUT$ ) and the wild-type ( $WT$ ) of the protein.

$\Delta\Delta G_{stability}$  values are calculated according to Equation (2). For the purpose of the model,  $\Delta\Delta G_{stability}$  was computed representing the unfolded ligand as a tripeptide,  $A-X-B$ , where  $X$  is the residue that is mutated, and  $A$  and  $B$  are its neighbours, capped with ACE and NMA. The assumption is that the remaining interactions in the unfolded state are negligible.

$$\Delta\Delta G_{Stability} = \left( E_{L(u)}^{MUT} - E_{L(f)}^{MUT} \right) - \left( E_{L(u)}^{WT} - E_{L(f)}^{WT} \right) \quad (2)$$

Where  $E$ , in this case, is the calculated energy for the unfolded parent ligand ( $L(u)$ ) and the folded parent ligand ( $L(f)$ ) considering the mutant form ( $MUT$ ) and the wild-type ( $WT$ ) of the protein [56]. The calculations are done with Prime MM-GBSA [63,64], which employs an implicit (continuum) solvation model.

### 3.3. MD Simulations of PBX1-DNA in Complex with HOXA9 Protein, HOXA9 Hexapeptide, HTL001 Core Peptide, Point-mutated Peptides and Combinatorial Peptides

In this work overall, thirty-five MD simulations of 200 ns per each were performed using Desmond [65] and they were run by applying the following settings. The systems were created using the "System builder" tool of the Schrödinger suite. TIP3P [66] was selected as a solvent model and the orthorhombic shape box was chosen. The selected box size calculation method was buffer and the box side distances were set 10 Å. The force field OPLS3 [67] was applied and the system was neutralized by adding  $\text{Na}^+$  ions. The outputs were further processed by performing MD simulations of 200 ns. The ensemble class NPT was chosen to maintain the number of atoms, the pressure and the temperature constant for the entire trajectories. The thermostat method employed was the Nose–Hoover chain with a relaxation time of 1.0 ps and a temperature of 300 K. The barostat method applied was Martyna-Tobias-Klein with a relaxation time of 2.0 ps and an isotropic coupling style. The timestep for numerical integration was 2.0 fs for bonded interactions, 2.0 fs for nonbonded near (van der Waals and short-range electrostatic interactions), and 6.0 fs for nonbonded-far (long-range electrostatic interactions). For Coulombic interactions, a cut-off radius of 9.0 Å

was tuned as a short-range method. Pressure and temperature were set at 1.01325 bar and 300 K, respectively. Finally, the systems were relaxed before starting simulations according to the following steps:

- (1) Minimization with the solute restrained;
- (2) Minimization without restraints;
- (3) Simulation in the NVT ensemble using a Berendsen thermostat with a simulation time of 12 ps, a temperature of 10 K, a fast temperature relaxation constant, velocity resampling every 1 ps, and non-hydrogen solute atoms restrained;
- (4) Simulation in the NPT ensemble using a Berendsen thermostat and a Berendsen barostat with a simulation time of 12 ps, a temperature of 10 K and a pressure of 1 atm, a fast temperature relaxation constant, a slow pressure relaxation constant, velocity resampling every 1 ps, and non-hydrogen solute atoms restrained;
- (5) Simulation in the NPT ensemble using a Berendsen thermostat and a Berendsen barostat with a simulation time of 24 ps, a temperature of 300 K and a pressure of 1 atm, a fast temperature relaxation constant, a slow pressure relaxation constant, velocity resampling every 1 ps, and non-hydrogen solute atoms restrained;
- (6) Simulation in the NPT ensemble using a Berendsen thermostat and a Berendsen barostat with a simulation time of 24 ps, a temperature of 300 K and a pressure of 1 atm, a fast temperature relaxation constant and a normal pressure relaxation constant.

### 3.4. MM-GBSA Calculations Performed on the Protein-peptide Complexes

The MD outputs of HOXA9 hexapeptide, HTL001 core peptide, point-mutated HOXA9 peptides and combinatorial peptides in complex with PBX1 protein and DNA were used to compute MM-GBSA calculations through the command line. For this purpose, the Python script "thermal\_mmgbsa.py" was run.

Overall thirty-four MM-GBSA calculations were carried out using VSGB as a solvation model and OPLS3 FF was set for each MD trajectory. The  $\Delta G_{binding}$  values were computed for each trajectory frames according to Equation (3).

$$\Delta G_{binding} = E_{A\Delta B (minimized)} - E_A (minimized) - E_B (minimized) \quad (3)$$

Where  $E$  is the calculated energy of complex ( $A \cdot B$ ) or each protein ( $A$  and  $B$ ) after minimization [68]. Finally, the average of  $\Delta G_{binding}$  values of the entire trajectories was calculated and the results are above reported in Tables 2–4 of the "Results and discussion" section.

