



Article

Construction of Histone–Protein Complex Structures by Peptide Growing

Balázs Zoltán Zsidó [†], Bayartsetseg Bayarsaikhan [†], Rita Börzsei and Csaba Hetényi ^{*ID}

Pharmacoinformatics Unit, Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, Szigeti Út 12, 7624 Pécs, Hungary; zsidobalazs@pte.hu (B.Z.Z.); bayartsetseg704@yahoo.com (B.B.); rita.borzsei@aok.pte.hu (R.B.)

* Correspondence: hetenyi.csaba@pte.hu; Tel.: +36-72-536-000 (ext. 38201)

[†] These authors contributed equally to this work.

Abstract: The structures of histone complexes are master keys to epigenetics. Linear histone peptide tails often bind to shallow pockets of reader proteins via weak interactions, rendering their structure determination challenging. In the present study, a new protocol, PepGrow, is introduced. PepGrow uses docked histone fragments as seeds and grows the full peptide tails in the reader-binding pocket, producing atomic-resolution structures of histone–reader complexes. PepGrow is able to handle the flexibility of histone peptides, and it is demonstrated to be more efficient than linking pre-docked peptide fragments. The new protocol combines the advantages of popular program packages and allows fast generation of solution structures. AutoDock, a force-field-based program, is used to supply the docked peptide fragments used as structural seeds, and the building algorithm of Modeller is adopted and tested as a peptide growing engine. The performance of PepGrow is compared to ten other docking methods, and it is concluded that in situ growing of a ligand from a seed is a viable strategy for the production of complex structures of histone peptides at atomic resolution.

Keywords: docking; histone; peptide; ligand; fragment; growing



Citation: Zsidó, B.Z.; Bayarsaikhan, B.; Börzsei, R.; Hetényi, C. Construction of Histone–Protein Complex Structures by Peptide Growing. *Int. J. Mol. Sci.* **2023**, *24*, 13831. <https://doi.org/10.3390/ijms241813831>

Academic Editor: Wolfgang Sippl

Received: 21 August 2023

Revised: 4 September 2023

Accepted: 5 September 2023

Published: 7 September 2023



Copyright: © 2023 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

Histones have a diverse interaction profile [1] and play a key role in epigenetic regulation via interactions with the DNA in the chromatin [2,3], as well as various protein partners [4,5]. Readers are important proteins that distinguish between the combinatorial numbers of post-translationally modified histone molecules commonly called as the “histone code” [6]. The atomic-resolution structures of histone–reader complexes are key to understanding the “histone code” and designing new drugs that affect epigenetic regulation [6–8]. The present study is focused on consisting of histone H3 peptides and their reader proteins, which play an important role in the pathophysiology of various autoimmune diseases, intellectual disabilities, cancer development, such as breast cancer, colorectal carcinoma and hematopoietic cancers, autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy, meiotic defects in spermatocytes, breast, prostate and colorectal cancers, and leukemia (Table S1 [9–18]). These pathophysiological involvements render histone reader proteins such as bromodomains [19] and the eleven–nineteen leukemia protein (ENL [20]) attractive targets for drug design purposes.

While knowledge of the structures of histone H3–reader complexes is necessary for understanding the pathomechanism of epigenetic diseases and designing new drugs to act against them, the determination of their atomic-resolution structures can be rather challenging [21]. Experimental difficulties are presented by the creation of well-diffracting and stable crystals in X-ray crystallography [22], the computational processing of noisy images in cryo-electron microscopy [23], and the isotopic labeling of proteins in NMR [24]. Histones are particularly problematic ligands for structural determination, as they have a linear N-terminal tail with a high degree of conformational flexibility [25–27] that sticks

out of the nucleosome structure (Figure 1). The protruding N-terminal tails of histones may interact with histone readers (like the proteins in Table S1) or with DNA [28]. Thus, the binding of the N-terminal tail of histone H3 with DNA may compete with the binding of histone N-terminal tails to histone reader proteins [29], which is further supported by the increased accessibility of histone H3 during nucleosome disassembly during transcription [30]. Like all peptides, histones are also extensively hydrated, which further complicates the determination of their interactions [5,31]. Moreover, there are shallow binding pockets on the reader side that result in the histone–reader complexes possessing moderate stability [32,33], with micromolar binding constants (see K_d values in Table S1 for examples). Long peptides such as histone tails are well-known problematic cases for fast computational docking [31,34], due to the inappropriateness of the scoring schemes [35–37] of their binding modes (position, orientation, and conformation) and the lack of explicit water models [38]. The complexes presented in Table S1 are good representatives for investigations of the above structural challenges.

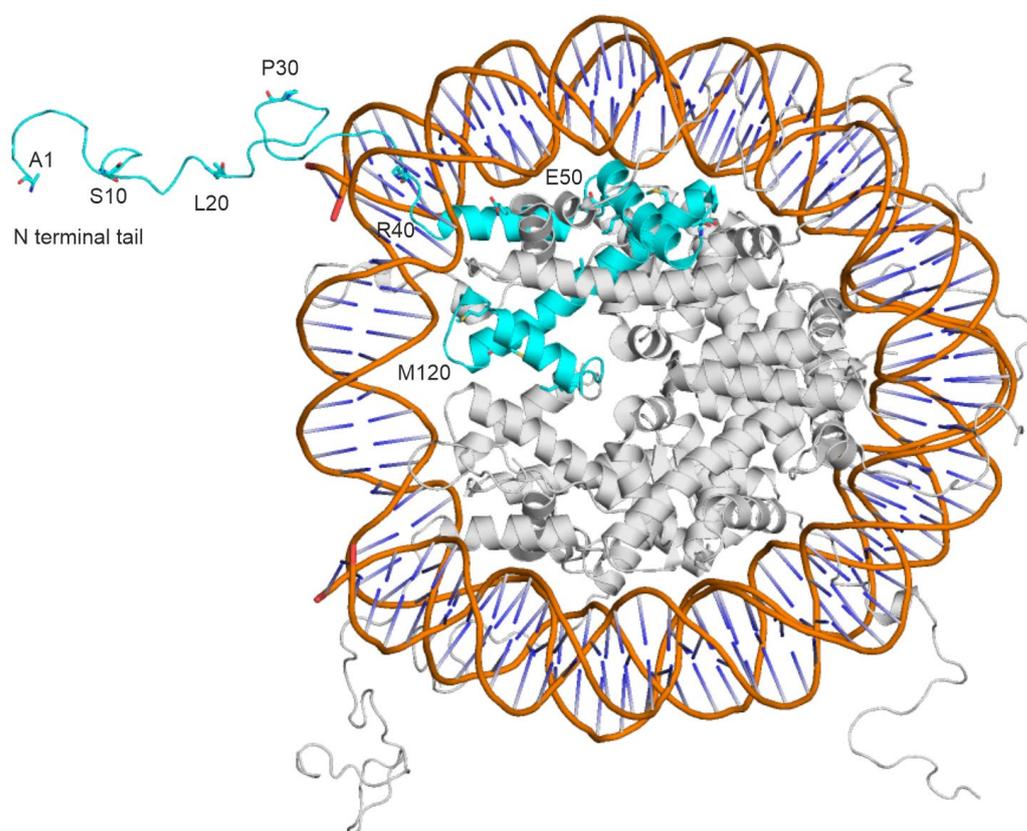


Figure 1. The terminal tails of histone proteins (teal and grey) stick out of the nucleosome core unit and have a flexible structure. The DNA backbone is colored in orange, base pairs are shown as dark blue sticks. A histone H3 protein is highlighted in teal. Every 10th amino acid of the histone H3 (teal) is marked. The figure was prepared from the PDB structure [39] 1kx5 using PyMol v2.0 [40].

The recognition of the above structural and methodological challenges accelerated the development of numerous fast docking methods for peptide ligands. At least three branches can be distinguished among the different methods: physico-chemical approaches, knowledge-based approaches, and their hybrid [41]. Physico-chemical approaches [42–44] calculate energy (scoring) values directly from the atomic positions of the molecules, without conducting further training or experiments. Knowledge-based methods [34,41] are relatively fast and are often restricted by their training set of known structures. Their scores are often based on similarities to the training set [45] and lack physical meaning, which hampers the interpretation of the results (validity problems).

Comprehensive reviews [31,34] and tests [46,47] have shown that the available approaches still have serious limitations with respect to the docking of peptide ligands.

Fragment-based docking is a popular and widely used approach in drug design [48–52], and is based on the linking of docked fragments into the whole bound ligand structure. The number of fragment-based docking methods applied for peptide ligands is still limited. The covalent linking of fragments [48,53] is a critical step in fragment docking, and its success largely depends on the actual steric situation, including the shape-wise matching and the gap between the two docked fragments. Thus, the available methods have multiple limitations, including the lack of full automation, and their dependency on the diversity and selection of linkers and anchoring fragments [54–61]. Inappropriate steric situations of the fragments often necessitate time-consuming follow-up efforts [62] to achieve a new and appropriate covalent bond between two fragments. Further details of the limitations of covalent linking approaches are summarized in Table S2.