## 4. Conclusions

This work was based on the use of computational methods to design new non-standard peptides to inhibit HOX-PBX1 interaction. The above-described methods started from literature evidence [3,14,18,27,31,32] as an initial guide for the in silico design of new peptide inhibitors. In details, MD simulations of the newly designed peptides, in complex with PBX1 protein and DNA, reported promising results according to the binding mode, the predicted  $\Delta G_{binding}$  average values and the physicochemical properties. From our extensive analysis, we got some important hints for the design of these new inhibitors. The first amino acid of the sequence should be hydrophobic and possibly aromatic to establish a  $\pi$ -stacking interaction with the DNA minor groove. The TRP in the fourth position seemed to be crucial for the stabilisation of the complex. From our computational analysis, the sixth amino acid (HIS) appeared not to be substitutable.

The most promising peptides found out from this work will be soon synthesized, to test their efficacy in preventing HOX-PBX1 cooperative binding. In detail, biological assays will be conducted on ovarian, breast, and melanoma cancer cell lines, because of the overexpression of *HOX* genes.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijms22115670/s1>.

**Author Contributions:** Conceptualization, M.R.G., J.J., and A.B.; methodology, M.R.G. and A.B.; formal analysis, U.P. and A.B.; investigation, M.R.G.; resources, A.B.; data curation and analysis, M.R.G., G.D.S., U.P., and A.B.; writing—original draft preparation, M.R.G.; writing—review and editing, M.R.G., U.P., and A.B.; visualization, U.P. and A.B.; supervision, A.B.; project administration, A.B.; funding acquisition, U.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research activity of M.R.G. was funded by the Regione Siciliana, Assessorship of Productive Activities—Department of Productive Activities, funds: FSC 2014/2020, project name: *Computational Molecular Design e Screening – CheMIST - CUPG77B17000110001*, Scientific Research within the “Patto per il sud” of the Sicily Region and the APC was funded by Regione Siciliana, Assessorship of Productive Activities—Department of Productive Activities, Action 1.1.5 of PO FESR Sicilia 2014/2020, Project n. 086202000366 – “OBIND”, CUP G29J18000700007 to U.P.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Pearson, J.C.; Lemons, D.; McGinnis, W. Modulating *Hox* gene functions during animal body patterning. *Nat. Rev. Genet.* **2005**, *6*, 893–904. [CrossRef]
- Krumlauf, R. *Hox* genes in vertebrate development. *Cell* **1994**, *78*, 191–201. [CrossRef]
- Gehring, W.J.; Qian, Y.Q.; Billeter, M.; Furukubo-Tokunaga, K.; Schier, A.F.; Resendez-Perez, D.; Affolter, M.; Otting, G.; Wüthrich, K. Homeodomain-DNA recognition. *Cell* **1994**, *78*, 211–223. [CrossRef]
- Duboule, D. The rise and fall of *Hox* gene clusters. *Development* **2007**, *134*, 2549–2560. [CrossRef]
- Mallo, M.; Alonso, C.R. The regulation of *Hox* gene expression during animal development. *Development* **2013**, *140*, 3951–3963. [CrossRef] [PubMed]
- McGinnis, N.; Kuziora, M.A.; McGinnis, W. Human *Hox-4.2* and *Drosophila Deformed* encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* **1990**, *63*, 969–976. [CrossRef]
- Quinonez, S.C.; Innis, J.W. Human *HOX* gene disorders. *Mol. Genet. Metab.* **2014**, *111*, 4–15. [CrossRef]
- Hombria, J.C.-G.; Lovegrove, B. Beyond homeosis—HOX function in morphogenesis and organogenesis. *Differentiation* **2003**, *71*, 461–476. [CrossRef] [PubMed]
- Sánchez-Herrero, E. *Hox* Targets and Cellular Functions. *Scientifica* **2013**, *2013*, 738257. [CrossRef] [PubMed]
- Mann, R.S.; Chan, S.-K. Extra specificity from extradenticle: The partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet.* **1996**, *12*, 258–262. [CrossRef]
- Hayashi, S.; Scott, M.P. What determines the specificity of action of *Drosophila* homeodomain proteins? *Cell* **1990**, *63*, 883–894. [CrossRef]
- Berger, M.F.; Badis, G.; Gehrke, A.R.; Talukder, S.; Philippakis, A.A.; Peña-Castillo, L.; Alleyne, T.M.; Mnaimneh, S.; Botvinnik, O.B.; Chan, E.T.; et al. Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell* **2008**, *133*, 1266–1276. [CrossRef] [PubMed]
- Noyes, M.B.; Christensen, R.G.; Wakabayashi, A.; Stormo, G.D.; Brodsky, M.H.; Wolfe, S.A. Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. *Cell* **2008**, *133*, 1277–1289. [CrossRef] [PubMed]
- Knoepfler, P.S.; Kamps, M.P. The pentapeptide motif of *Hox* proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. *Mol. Cell. Biol.* **1995**, *15*, 5811–5819. [CrossRef] [PubMed]
- Moens, C.B.; Selli, L. *Hox* cofactors in vertebrate development. *Dev. Biol.* **2006**, *291*, 193–206. [CrossRef] [PubMed]
- Longobardi, E.; Penkov, D.; Mateos, D.; De Florian, G.; Torres, M.; Blasi, F. Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. *Dev. Dyn.* **2014**, *243*, 59–75. [CrossRef]
- Peifer, M.; Wieschaus, E. Mutations in the *Drosophila* gene extradenticle affect the way specific homeo domain proteins regulate segmental identity. *Genes Dev.* **1990**, *4*, 1209–1223. [CrossRef]
- Piper, D.E.; Batchelor, A.H.; Chang, C.-P.; Cleary, M.L.; Wolberger, C. Structure of a HoxB1–Pbx1 Heterodimer Bound to DNA. *Cell* **1999**, *96*, 587–597. [CrossRef]
- Brendolan, A. A Pbx1-dependent genetic and transcriptional network regulates spleen ontogeny. *Development* **2005**, *132*, 3113–3126. [CrossRef]
- Allen, T.D.; Zhlp, Y.-X.; Hawley, T.S.; Hawley, R.G. TALE Homeoproteins as HOX11-Interacting Partners in T-cell Leukemia. *Leuk. Lymphoma* **2000**, *39*, 241–256. [CrossRef]