In the present study, a new protocol, PepGrow, is introduced and tested for the docking of histone H3 peptide tails to their target reader proteins. PepGrow aims to overcome the limitations of fragment-based docking techniques described above by replacing the fragment linking steps with a growing procedure. Thus, the new protocol is based on the in situ growing of a fragment seed of the peptide ligand in the binding pocket of the reader protein. In drug design, growing steps have been applied for the attachment of small functional groups to ligands [63,64], so as to increase the strength of target–ligand interactions [52]. On the other hand, the growing of a full peptide ligand structure from a small fragment seed is a more difficult task than that handled in the present study. We report the answers to the above challenges, and present a description and validation of PepGrow, comparing its performing with that of ten other docking methods.

2. Results and Discussion

2.1. Histone Systems and Benchmark Methods

Ten complexes of histone H3 peptides and reader proteins (Table S1) of physiological importance, a complete N-terminal end, and available apo forms of the reader proteins were collected from the Protein Data Bank (PDB [65]) as test systems for the development and evaluation of PepGrow. Due to problems regarding their structural determination (see Section 1), there are relatively few complexes in the PDB with histone ligands of a complete N-terminal end, that is, starting with the first amino acid. Notably, the use of apo target structures allowed a truly unbiased test, excluding any help of the ligand-bound conformation of the pre-formed target-binding sites that may be present in the holo structures.

Histone H3 peptides contain up to ca. 50 rotatable bonds (Table S1), that is a challenge of computational docking. The challenges are further increased by the unique binding pattern of histones. Reader proteins often have a shallow binding surface, as in the case of the UHRF1 PHD finger (System 3sou, Figure 2A) [66–68]. A considerable part of the linear [69,70] N-terminal region of histone H3 is not able to find anchor points on this shallow target, tending rather to remain unbound in the bulk (Figure 2A). Quantitative analyses of the per-residue interaction energy (E_{inter} ; see Section 3) distribution of the experimental holo structures in Table S1 show (Figure 2B) that mostly the first five amino acids of the N-terminal of histone H3 are involved in the interaction with the target, while the C-terminal end is exposed to the bulk, and often has a high degree of conformational freedom, which is also reflected by the large atomic B-factors (red in Figure 2A). This finding also emphasizes that only complexes with a full histone tail (i.e., a complete N-terminal end) are useful as test systems.

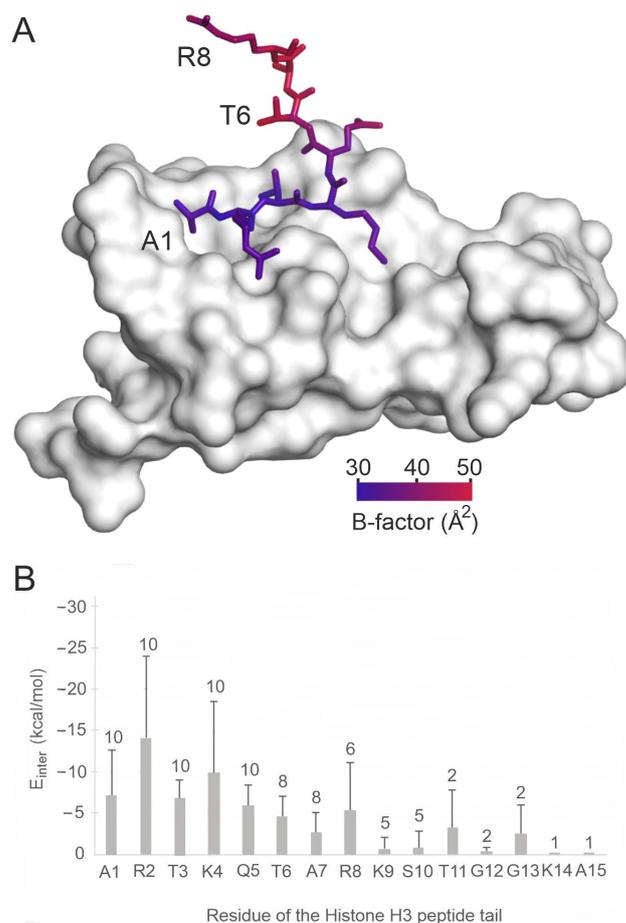


Figure 2. Per-residue energetic and structural analysis of histone H3 peptide ligands bound to their reader proteins. **(A)** The experimental structure of reader UHRF1 PHD finger (grey surface, PDB ID 3sou) in complex with a histone H3 peptide (sticks, colored by C α B-factors). **(B)** The mean (columns) and standard deviations (error bars) of E_{inter} values for the respective residues calculated for the energy-minimized experimental histone complexes presented in Table S1. The numbers on top of the error bars show the number of systems used for calculation of the averages. The numbers are smaller than the maximum of 10 if histone peptides shorter than 15 amino acids in length were measured experimentally.

Besides the PepGrow protocol, a benchmark set of ten available docking methods (Table S3 [41–45,71–77]) was assembled for the present study. Physico-chemical and hybrid (i.e., incorporating knowledge-based elements into their algorithms; see Section 1) methods were included in the benchmark. The same target and ligand structures were used as inputs for the PepGrow and the benchmark methods.

2.2. The PepGrow Protocol

The PepGrow protocol builds the structures of target–peptide complexes at atomic resolutions (Figure 3) without prior knowledge of the binding site residues of the target. PepGrow starts with the selection of a seed molecule that is a fragment of the ligand peptide. As the ligand used in our cases is the same histone H3 tail (Table S1), the selection of an appropriate seed needs to be carried out only once. For the seed selection procedure, the use of only one holo complex structure (2ke1) proved to be sufficient to pick the best dipeptide fragment from among all of the possible dipeptides (Figure 4A) derived from the H3 peptide (Table S4). In the case of histone H3, Fragment 1 (AR) produced the best results (Figure 4B), and therefore, it was selected as the seed for H3 peptide docking for all complexes except for System 2fuu, for which Fragment 4 (KQ) was used. The selection

of Fragments 1 (AR) and 4 (KQ) as seeds is also reflected by the per-residue E_{inter} plot (Figure 2B), where R2 and K4 have the largest E_{inter} contribution among the residues of histone H3. (Thus, the fast, per-residue E_{inter} scoring (Figure 2B) plot of a single strong complex is also applicable for seed selection in PepGrow).

In the next step, the seed was docked on the target protein using a fast method utilizing AutoDock 4.2.6, focusing on the peptide binding area [78], which resulted in several binding modes (where the binding mode refers to the position, orientation, and conformation of a ligand). The binding modes were ranked according to the calculated free energy of their binding and their structural similarity. The representative binding modes were produced for all ranks (see Tables S5 and S6 for a list of the rank counts of all systems). All representative binding modes then proceeded to the fragment growing step, which was accomplished using the builder routine of the homology modeling program Modeller [79]. The experimental target structure with the docked peptide fragment (seed) served as a starting template for growing fragments in the binding pocket. In this way, all docking ranks were used to generate thousands of target–peptide complex models in a matter of minutes, resulting in a large enough pool of peptide binding modes (see Tables S5 and S6 for a list of the binding mode counts of all systems). The complex models of the pool were scored and ranked based on the target–ligand intermolecular interaction energy (E_{inter} , Section 3) values calculated for the full peptide and for the five N-terminal amino acid residues, respectively. The representative peptide structure with coordinates closest to the average coordinates calculated for the peptide structures ranked in the top 1% (Rank 1) according to E_{inter} was selected as the solution. It was observed that in many cases, the top 1% of solution structures contained the best one, but not necessarily the best E_{inter} of all of the structures. Thus, it was reasonable to consider a structure that was representative of the top 1%, rather than a single top structure. Technical details of the PepGrow protocol are provided in the Section 3. Example in- and output files and computational details of the PepGrow protocol are available in the Public Repository files Protocol.pdf and Protocol.tgz (see Data Availability Statement).

2.3. Performance

The structural accuracies of PepGrow and 10 other docking methods are expressed as the root mean square deviation (RMSD; see Section 3) measured between the docked and the experimental (reference) ligand-binding modes. As the experimental complexes mostly show stable (reference) conformations at the first five amino acids of histone H3 (Figure 2), RMSD values were calculated for the full ligand and for the first five amino acids of the N-terminal, respectively. The lowest RMSD of all docked binding modes is referred to as $\text{RMSD}_{\text{best}}$. The statistics (mean and standard deviation) for the $\text{RMSD}_{\text{best}}$ values of docking results to the apo targets for all systems in Table S1 are presented in Figure 5. Due to the high mobility (and structural uncertainty) of peptide ligands outside the binding interface, it is common to use only the interfacial (strongly bound core) amino acids [75] for RMSD calculation. In the case of the histone H3 ligand, this core region corresponds to (see Section 2.1) the first five amino acids (full bars in Figure 5). For comparison, the RMSD values measured for all amino acids (empty bars in Figure 5) of the docked histone H3 ligands are also shown. In general, the $\text{RMSD}_{\text{best}}$ values calculated for the first five amino acids of H3 reflect a much better performance for all methods than the $\text{RMSD}_{\text{best}}$ values calculated for the full ligand (Figure 5A), due to the natural flexibility of the extended C-terminal region described above (Figure 2).

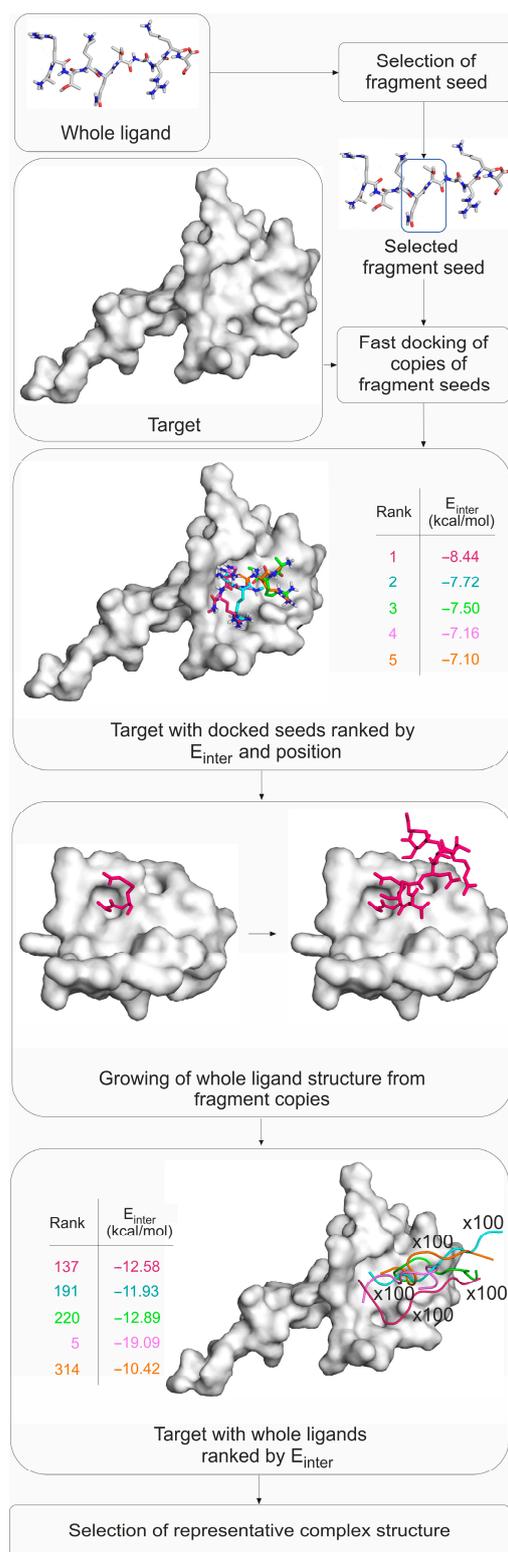


Figure 3. The flow chart of the PepGrow protocol. The different fragment colors correspond to different fragment seed ranks acquired during the fast-docking and seed ranking steps. A close-up of the growth of Rank 1 fragments (purple) only during the growth step is shown for clarity.

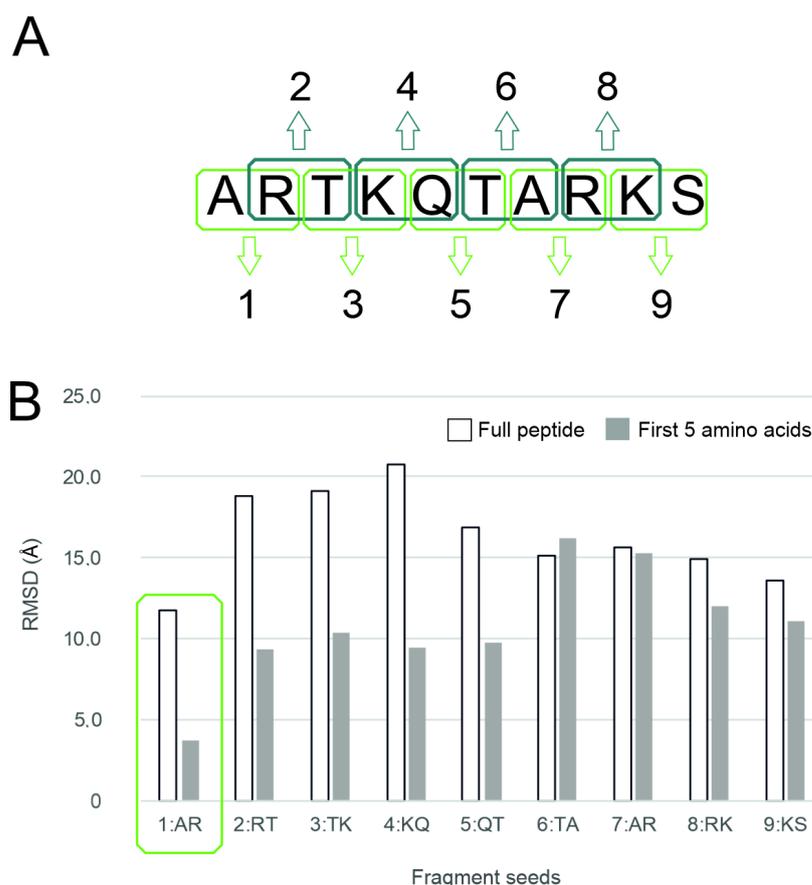


Figure 4. Seed selection. (A) All possible (nine) dipeptide fragments were produced from the histone H3 peptide N terminal sequence. Note that Fragment 1 (AR) was capped with an N-methyl group (-NHMe) at the R residue, and Fragment 7 (AR) was capped with an additional acetyl group at the A residue. The capping of the other fragments (2–9) was performed on both ends. (B) The PepGrow results for each fragment for System 2ke1. The fragment with the lowest $RMSD_{top}$ is marked with a green frame. See Table S4 for details.

The statistics regarding the apo targets show that PepGrow outperformed all of the other fast docking approaches (Figure 5A), with an $RMSD_{best}$ of $5.36 (\pm 1.47)$ Å being calculated for all of the amino acids of the docked histone H3 peptide fragments. Furthermore, PepGrow achieved an excellent $RMSD_{best}$ of $4.09 (\pm 1.18)$ Å, when calculated for the first five amino acids, as well. The per-system analysis of the PepGrow results (Figure 5B) indicates that the best performance was obtained in the case of the target human BAZ2A PHD zinc finger (System 4qf2). Here, the AR-NHMe dipeptide seed was accurately docked (Figure 6), providing a good starting point for ligand growing. The docking of such dipeptides can be accomplished precisely [80] using fast docking techniques. Thus, they provide a good starting point for growing peptide ligands, which is a better alternative than the problematic linking of several, often inadequately docked large-peptide fragments. The accurately docked dipeptide seeds also have the best E_{inter} values (Figure 2), determining the success of PepGrow.

Target flexibility poses a great challenge for docking methods [81]. To check the sensitivity of the investigated docking methods to target conformation, all docking calculations were repeated for the holo structures of the target molecules. As the holo structures have a pre-formed conformation that is ideal for binding to a certain ligand, large differences between the results when docking to the apo and when docking to the holo forms may indicate a high (unwanted) sensitivity to target conformation and moderate robustness of the method. In the case of PepGrow, no significant differences could be detected

(Figure 5 vs. Figure S1) between the results on the apo and holo targets, indicating the robustness of the method.