21. Chan, S.-K.; Jaffe, L.; Capovilla, M.; Botas, J.; Mann, R.S. The DNA binding specificity of ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* **1994**, *78*, 603–615. [[CrossRef](#)]
22. van Dijk, M.A.; Murre, C. extradenticle Raises the DNA binding specificity of homeotic selector gene products. *Cell* **1994**, *78*, 617–624. [[CrossRef](#)]
23. Phelan, M.L.; Rambaldi, I.; Featherstone, M.S. Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol. Cell. Biol.* **1995**, *15*, 3989–3997. [[CrossRef](#)]
24. Abu-Shaar, M.; Ryoo, H.D.; Mann, R.S. Control of the nuclear localization of Extradenticle by competing nuclear import and export signals. *Genes Dev.* **1999**, *13*, 935–945. [[CrossRef](#)]
25. Stevens, K.E.; Mann, R.S. A Balance Between Two Nuclear Localization Sequences and a Nuclear Export Sequence Governs Extradenticle Subcellular Localization. *Genetics* **2007**, *175*, 1625–1636. [[CrossRef](#)]
26. Saleh, M.; Huang, H.; Green, N.C.; Featherstone, M.S. A Conformational Change in PBX1A Is Necessary for Its Nuclear Localization. *Exp. Cell Res.* **2000**, *260*, 105–115. [[CrossRef](#)]
27. LaRonde-LeBlanc, N.A.; Wolberger, C. Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev.* **2003**, *17*, 2060–2072. [[CrossRef](#)]
28. Merabet, S.; Hudry, B.; Saadaoui, M.; Graba, Y. Classification of sequence signatures: A guide to Hox protein function. *BioEssays* **2009**, *31*, 500–511. [[CrossRef](#)]
29. Johnson, F.B.; Parker, E.; Krasnow, M.A. Extradenticle protein is a selective cofactor for the Drosophila homeotics: Role of the homeodomain and YPWM amino acid motif in the interaction. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 739–743. [[CrossRef](#)]
30. Joshi, R.; Passner, J.M.; Rohs, R.; Jain, R.; Sosinsky, A.; Crickmore, M.A.; Jacob, V.; Aggarwal, A.K.; Honig, B.; Mann, R.S. Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. *Cell* **2007**, *131*, 530–543. [[CrossRef](#)]
31. Passner, J.M.; Ryoo, H.D.; Shen, L.; Mann, R.S.; Aggarwal, A.K. Structure of a DNA-bound Ultrabithorax–Extradenticle homeodomain complex. *Nature* **1999**, *397*, 714–719. [[CrossRef](#)]
32. Shen, W.-F.; Chang, C.-P.; Rozenfeld, S.; Sauvageau, G.; Humphries, R.K.; Lu, M.; Lawrence, H.J.; Cleary, M.L.; Largman, C. Hox Homeodomain Proteins Exhibit Selective Complex Stabilities with Pbx and DNA. *Nucleic Acids Res.* **1996**, *24*, 898–906. [[CrossRef](#)]
33. Joshi, R.; Sun, L.; Mann, R. Dissecting the functional specificities of two Hox proteins. *Genes Dev.* **2010**, *24*, 1533–1545. [[CrossRef](#)]
34. Chang, C.P.; Brocchieri, L.; Shen, W.F.; Largman, C.; Cleary, M.L. Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. *Mol. Cell. Biol.* **1996**, *16*, 1734–1745. [[CrossRef](#)]
35. Lu, Q.; Kamps, M.P. Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: Proposal for a model of a Pbx1–Hox–DNA complex. *Mol. Cell. Biol.* **1996**, *16*, 1632–1640. [[CrossRef](#)]
36. Peltenburg, L.T.; Murre, C. Specific residues in the Pbx homeodomain differentially modulate the DNA-binding activity of Hox and Engrailed proteins. *Development* **1997**, *124*, 1089–1098. [[CrossRef](#)]