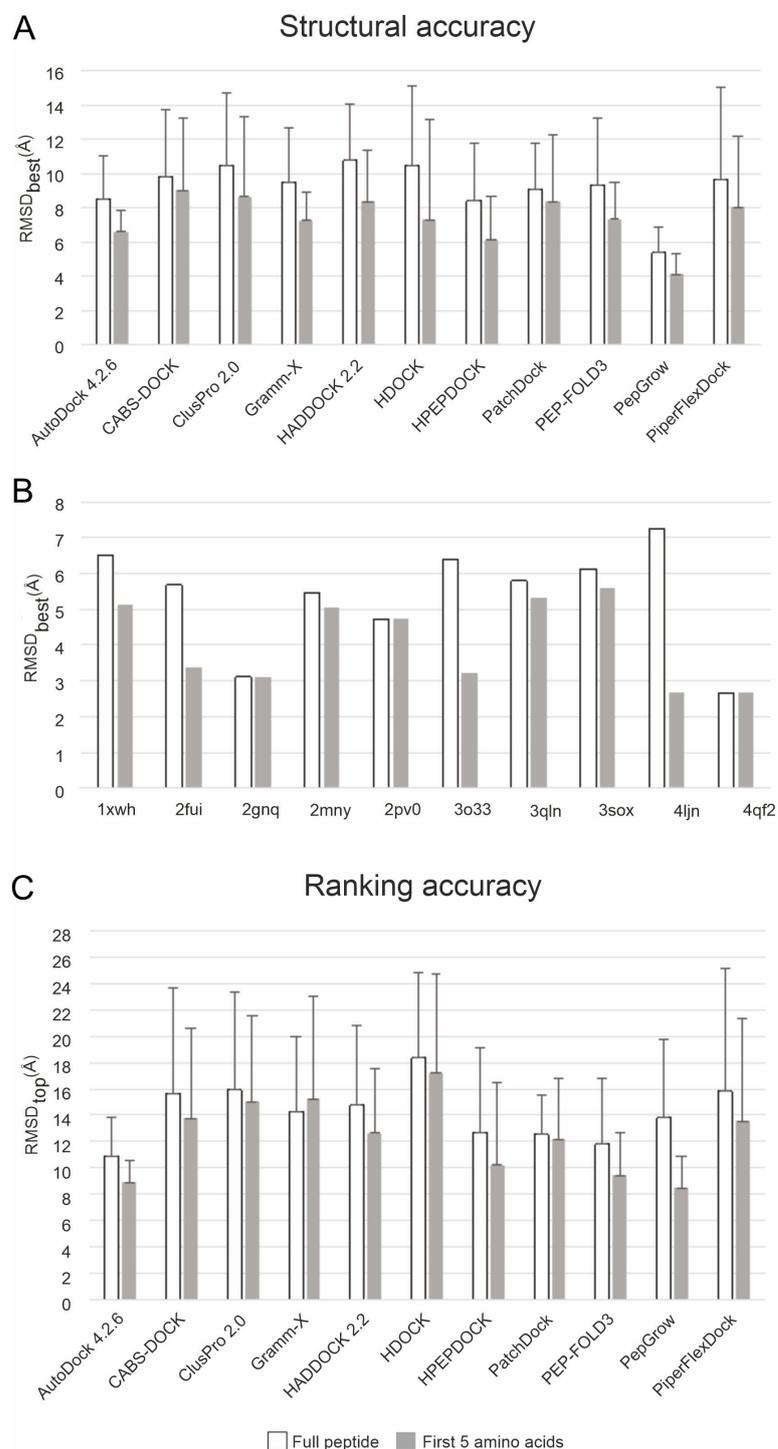


Figure 5. The statistics of docking results obtained for all test systems of Table S1 using all apo target structures. (A) Columns represent the mean $RMSD_{best}$ values (of all test systems) calculated for ligand-binding modes supplied by PepGrow and the 10 benchmark methods. Error bars represent standard deviations (see also Table S7a). (B) Structural performance of PepGrow on the individual test systems (see also Table S5). (C) Columns represent the mean $RMSD_{top}$ values (of all test systems) calculated for ligand-binding modes supplied by PepGrow and the 10 benchmark methods. Error bars represent standard deviations (Table S7a).

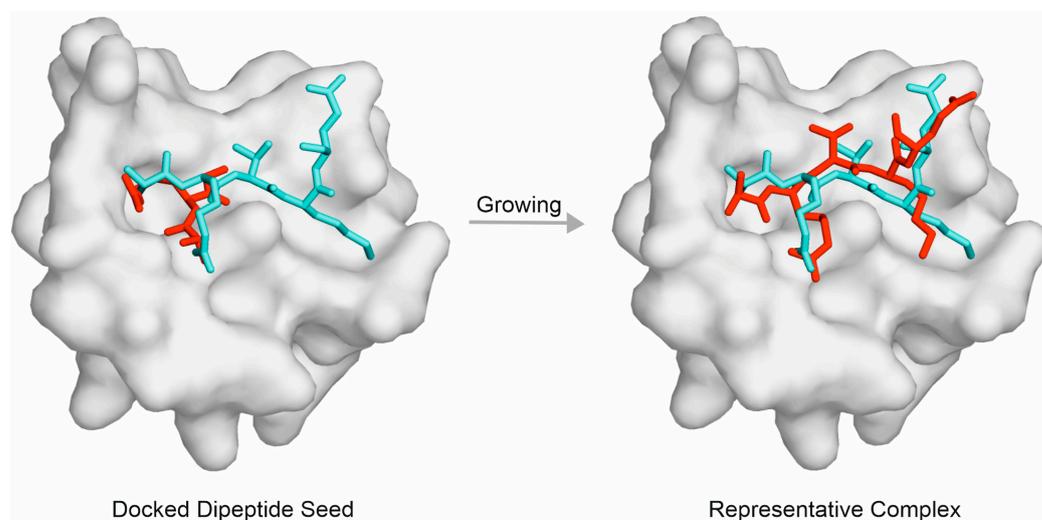


Figure 6. Fragment growing of the fast-docked seed for the complex of the human BAZ2A PHD zinc finger reader (grey surface)–histone H3 peptide (sticks, System 4qf2). The fast-docked seed AR-NHMe of an RMSD of 3.79 Å is shown as red sticks (**left**), representing a good basis of peptide growing. The ligand structure corresponding to an RMSD_{best} of 2.67 Å is shown as red sticks (**right**), representing the results of the growing. The crystallographic ligand-binding mode is shown as teal sticks for comparison.

The acceptable level of RMSD_{best} was concluded to be 4.0 ± 3.0 Å on the basis of data (Table S8) collected from publications related to the benchmark methods (Table S3), in which RMSD was calculated only for the peptide backbone. Notably, side-chain atoms were also included in the RMSD calculations in the present study. Thus, the above performance of PepGrow can be considered to be as good as or above average when compared to the RMSD values produced by the benchmark methods (Figure 5).

Besides the structural accuracy of the methods, their ranking performance was also measured on the basis of their respective RMSD values. The docked-ligand-binding modes were ranked by the default scoring functions of the respective methods (Table S3, Supplementary Materials). The RMSD value of the ligand with the best score (representative of the first rank) is referred to as RMSD_{top}. In the case of a method with perfect scoring and ranking, RMSD_{top} is equal to RMSD_{best} per definitionem. Unfortunately, such an ideal situation was not observed with the methods investigated, as RMSD_{top} considerably exceeded RMSD_{best} in all dockings to the apo targets (Figure 5C), and the same trend was observed in the cases of holo forms (Figure S1). A comparison of the ranking performance of all of the methods (Figure 5C) shows that PepGrow achieved the best results when compared to the benchmark methods. Thus, the E_{inter} -based representative selection method of PepGrow is a viable ranking alternative. Notably, the separate components of E_{inter} (Lennard-Jones and Coulomb terms, respectively) showed a drop in performance (Table S9), and therefore, E_{inter} including both terms (see Section 3) was used in the ranking throughout the present study.

The above results indicate that the structural (Figure 5A) and ranking (Figure 5C) performances of PepGrow are better than/comparable to those of the 10 benchmark methods presented in Table S3. PepGrow can also be considered a physico-chemical method, with energy-based scoring and ranking of the ligand-binding mode (Section 2). In theory, physico-chemical methods are generally applicable for any ligand type with appropriate molecular mechanics parametrization. The efficient sampling of the conformational space of flexible peptide ligands [82] like histone H3 tails is a common problem for all fast docking methods. Knowledge-based and hybrid methods (Table S3) attempt to solve this problem using a training set of experimentally determined structures as templates for achieving the correct bound ligand conformation. However, their performance is limited by the availability and reliability of templates for use in training.

In addition to the above sampling problem, the scoring functions of fast docking methods (Table S3) tend to maximize the interactions of the entire ligand with the target, and therefore cannot handle non-interacting parts (see Section 1). Fragment docking methods may provide a solution for this scoring problem by docking only short fragments instead of the entire ligand. This may be a divide-and-conquer strategy for addressing the limitation of linking fragments (see Section 1). For example, PIPER-FlexPepDock is a fragment-based, hybrid approach in which an ensemble of short peptide fragments is collected from experimentally determined structures with a high degree of sequence and (predicted) secondary structure similarity to the actual ligand. However, such methods are also limited by the lack of structures of peptide fragments of large size and/or unusual conformations. Similar to PIPER-FlexPepDock, PepGrow utilizes the potential of physico-chemical methods to accurately dock small peptide fragments, but instead of all possible fragments in the peptide, it focuses on the anchoring fragment of a good E_{inter} (see Section 2, Figure 2) and grows the remaining part of the peptide in situ in the binding pocket. Thus, PepGrow addresses both the sampling and scoring (ranking) problems via its fragment docking strategy and the focused growing of a ligand from the docked seed the strongest interaction with the target.

Data files of the performance tests of PepGrow and the benchmark methods are available in the Public Repository files `PepGrow.tgz` and `Benchmark.tgz` (see Data Availability Statement).

3. Materials and Methods

3.1. Selection of Test Systems and Benchmark Methods

All atomic coordinates of the targets were acquired from the PDB. Apart from their physiological relevance, histone-target systems were preferentially selected that exhibited high resolution ($<4 \text{ \AA}$) and the availability of a non-covalently bound histone H3 N-terminal peptide tail, starting from the first amino acid (A). The availability of both complexed (holo) and apo forms was a selection criterion, as well. For the benchmark methods, fast docking engines were selected that were designed to model interactions in protein–peptide or macromolecular complexes (except AutoDock) and had previously been evaluated on protein–peptide complexes. A further selection criterion was their free availability for academic purposes via web servers or as standalone programs. The investigated docking engines can be roughly sorted into knowledge-based, physico-chemical, and hybrid categories (Table S3).

3.2. Performance Metrics

Both structural and ranking performance are expressed in terms of root mean square deviation (RMSD), a commonly used measure for the comparison of the conformational match of two molecules. In the present study, the bound conformation of a peptide ligand produced by PepGrow (P) was compared to the bound conformation of the same ligand in the experimental complex (E) structure used as a reference (Equation (1)).

$$RMSD = \sqrt{\frac{1}{N} \sum_{n=1}^N |P_n - E_n|^2} \quad (1)$$

N is the number of ligand heavy atoms, E is the space vector of the n th heavy atom of the experimental reference ligand molecule, and P is the space vector of the n th heavy atom of the PepGrow-calculated ligand conformation. Crystallographic structures were mostly used as references (Table S1). In 3 cases, NMR structures were also employed, where the first model was selected as a reference. RMSD values were calculated after superimposition of the target parts (Table S10).

3.3. Application of Benchmark Methods

The general and specific settings, and the preparation of targets and ligands are detailed for all benchmark methods in the Supplementary Materials Methods [83–89].

3.4. PepGrow

Target preparation. The atomic coordinate structure files for the selected target protein (Table S1) were downloaded from the PDB. All non-protein parts (ligands, waters, etc.) were removed from all selected target structures prior to docking. If the structure was a homo-oligomer, then only one selected chain was used (the first protein chain in the PDB file). The rest of the target molecule was equipped with polar hydrogen atoms and Gasteiger–Marsilli [90] partial charges in AutoDock Tools [44].

Ligand preparation. An initial fragmenting step was used to create dipeptide-sized fragments of the original histone H3 peptide. The fragments were built using the Tinker program package [91] with the protein, newton and xyzpdb commands. The cut was made between the carbon and nitrogen atoms of the amide bond, acetyl (Ac-) and N-methyl (-NHMe) groups were used to block the N- and C-terminal cut ends (the 1:AR fragment was not capped at the N-terminal end, but the 7:AR fragment was capped at both ends). These blocking groups were added in Tinker [91]. The acquired ligand structures were then energy minimized using Open Babel [92] with the Amber99 force field [93] using the steepest descent optimization with 10^4 steps; the convergence threshold was set to 10^3 kJ mol⁻¹ nm⁻¹. The next step was conjugate gradient minimization; a maximum of 10^4 steps was used, and the convergence threshold was set to 10 kJ mol⁻¹ nm⁻¹. Gasteiger–Marsilli charges [90] were added to the fragments with AutoDock Tools [44].

Fragment docking. The fragment docking was performed using AutoDock 4.2.6 [44]. The previously prepared target was handled as a rigid body. All active torsions were allowed on the prepared ligand fragments. All ligand structures were docked to the interacting site defined by the experimental structure, where the docking box was set to a size that would fit the whole peptide inside. The number of grid points was set to $60 \times 60 \times 60$, with a grid spacing of 0.375 Å; the middle of the box was set to the center of the respective experimental full ligand conformation in a manner similar to the procedure used for the benchmark methods. The Lamarckian genetic algorithm was used to perform a global search. Ten docking runs were performed, and the resulting fragment conformations were ranked [94], and representatives of each rank were used in Step 4.

Fragment growing with a homology modelling tool. All docked fragment copies were processed using Modeller 9.22 [79], a homology modeling program. The template structure was the experimental structure of the target protein with the docked (previous step) fragment seed of the ligand peptide. The query sequence was the respective sequence of each system and the histone H3 peptide tail matching the sequence length seen in the corresponding experimental structures (Table S1). The target and ligand sequences were taken from the UniProt database. The alignment between the template structure and the query sequence was manually optimized if necessary to obtain identical regions that correctly matched each other. This was necessary when fitting the sequence of the docked dipeptide seed to the sequence of the whole ligand. The Modeller 9.22 software package was applied to generate 100 models per step, following the final PepGrow protocol. Explicit manual restraints were not added to access additional energy calculation features. During the method development phase of the present work, restraints, energy calculating features, and seed number variation steps were evaluated thoroughly (Table S11). When the rapid generation of 100 models with default building settings was compared with the generation of fewer models (20) with slower refinement, the results were similar, so the faster method (with 100 models) was selected as the main PepGrow protocol step. The robustness of the building procedure was further challenged by changing the random seed number, which did not affect the results (Table S11). For System 2fuu, fragment 4:KQ was selected, due to the special interaction of the trimethylated K4 with the target. In addition, fragment 4:KQ had the second-best performance (after 1:AR) when compared with the other seeds (Figure 4).

Scoring. To extend the use of the method to apo structures with previously unknown N-terminal histone tail ligand positions, it is important to apply a scoring function that is able to select the bound ligand conformation closest to the real structure. The discrete optimized

protein energy (DOPE [79,95]), the Modeller probability density function (molpdf [79,95]), and the Lennard-Jones, Coulomb and E_{inter} interaction energy scores (Equation (2)) of each model were calculated. The E_{inter} interaction energy score calculated for the first five amino acids was the basis of the representative model selection (Tables S7a,b and S12). Table S13 details the scoring functions of the benchmark methods; the differences between the physico-chemical, knowledge-based and hybrid methods were determined based on these scoring functions. Notably, the DOPE and molpdf scores were developed on a benchmark set containing only single-chain proteins, according to the User's Manual of Modeller 9.22 [79]; there is no guarantee of their applicability to multi-chain structures. The calculated DOPE and molpdf scores were therefore only used to test the effect of changing the random seed number for model generation during the initial steps of testing Modeller, as these two scores are the default scoring functions of the software (Table S11).

3.5. Calculation of E_{inter} and Energy Analyses

Experimental, Modeller-built, and energy-minimized experimental structures were subjected to per-residue interaction energy scoring. The missing atoms of all crystallographic targets were modeled using SWISS-Model [96]; for a detailed list of the missing atoms and residues, please see the respective pdb structure files. However, these missing atoms did not affect the binding site. The experimental structures were equipped with polar hydrogen atoms and Gasteiger–Marsilli partial charges [90] using Open Babel 2.4.0 [92], and were converted from pdb files to mol2 files. The mol2 files were then subjected to per-residue interaction energy calculation using Equation (2), implemented in an energy calculator program, which is available as a binary version, downloadable as PepGrow.tgz (see Data Availability Statement). Lennard-Jones and Coulomb energies were calculated and summarized to obtain the total E_{inter} for each residue, and the whole ligand according to Equation (2). The Coulomb term was calculated with a distance-dependent dielectric function of Mehler and Solmajer [97] (Equation (3)), and Amber 2012 van der Waals parameters and atom types were used [98].

$$E_{inter} = E_{LJ} + E_{Coulomb} = \sum_{i,j}^{N_T N_L} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \right)$$

$$A_{ij} = \epsilon_{ij} R_{ij}^{12}$$

$$B_{ij} = 2\epsilon_{ij} R_{ij}^6$$

$$R_{ij} = R_i + R_j$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$
(2)

where ϵ_{ij} is the potential well depth at equilibrium between the *i*th (ligand) and *j*th (target) atoms; ϵ_0 is the permittivity of vacuum; ϵ_r is the distance-dependent relative permittivity (Equation (3)); R_{ij} is the inter-nuclear distance at equilibrium between the *i*th (ligand) and *j*th (target) atoms; q is the partial charge of an atom; r_{ij} is the actual distance between the *i*th (ligand) and *j*th (target) atoms; N_T is the number of target atoms; N_L is the number of ligand atoms.

$$\epsilon_r = A + \frac{B}{1 + ke^{-\lambda Br}}$$
(3)

where $B = \epsilon_0 - A$, ϵ_0 is the dielectric constant of water at 25 °C, and A , λ and k are constants [97].

4. Conclusions

Although fast docking methods have proven successful in the design of small-molecule ligands [99,100], they face persistent challenges [99–102]. While long peptides are often used as templates for the development of new drugs [103–105], they are especially challenging ligands due to their high degree of flexibility and hydration, which cannot properly be handled by fast docking methods. In the present study, a popular fast docking method,

AutoDock 4.2.6, and the fast model building function of the widely used program Modeller were combined into a new protocol PepGrow.

A comparison of the results with those obtained using ten other benchmark methods showed that PepGrow offers a real alternative for the construction of histone complexes. The relatively good performance of PepGrow is based on at least two key components of the algorithm. Firstly, the docking of very short and strongly interacting (di)peptide seeds can be reliably achieved [80] using currently available fast docking methods like AutoDock 4.2.6 (unlike large peptide ligands, where fast docking presents problems [31]). Secondly, instead of the problematic linking step of all fragments of the ligand, a robust ligand growing step is implemented.

PepGrow constructs the complex structures of histone H3 peptides of various lengths with various targets. While the number of such complexes is expected to be very high (histone code), only a small number of structures have been determined. Thus, PepGrow can help to accelerate the structural exploration of the histone code, as well as the prediction of the outcome of the reader–DNA binding competition mentioned in the Introduction. The disordered nature of histone peptides presented a real challenge for all eleven methods compared. The structural performance of PepGrow was better than that of the other methods, the ranking of such large ligands still remains [34,37] a challenging task for all methods. Our results also indicate that physico-chemical scores like E_{inter} are a necessary component of the ranking and selection of representative structures. The histone complexes selected for the present work can be recommended as a particularly challenging test set for future method development studies.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms241813831/s1>.

Author Contributions: Conceptualization, C.H.; writing—original draft preparation, B.Z.Z., B.B., R.B. and C.H.; writing—review and editing, B.Z.Z., B.B., R.B. and C.H.; visualization, B.Z.Z., B.B., R.B. and C.H.; supervision, C.H.; project administration, B.Z.Z. and C.H.; funding acquisition, B.Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the PTE ÁOK KA-2022-26 grant. The project was supported by the European Union, co-financed by the European Social Fund. Project name and code: Comprehensive Development for Implementing Smart Specialization Strategies at the University of Pécs, EFOP-3.6.1-16-2016-00004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data files are available at <https://zenodo.org/record/8251745> (access date: 20 August 2023). A compressed data file PepGrow.tgz contains the in- and output files of PepGrow for both the holo and apo systems, the scripts and the programs necessary to produce them, and a Table of Contents file with detailed folder descriptions. A portable document file Protocol.pdf contains detailed description how to perform the PepGrow protocol. A compressed data file Protocol.tgz contains example folders of PepGrow on a system with in- and output files, scripts and the programs necessary to produce them, and a Table of Contents file with detailed folder descriptions. A compressed data file Benchmark.tgz contains the in- and output files of the 10 benchmark methods for both holo and apo systems, and a Table of Contents file with detailed folder descriptions.