37. Chang, C.P.; Shen, W.F.; Rozenfeld, S.; Lawrence, H.J.; Largman, C.; Cleary, M.L. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev.* **1995**, *9*, 663–674. [[CrossRef](#)]
38. Perricone, U.; Gulotta, M.R.; Lombino, J.; Parrino, B.; Cascioferro, S.; Diana, P.; Cirrincione, G.; Padova, A. An overview of recent molecular dynamics applications as medicinal chemistry tools for the undruggable site challenge. *MedChemComm* **2018**, *9*, 920–936. [[CrossRef](#)]
39. Gulotta, M.R.; Vittorio, S.; Gitto, R.; Perricone, U.; De Luca, L. Exploring Molecular Contacts of MUC1 at CIN85 Binding Interface to Address Future Drug Design Efforts. *Int. J. Mol. Sci.* **2021**, *22*, 2208. [[CrossRef](#)]
40. Ji, T.; Lee, M.; Pruitt, S.C.; Hangauer, D.G. Privileged scaffolds for blocking protein–protein interactions: 1,4-disubstituted naphthalene antagonists of transcription factor complex HOX–PBX/DNA. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3875–3879. [[CrossRef](#)]
41. Morgan, R.G.L.; Sohal, J. HOX peptides as PBX modulators for the treatment of cancer. European Patent EP1909811B1, 2009. Available online: <https://patentimages.storage.googleapis.com/09/d8/95/6ccfb3d2ef2f38/EP1909811B1.pdf> (accessed on 21 May 2020).
42. Morgan, R.; Pirard, P.M.; Shears, L.; Sohal, J.; Pettengell, R.; Pandha, H.S. Antagonism of HOX/PBX Dimer Formation Blocks the In vivo Proliferation of Melanoma. *Cancer Res.* **2007**, *67*, 5806–5813. [[CrossRef](#)]
43. Plowright, L.; Harrington, K.J.; Pandha, H.S.; Morgan, R. HOX transcription factors are potential therapeutic targets in non-small-cell lung cancer (targeting HOX genes in lung cancer). *Br. J. Cancer* **2009**, *100*, 470–475. [[CrossRef](#)]
44. Morgan, R.; Boxall, A.; Harrington, K.J.; Simpson, G.R.; Gillett, C.; Michael, A.; Pandha, H.S. Targeting the HOX/PBX dimer in breast cancer. *Breast Cancer Res. Treat.* **2012**, *136*, 389–398. [[CrossRef](#)] [[PubMed](#)]
45. Morgan, R.; Boxall, A.; Harrington, K.J.; Simpson, G.R.; Michael, A.; Pandha, H.S. Targeting HOX transcription factors in prostate cancer. *BMC Urol.* **2014**, *14*, 17. [[CrossRef](#)] [[PubMed](#)]
46. Morgan, R.; Simpson, G.; Gray, S.; Gillett, C.; Tabi, Z.; Spicer, J.; Harrington, K.J.; Pandha, H.S. HOX transcription factors are potential targets and markers in malignant mesothelioma. *BMC Cancer* **2016**, *16*, 85. [[CrossRef](#)]
47. Errico, M.C.; Felicetti, F.; Bottero, L.; Mattia, G.; Boe, A.; Felli, N.; Petrini, M.; Bellenghi, M.; Pandha, H.S.; Calvaruso, M.; et al. The abrogation of the HOXB7/PBX2 complex induces apoptosis in melanoma through the miR-221&222-c-FOS pathway. *Int. J. Cancer* **2013**, *133*, 879–892. [[PubMed](#)]
48. Ando, H.; Natsume, A.; Senga, T.; Watanabe, R.; Ito, I.; Ohno, M.; Iwami, K.; Ohka, F.; Motomura, K.; Kinjo, S.; et al. Peptide-based inhibition of the HOXA9/PBX interaction retards the growth of human meningioma. *Cancer Chemother. Pharmacol.* **2014**, *73*, 53–60. [[CrossRef](#)]