Acknowledgments: We acknowledge the support from the Governmental Information Technology Development Agency, Hungary. We acknowledge that the results of this research were achieved using the DECI resource Archer2, based in the UK at the National Supercomputing Service, with support from the PRACE aisbl.

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

References

1. Shvedunova, M.; Akhtar, A. Modulation of Cellular Processes by Histone and Non-Histone Protein Acetylation. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 329–349. [[CrossRef](#)]
2. Enetics, E.P.I.G.; Gamblin, S.J.; Wilson, J.O.N.R. A Key to Unlock Chromatin. *Nature* **2019**, *573*, 355–356.
3. Izzo, L.T.; Wellen, K.E. Histone Lactylation Links Metabolism and Gene Regulation. *Nature* **2019**, *574*, 492–493. [[CrossRef](#)] [[PubMed](#)]
4. Org, T.; Chignola, F.; Hetényi, C.; Gaetani, M.; Rebane, A.; Liiv, I.; Maran, U.; Mollica, L.; Bottomley, M.J.; Musco, G.; et al. The Autoimmune Regulator PHD Finger Binds to Non-Methylated Histone H3K4 to Activate Gene Expression. *EMBO Rep.* **2008**, *9*, 370–376. [[CrossRef](#)] [[PubMed](#)]
5. Zsidó, B.Z.; Hetényi, C. Molecular Structure, Binding Affinity, and Biological Activity in the Epigenome. *Int. J. Mol. Sci.* **2020**, *21*, 4134. [[CrossRef](#)]
6. Strahl, B.D.; Allis, C.D. The Language of Covalent Histone Modifications. *Nature* **2000**, *403*, 41–45. [[CrossRef](#)] [[PubMed](#)]
7. Musselman, C.A.; Lalonde, M.E.; Côté, J.; Kutateladze, T.G. Perceiving the Epigenetic Landscape through Histone Readers. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1218–1227. [[CrossRef](#)]
8. Arrowsmith, C.H.; Bountra, C.; Fish, P.V.; Lee, K.; Schapira, M. Epigenetic Protein Families: A New Frontier for Drug Discovery. *Nat. Rev. Drug Discov.* **2012**, *11*, 384–400. [[CrossRef](#)]
9. Bortoluzzi, A.; Amato, A.; Lucas, X.; Blank, M.; Ciulli, A. Structural Basis of Molecular Recognition of Helical Histone H3 Tail by PHD Finger Domains. *Biochem. J.* **2017**, *474*, 1633–1651. [[CrossRef](#)]
10. Ruthenburg, A.J.; Wang, W.; Graybosch, D.M.; Li, H.; Allis, C.D.; Patel, D.J.; Verdine, G.L. Histone H3 Recognition and Presentation by the WDR5 Module of the MLL1 Complex. *Nat. Struct. Mol. Biol.* **2006**, *13*, 704–712. [[CrossRef](#)]
11. Ooi, S.K.T.; Qiu, C.; Bernstein, E.; Li, K.; Jia, D.; Yang, Z.; Erdjument-Bromage, H.; Tempst, P.; Lin, S.P.; Allis, C.D.; et al. DNMT3L Connects Unmethylated Lysine 4 of Histone H3 to de Novo Methylation of DNA. *Nature* **2007**, *448*, 714–717. [[CrossRef](#)] [[PubMed](#)]
12. Iwase, S.; Xiang, B.; Ghosh, S.; Ren, T.; Lewis, P.W.; Cochrane, J.C.; Allis, C.D.; Picketts, D.J.; Patel, D.J.; Li, H.; et al. ATRX ADD Domain Links an Atypical Histone Methylation Recognition Mechanism to Human Mental-Retardation Syndrome. *Nat. Struct. Mol. Biol.* **2011**, *18*, 769–776. [[CrossRef](#)] [[PubMed](#)]
13. Rajakumara, E.; Wang, Z.; Ma, H.; Hu, L.; Chen, H.; Lin, Y.; Guo, R.; Wu, F.; Li, H.; Lan, F.; et al. PHD Finger Recognition of Unmodified Histone H3R2 Links UHRF1 to Regulation of Euchromatic Gene Expression. *Mol. Cell* **2011**, *43*, 275–284. [[CrossRef](#)] [[PubMed](#)]
14. Tsai, W.W.; Wang, Z.; Yiu, T.T.; Akdemir, K.C.; Xia, W.; Winter, S.; Tsai, C.Y.; Shi, X.; Schwarzer, D.; Plunkett, W.; et al. TRIM24 Links a Non-Canonical Histone Signature to Breast Cancer. *Nature* **2010**, *468*, 927–932. [[CrossRef](#)]
15. Chignola, F.; Gaetani, M.; Rebane, A.; Org, T.; Mollica, L.; Zucchelli, C.; Spitaleri, A.; Mannella, V.; Peterson, P.; Musco, G. The Solution Structure of the First PHD Finger of Autoimmune Regulator in Complex with Non-Modified Histone H3 Tail Reveals the Antagonistic Role of H3R2 Methylation. *Nucleic Acids Res.* **2009**, *37*, 2951–2961. [[CrossRef](#)]
16. Zhang, Y.; Yang, H.; Guo, X.; Rong, N.; Song, Y.; Xu, Y.; Lan, W.; Zhang, X.; Liu, M.; Xu, Y.; et al. The PHD1 Finger of KDM5B Recognizes Unmodified H3K4 during the Demethylation of Histone H3K4me2/3 by KDM5B. *Protein Cell* **2014**, *5*, 837–850. [[CrossRef](#)] [[PubMed](#)]
17. Li, H.; Ilin, S.; Wang, W.; Duncan, E.M.; Wysocka, J.; Allis, C.D.; Patel, D.J. Molecular Basis for Site-Specific Read-out of Histone H3K4me3 by the BPTF PHD Finger of NURF. *Nature* **2006**, *442*, 91–95. [[CrossRef](#)] [[PubMed](#)]
18. Dreveny, I.; Deeves, S.E.; Fulton, J.; Yue, B.; Messmer, M.; Bhattacharya, A.; Collins, H.M.; Heery, D.M. The Double PHD Finger Domain of MOZ/MYST3 Induces α -Helical Structure of the Histone H3 Tail to Facilitate Acetylation and Methylation Sampling and Modification. *Nucleic Acids Res.* **2014**, *42*, 822–835. [[CrossRef](#)]
19. Sanchez, R.; Meslamani, J.; Zhou, M.-M. The Bromodomain: From Epigenome Reader to Druggable Target. *Biochim. Biophys. Acta BBA Gene Regul. Mech.* **2014**, *1839*, 676–685. [[CrossRef](#)]
20. Li, X.; Yao, Y.; Wu, F.; Song, Y. A Proteolysis-Targeting Chimera Molecule Selectively Degrades ENL and Inhibits Malignant Gene Expression and Tumor Growth. *J. Hematol. Oncol.* **2022**, *15*, 41. [[CrossRef](#)]
21. Mosca, R.; Céol, A.; Aloy, P. Interactome3D: Adding Structural Details to Protein Networks. *Nat. Methods* **2013**, *10*, 47–53. [[CrossRef](#)]
22. Srivastava, A.; Nagai, T.; Srivastava, A.; Miyashita, O.; Tama, F. Role of Computational Methods in Going beyond X-Ray Crystallography to Explore Protein Structure and Dynamics. *Int. J. Mol. Sci.* **2018**, *19*, 3401. [[CrossRef](#)]
23. Frank, J. Electron Microscopy Applied to Molecular Machines. *Biopolymers* **2013**, *99*, 832–836. [[CrossRef](#)]
24. Verardi, R.; Traaseth, N.J.; Masterson, L.R.; Vostrikov, V.V.; Veglia, G. *Isotope Labeling in Biomolecular NMR.*; Springer: Dordrecht, The Netherlands, 2012; Volume 992, ISBN 978-94-007-4953-5.
25. Antunes, D.A.; Devaurs, D.; Kavraki, L.E. Understanding the Challenges of Protein Flexibility in Drug Design. *Expert Opin. Drug Discov.* **2015**, *10*, 1301–1313. [[CrossRef](#)]
26. Du, X.; Li, Y.; Xia, Y.L.; Ai, S.M.; Liang, J.; Sang, P.; Ji, X.L.; Liu, S.Q. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17*, 144. [[CrossRef](#)]
27. Hauser, A.S.; Windshügel, B. LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. *J. Chem. Inf. Model.* **2016**, *56*, 188–200. [[CrossRef](#)] [[PubMed](#)]

28. Lehmann, K.; Felekyan, S.; Kühnemuth, R.; Dimura, M.; Tóth, K.; Seidel, C.A.M.; Langowski, J. Dynamics of the Nucleosomal Histone H3 N-Terminal Tail Revealed by High Precision Single-Molecule FRET. *Nucleic Acids Res.* **2020**, *48*, 1551–1571. [[CrossRef](#)] [[PubMed](#)]
29. Morrison, E.A.; Bowerman, S.; Sylvers, K.L.; Wereszczynski, J.; Musselman, C.A. The Conformation of the Histone H3 Tail Inhibits Association of the BPTF PHD Finger with the Nucleosome. *eLife* **2018**, *7*, e31481. [[CrossRef](#)]
30. Morrison, E.A.; Baweja, L.; Poirier, M.G.; Wereszczynski, J.; Musselman, C.A. Nucleosome Composition Regulates the Histone H3 Tail Conformational Ensemble and Accessibility. *Nucleic Acids Res.* **2021**, *49*, 4750–4767. [[CrossRef](#)] [[PubMed](#)]
31. Rentzsch, R.; Renard, B.Y. Docking Small Peptides Remains a Great Challenge: An Assessment Using AutoDock Vina. *Brief. Bioinform.* **2015**, *16*, 1045–1056. [[CrossRef](#)] [[PubMed](#)]
32. Peach, C.J.; Mignone, V.W.; Arruda, M.A.; Alcobia, D.C.; Hill, S.J.; Kilpatrick, L.E.; Woolard, J. Molecular Pharmacology of VEGF-A Isoforms: Binding and Signalling at VEGFR2. *Int. J. Mol. Sci.* **2018**, *19*, 1264. [[CrossRef](#)]
33. Weaver, T.M.; Morrison, E.A.; Musselman, C.A. Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains. *Molecules* **2018**, *23*, 2614. [[CrossRef](#)] [[PubMed](#)]
34. Ciemny, M.; Kurcinski, M.; Kamel, K.; Kolinski, A.; Alam, N.; Schueler-Furman, O.; Kmiecik, S. Protein–Peptide Docking: Opportunities and Challenges. *Drug Discov. Today* **2018**, *23*, 1530–1537. [[CrossRef](#)] [[PubMed](#)]
35. Lee, A.C.L.; Harris, J.L.; Khanna, K.K.; Hong, J.H. A Comprehensive Review on Current Advances in Peptide Drug Development and Design. *Int. J. Mol. Sci.* **2019**, *20*, 2383. [[CrossRef](#)] [[PubMed](#)]
36. Peterson, L.X.; Roy, A.; Christoffer, C.; Terashi, G.; Kihara, D. Modeling Disordered Protein Interactions from Biophysical Principles. *PLoS Comput. Biol.* **2017**, *13*, e1005485. [[CrossRef](#)]
37. Xiong, G.-L.; Ye, W.-L.; Shen, C.; Lu, A.-P.; Hou, T.-J.; Cao, D.-S. Improving Structure-Based Virtual Screening Performance via Learning from Scoring Function Components. *Brief. Bioinform.* **2021**, *22*, bbaa094. [[CrossRef](#)]
38. Zsidó, B.Z.; Hetényi, C. The Role of Water in Ligand Binding. *Curr. Opin. Struct. Biol.* **2021**, *67*, 1–8. [[CrossRef](#)]
39. Richmond, T.J.; Davey, C.A. The Structure of DNA in the Nucleosome Core. *Nature* **2003**, *423*, 145–150. [[CrossRef](#)]
40. DeLano, W.L. *The PyMOL Molecular Graphics System, Version 2.0*; Schrödinger, LLC.: New York, NY, USA, 2021.
41. Dominguez, C.; Boelens, R.; Bonvin, A.M.J.J. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. [[CrossRef](#)]
42. Alam, N.; Goldstein, O.; Xia, B.; Porter, K.A.; Kozakov, D.; Schueler-Furman, O. High-Resolution Global Peptide-Protein Docking Using Fragments-Based PIPER-FlexPepDock. *PLoS Comput. Biol.* **2017**, *13*, e1005905. [[CrossRef](#)]
43. Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-Dock Web Server for the Flexible Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Nucleic Acids Res.* **2015**, *43*, W419–W424. [[CrossRef](#)]
44. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [[CrossRef](#)] [[PubMed](#)]
45. Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tufféry, P. PEP-FOLD3: Faster de Novo Structure Prediction for Linear Peptides in Solution and in Complex. *Nucleic Acids Res.* **2016**, *44*, W449–W454. [[CrossRef](#)]
46. Castro-Alvarez, A.; Costa, A.M.; Vilarrasa, J. The Performance of Several Docking Programs at Reproducing Protein-Macrolide-like Crystal Structures. *Molecules* **2017**, *22*, 136. [[CrossRef](#)] [[PubMed](#)]
47. Hetényi, C.; Körtvélyesi, T.; Penke, B. Mapping of Possible Binding Sequences of Two Beta-Sheet Breaker Peptides on Beta Amyloid Peptide of Alzheimer’s Disease. *Bioorg. Med. Chem.* **2002**, *10*, 1587–1593. [[CrossRef](#)] [[PubMed](#)]
48. Bian, Y.; Xie, X.Q. Computational Fragment-Based Drug Design: Current Trends, Strategies, and Applications. *AAPS J.* **2019**, *20*, 59. [[CrossRef](#)]
49. Evans, D.J.; Yovanno, R.A.; Rahman, S.; Cao, D.W.; Beckett, M.Q.; Patel, M.H.; Bandak, A.F.; Lau, A.Y. Finding Druggable Sites in Proteins Using TACTICS. *J. Chem. Inf. Model.* **2021**, *61*, 2897–2910. [[CrossRef](#)]
50. Aguayo-Ortiz, R.; Dominguez, L. Unveiling the Possible Oryzalin-Binding Site in the α -Tubulin of *Toxoplasma Gondii*. *ACS Omega* **2022**, *7*, 18434–18442. [[CrossRef](#)]
51. Aguayo-Ortiz, R.; Guzmán-Ocampo, D.C.; Dominguez, L. Insights into the Binding of Morin to Human Γ D-Crystallin. *Biophys. Chem.* **2022**, *282*, 106750. [[CrossRef](#)]
52. Pires, D.E.V.; Portelli, S.; Rezende, P.M.; Veloso, W.N.P.; Xavier, J.S.; Karmakar, M.; Myung, Y.; Linhares, J.P.V.; Rodrigues, C.H.M.; Silk, M.; et al. A Comprehensive Computational Platform to Guide Drug Development Using Graph-Based Signature Methods. *Methods Mol. Biol.* **2020**, *2112*, 91–106.
53. Lamoree, B.; Hubbard, R.E. Current Perspectives in Fragment-Based Lead Discovery (FBLD). *Essays Biochem.* **2017**, *61*, 453–464. [[CrossRef](#)]
54. de Beauchene, I.C.; de Vries, S.J.; Zacharias, M. Binding Site Identification and Flexible Docking of Single Stranded RNA to Proteins Using a Fragment-Based Approach. *PLoS Comput. Biol.* **2016**, *12*, e1004697. [[CrossRef](#)]
55. Liao, J.M.; Wang, Y.T.; Lin, C.L.S. A Fragment-Based Docking Simulation for Investigating Peptide-Protein Bindings. *Phys. Chem. Chem. Phys.* **2017**, *19*, 10436–10442. [[CrossRef](#)] [[PubMed](#)]
56. Budin, N.; Majeux, N.; Cafilisch, A. Fragment-Based Flexible Ligand Docking by Evolutionary Optimization. *Biol. Chem.* **2001**, *382*, 1365–1372. [[CrossRef](#)]
57. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S.B.; Johnson, A.P. EHiTS: A New Fast, Exhaustive Flexible Ligand Docking System. *J. Mol. Graph. Model.* **2007**, *26*, 198–212. [[CrossRef](#)]