49. Primon, M.; Hoffman, E.; Pandha, H.S.; Morgan, R. HTL001, a novel inhibitor of HOX/PBX binding, is highly cytotoxic to prostate and breast cancer cells. *Eur. J. Cancer* **2016**, *69*, S133. [CrossRef]
50. Shears, L.; Plowright, L.; Harrington, K.; Pandha, H.S.; Morgan, R. Disrupting the Interaction Between HOX and PBX Causes Necrotic and Apoptotic Cell Death in the Renal Cancer Lines CaKi-2 and 769-P. *J. Urol.* **2008**, *180*, 2196–2201. [CrossRef]
51. Protein Data Bank. Available online: <https://pdb101.rcsb.org> (accessed on 21 May 2020).
52. Perricone, U.; Wieder, M.; Seidel, T.; Langer, T.; Padova, A.; Almerico, A.M.; Tutone, M. A Molecular Dynamics–Shared Pharmacophore Approach to Boost Early-Enrichment Virtual Screening: A Case Study on Peroxisome Proliferator-Activated Receptor  $\alpha$ . *ChemMedChem* **2017**, *12*, 1399–1407. [CrossRef]
53. Database of Non-Standard Amino Acids. Available online: [https://www.swissidechain.ch/](https://www.swisssidechain.ch/) (accessed on 27 July 2020).
54. Gfeller, D.; Michielin, O.; Zoete, V. Expanding molecular modeling and design tools to non-natural sidechains. *J. Comput. Chem.* **2012**, *33*, 1525–1535. [CrossRef]
55. Gulotta, M.R.; Lombino, J.; Perricone, U.; De Simone, G.; Mekni, N.; De Rosa, M.; Diana, P.; Padova, A. Targeting SARS-CoV-2 RBD Interface: A Supervised Computational Data-Driven Approach to Identify Potential Modulators. *ChemMedChem* **2020**, *15*, 1921–1931. [CrossRef] [PubMed]
56. Beard, H.; Cholleti, A.; Pearlman, D.; Sherman, W.; Loving, K.A. Applying Physics-Based Scoring to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes. *PLoS ONE* **2013**, *8*, e82849. [CrossRef]
57. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581*, 215–220. [CrossRef]
58. Matsson, P.; Kihlberg, J. How Big Is Too Big for Cell Permeability? *J. Med. Chem.* **2017**, *60*, 1662–1664. [CrossRef] [PubMed]
59. Duffy, E.M.; Jorgensen, W.L. Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000**, *122*, 2878–2888. [CrossRef]
60. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221–234. [CrossRef] [PubMed]
61. Greenwood, J.R.; Calkins, D.; Sullivan, A.P.; Shelley, J.C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **2010**, *24*, 591–604. [CrossRef] [PubMed]
62. Olsson, M.H.M.; Søndergaard, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537. [CrossRef]
63. Jacobson, M.P.; Friesner, R.A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J. Mol. Biol.* **2002**, *320*, 597–608. [CrossRef]
64. Jacobson, M.P.; Pincus, D.L.; Rapp, C.S.; Day, T.J.F.; Honig, B.; Shaw, D.E.; Friesner, R.A. A hierarchical approach to all-atom protein loop prediction. *Proteins Struct. Funct. Bioinform.* **2004**, *55*, 351–367. [CrossRef]
65. Bowers, K.J.; Chow, E.; Xu, H.; Dror, R.O.; Eastwood, M.P.; Gregersen, B.A.; Klepeis, J.L.; Kolossvary, I.; Moraes, M.A.; Sacerdoti, F.D.; et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, Tampa, FL, USA, November 2006; Association for Computing Machinery: New York, NY, USA.
66. Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960. [CrossRef]
67. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J.Y.; Wang, L.; Lupyan, D.; Dahlgren, M.K.; Knight, J.L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296. [CrossRef] [PubMed]
68. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461. [CrossRef] [PubMed]