58. Thompson, D.C.; Denny, R.A.; Nilakantan, R.; Humblet, C.; Joseph-McCarthy, D.; Feyfant, E. CONFIRM: Connecting Fragments Found in Receptor Molecules. *J. Comput. Aided Mol. Des.* **2008**, *22*, 761–772. [[CrossRef](#)]
59. Samsonov, S.A.; Zacharias, M.; de Chauvot Beauchene, I. Modeling Large Protein–Glycosaminoglycan Complexes Using a Fragment-Based Approach. *J. Comput. Chem.* **2019**, *40*, 1429–1439. [[CrossRef](#)] [[PubMed](#)]
60. Cross, S.S.J. Improved FlexX Docking Using FlexS-Determined Base Fragment Placement. *J. Chem. Inf. Model.* **2005**, *45*, 993–1001. [[CrossRef](#)] [[PubMed](#)]
61. Bálint, M.; Horváth, I.; Mészáros, N.; Hetényi, C. Towards Unraveling the Histone Code by Fragment Blind Docking. *Int. J. Mol. Sci.* **2019**, *20*, 422. [[CrossRef](#)]
62. Hoffer, L.; Muller, C.; Roche, P.; Morelli, X. Chemistry-Driven Hit-to-Lead Optimization Guided by Structure-Based Approaches. *Mol. Inform.* **2018**, *37*, e1800059. [[CrossRef](#)]
63. Yuan, Y.; Pei, J.; Lai, L. LigBuilder V3: A Multi-Target de Novo Drug Design Approach. *Front. Chem.* **2020**, *8*, 142. [[CrossRef](#)] [[PubMed](#)]
64. Perez, C.; Soler, D.; Soliva, R.; Guallar, V. FragPELE: Dynamic Ligand Growing within a Binding Site. A Novel Tool for Hit-To-Lead Drug Design. *J. Chem. Inf. Model.* **2020**, *60*, 1728–1736. [[CrossRef](#)]
65. Berman, H.M.; Battistuz, T.; Bhat, T.N.; Bluhm, W.F.; Bourne, P.E.; Burkhardt, K.; Feng, Z.; Gilliland, G.L.; Iype, L.; Jain, S.; et al. The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 899–907. [[CrossRef](#)] [[PubMed](#)]
66. Bálint, M.; Jeszenoi, N.; Horváth, I.; Van Der Spoel, D.; Hetényi, C. Systematic Exploration of Multiple Drug Binding Sites. *J. Cheminform.* **2017**, *9*, 65. [[CrossRef](#)] [[PubMed](#)]
67. Jeszenoi, N.; Bálint, M.; Horváth, I.; Van Der Spoel, D.; Hetényi, C. Exploration of Interfacial Hydration Networks of Target-Ligand Complexes. *J. Chem. Inf. Model.* **2016**, *56*, 148–158. [[CrossRef](#)]
68. Zhao, S.; Yang, M.; Zhou, W.; Zhang, B.; Cheng, Z.; Huang, J.; Zhang, M.; Wang, Z.; Wang, R.; Chen, Z.; et al. Kinetic and High-Throughput Profiling of Epigenetic Interactions by 3D-Carbene Chip-Based Surface Plasmon Resonance Imaging Technology. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7245–E7254. [[CrossRef](#)] [[PubMed](#)]
69. Van Roey, K.; Uyar, B.; Weatheritt, R.J.; Dinkel, H.; Seiler, M.; Budd, A.; Gibson, T.J.; Davey, N.E. Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev.* **2014**, *114*, 6733–6778. [[CrossRef](#)]
70. Davis, A.M.; Teague, S.J.; Kleywegt, G.J. Application and Limitations of X-Ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem. Int. Ed.* **2003**, *42*, 2718–2736. [[CrossRef](#)]
71. Kozakov, D.; Hall, D.R.; Xia, B.; Porter, K.A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein–Protein Docking. *Nat. Protoc.* **2017**, *12*, 255–278. [[CrossRef](#)]
72. Tovchigrechko, A.; Vakser, I.A. GRAMM-X Public Web Server for Protein–Protein Docking. *Nucleic Acids Res.* **2006**, *34*, W310–W314. [[CrossRef](#)]
73. van Zundert, G.C.P.; Rodrigues, J.P.G.L.M.; Trellet, M.; Schmitz, C.; Kastiris, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; de Vries, S.J.; Bonvin, A.M.J.J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725. [[CrossRef](#)] [[PubMed](#)]
74. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: Servers for Rigid and Symmetric Docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367. [[CrossRef](#)] [[PubMed](#)]
75. Zhou, P.; Jin, B.; Li, H.; Huang, S.-Y. HPEPDOCK: A Web Server for Blind Peptide–Protein Docking Based on a Hierarchical Algorithm. *Nucleic Acids Res.* **2018**, *46*, W443–W450. [[CrossRef](#)] [[PubMed](#)]
76. Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S.-Y. HDOCK: A Web Server for Protein–Protein and Protein–DNA/RNA Docking Based on a Hybrid Strategy. *Nucleic Acids Res.* **2017**, *45*, W365–W373. [[CrossRef](#)]
77. Kozakov, D.; Brenke, R.; Comeau, S.R.; Vajda, S. PIPER: An FFT-Based Protein Docking Program with Pairwise Potentials. *Proteins Struct. Funct. Bioinform.* **2006**, *65*, 392–406. [[CrossRef](#)] [[PubMed](#)]
78. Wang, P.; Wu, R.; Jia, Y.; Tang, P.; Wei, B.; Zhang, Q.; Wang, V.Y.-F.; Yan, R. Inhibition and Structure-Activity Relationship of Dietary Flavones against Three Loop 1-Type Human Gut Microbial β -Glucuronidases. *Int. J. Biol. Macromol.* **2022**, *220*, 1532–1544. [[CrossRef](#)]
79. Fiser, A.; Do, R.K.G.; Šali, A. Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, *9*, 1753–1773. [[CrossRef](#)]
80. Hetényi, C.; van der Spoel, D. Efficient Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Protein Sci.* **2009**, *11*, 1729–1737. [[CrossRef](#)]
81. Basciu, A.; Callea, L.; Motta, S.; Bonvin, A.M.J.J.; Bonati, L.; Vargiu, A.V. No Dance, No Partner! A Tale of Receptor Flexibility in Docking and Virtual Screening. *Annu. Rep. Med. Chem.* **2022**, *59*, 43–97.
82. Li, C.; Sun, J.; Li, L.-W.; Wu, X.; Palade, V. An Effective Swarm Intelligence Optimization Algorithm for Flexible Ligand Docking. *IEEE ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 2672–2684. [[CrossRef](#)]
83. The UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [[CrossRef](#)] [[PubMed](#)]
84. Kolinski, A. Protein Modeling and Structure Prediction with a Reduced Representation. *Acta Biochim. Pol.* **2004**, *51*, 349–371. [[CrossRef](#)]
85. Huang, S.-Y.; Zou, X. MDockPP: A Hierarchical Approach for Protein–Protein Docking and Its Application to CAPRI Rounds 15–19. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 3096–3103. [[CrossRef](#)] [[PubMed](#)]

86. Yan, Y.; Zhang, D.; Huang, S.-Y. Efficient Conformational Ensemble Generation of Protein-Bound Peptides. *J. Cheminform.* **2017**, *9*, 59. [[CrossRef](#)]
87. Huang, S.-Y.; Zou, X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins Struct. Funct. Bioinform.* **2006**, *66*, 399–421. [[CrossRef](#)]
88. Gront, D.; Kulp, D.W.; Vernon, R.M.; Strauss, C.E.M.; Baker, D. Generalized Fragment Picking in Rosetta: Design, Protocols and Applications. *PLoS ONE* **2011**, *6*, e23294. [[CrossRef](#)]
89. Raveh, B.; London, N.; Schueler-Furman, O. Sub-Angstrom Modeling of Complexes between Flexible Peptides and Globular Proteins. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 2029–2040. [[CrossRef](#)]
90. Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228. [[CrossRef](#)]
91. Rackers, J.A.; Wang, Z.; Lu, C.; Laury, M.L.; Lagardère, L.; Schnieders, M.J.; Piquemal, J.-P.; Ren, P.; Ponder, J.W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289. [[CrossRef](#)]
92. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3*, 33. [[CrossRef](#)]
93. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [[CrossRef](#)]
94. Hetényi, C.; Van Der Spoel, D. Blind Docking of Drug-Sized Compounds to Proteins with up to a Thousand Residues. *FEBS Lett.* **2006**, *580*, 1447–1450. [[CrossRef](#)]
95. Shen, M.; Sali, A. Statistical Potential for Assessment and Prediction of Protein Structures. *Protein Sci.* **2006**, *15*, 2507–2524. [[CrossRef](#)]
96. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [[CrossRef](#)]
97. Mehler, E.L.; Solmajer, T. Electrostatic Effects in Proteins: Comparison of Dielectric and Charge Models. *Protein Eng. Des. Sel.* **1991**, *4*, 903–910. [[CrossRef](#)] [[PubMed](#)]
98. Wang, J.; Cieplak, P.; Li, J.; Cai, Q.; Hsieh, M.-J.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations. 4. van Der Waals Parametrization. *J. Phys. Chem. B* **2012**, *116*, 7088–7101. [[CrossRef](#)]
99. Ferreira, L.; dos Santos, R.; Oliva, G.; Andricopulo, A. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, *20*, 13384–13421. [[CrossRef](#)] [[PubMed](#)]
100. Vitoria, M.; Granich, R.; Gilks, C.F.; Gunneberg, C.; Hosseini, M.; Were, W.; Raviglione, M.; De Cock, K.M. The Global Fight Against HIV/AIDS, Tuberculosis, and Malaria. *Am. J. Clin. Pathol.* **2009**, *131*, 844–848. [[CrossRef](#)] [[PubMed](#)]
101. Torres, P.H.M.; Sodero, A.C.R.; Jofily, P.; Silva, F.P., Jr. Key Topics in Molecular Docking for Drug Design. *Int. J. Mol. Sci.* **2019**, *20*, 4574. [[CrossRef](#)]
102. de Ruyck, J.; Brysbaert, G.; Blossey, R.; Lensink, M. Molecular Docking as a Popular Tool in Drug Design, an in Silico Travel. *Adv. Appl. Bioinform. Chem.* **2016**, *9*, 1–11. [[CrossRef](#)]
103. Schreiber, G.; Fleishman, S.J. Computational Design of Protein–Protein Interactions. *Curr. Opin. Struct. Biol.* **2013**, *23*, 903–910. [[CrossRef](#)] [[PubMed](#)]
104. Grosdidier, S.; Fernandez-Recio, J. Protein-Protein Docking and Hot-Spot Prediction for Drug Discovery. *Curr. Pharm. Des.* **2012**, *18*, 4607–4618. [[CrossRef](#)] [[PubMed](#)]
105. Bienstock, R.J. Computational Drug Design Targeting Protein-Protein Interactions. *Curr. Pharm. Des.* **2012**, *18*, 1240–1254. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.