

Article

Solubility-Aware Protein Binding Peptide Design Using AlphaFold

Takatsugu Kosugi  and Masahito Ohue * 

Department of Computer Science, School of Computing, Tokyo Institute of Technology, G3-56-4259 Nagatsutacho, Midori-ku, Yokohama City 226-8501, Kanagawa, Japan; kosugi@li.c.titech.ac.jp

* Correspondence: ohue@c.titech.ac.jp; Tel.: +81-45-924-5522

Abstract: New protein–protein interactions (PPIs) are identified, but PPIs have different physicochemical properties compared with conventional targets, making it difficult to use small molecules. Peptides offer a new modality to target PPIs, but designing appropriate peptide sequences by computation is challenging. Recently, AlphaFold and RoseTTAFold have made it possible to predict protein structures from amino acid sequences with ultra-high accuracy, enabling de novo protein design. We designed peptides likely to have PPI as the target protein using the “binder hallucination” protocol of AfDesign, a de novo protein design method using AlphaFold. However, the solubility of the peptides tended to be low. Therefore, we designed a solubility loss function using solubility indices for amino acids and developed a solubility-aware AfDesign binder hallucination protocol. The peptide solubility in sequences designed using the new protocol increased with the weight of the solubility loss function; moreover, they captured the characteristics of the solubility indices. Moreover, the new protocol sequences tended to have higher affinity than random or single residue substitution sequences when evaluated by docking binding affinity. Our approach shows that it is possible to design peptide sequences that can bind to the interface of PPI while controlling solubility.

Keywords: solubility; peptide design; AlphaFold; AfDesign



Citation: Kosugi, T.; Ohue, M.

Solubility-Aware Protein Binding

Peptide Design Using AlphaFold.

Biomedicines **2022**, *10*, 1626. <https://doi.org/10.3390/biomedicines10071626>

Academic Editors: William D. Lubell, Jackie Wilce and Johan Rosengren

Received: 13 May 2022

Accepted: 5 July 2022

Published: 7 July 2022

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



Copyright: © 2022 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

Protein–protein interactions (PPIs) are associated with various biological processes, such as signal transduction and metabolism, and play a fundamental role in many cellular activities [1]. Approximately 810,000 human non-redundant PPIs have been identified [2] (BioGrid 4.4.208—April 2022). These PPIs regulate many important intracellular pathways in human diseases and have attracted attention as potential drug targets since the early 2000s [3–7]. There have been several successful small-molecule drug designs targeting PPIs, including the Federal Drug Administration (FDA)-approved PPI inhibitor Venetoclax [8], which is considered a BH3 mimetic, and others in the clinical stage [9,10].

However, the physicochemical properties of PPI drug targets are significantly different from those of conventional drug targets, and PPI-targeting drug discovery remains a severe challenge [9,11]. One of the most important reasons for the difficulty in PPI-targeting drug development is shallow and large PPI interfaces. Small molecule binding sites are relatively deep and have an area of only approximately 500 Å², while PPI-binding interfaces are typically flat and wide, with a surface area of approximately 1000–4000 Å² [12–14]. Another difficulty in PPI-targeting drug discovery is that small molecule drugs tend to bind to the deep folding pockets of proteins rather than the large, flat, and, hydrophobic binding interface of the PPI [15]. Antibodies are an effective modality for recognizing PPI interfaces; although they are effective for targeting receptors on the cell membrane surface, they are not effective against intracellular PPI targets because they have difficulty penetrating the cell membrane [16].

In recent years, peptides with a balance of flexibility and binding affinity have attracted attention as a modality for targeting PPIs. Research on therapeutic peptides began with natural human hormones such as insulin, vasopressin, oxytocin, and gonadotropin-releasing hormone. To date, approximately 80 peptide drugs have been approved, with approximately 30 non-insulinic peptide drugs approved since 2000. More than 170 peptides have been in clinical development. As of 2021, there are 58 therapeutic peptides targeting PPIs in clinical stages, of which 13 are in Phase 1, 26 are in Phase 2, 15 are in Phase 3, and 4 have New Drug Applications pending and are close to approval [17,18].

There are several methods for the experimental design of PPI-targeting peptides, such as phage display, mRNA display, and other methods for screening large numbers of peptides that interact with proteins experimentally, but the cost is very high [19–21]. Therefore, a method is needed to computationally design candidate peptide sequences that interact with target proteins. While there are examples of designing sequences that have a function, such as the design of antimicrobial peptide sequences [22,23], a significant amount of experimental data is needed, and it is difficult to obtain a large amount of data on peptide sequences that interact with specific proteins targeting PPIs. There is also a method called PDGA that uses genetic algorithms to generate peptide sequences with similar small molecule footprints, but this method also requires at least one example of a small molecule compound that is the target of a specific protein, and the footprints do not consider actual 3D structure [24]. In general, protein designs are designed using the three-dimensional (3D) structure of a protein that binds to a specific site on the surface of the target protein [25–28]. For a more efficient design of PPI-targeted peptides, it is important to utilize conformational information on the target protein and the PPI-binding interface.

Recently, protein structure prediction using deep neural networks, which estimate 3D structures from the co-evolution of amino acid residues, has been widely used [29,30]. It is also said that AlphaFold can be used for modeling peptide structures smaller than 60 amino acids if the target peptide has a well-defined secondary structure and does not have multiple turned regions or flexible regions that may take different conformations [31]. Furthermore, it has been shown that docking of proteins and peptides is possible using AlphaFold [32]. It has recently become possible to predict sequences from protein structures. For example, it was reported that Monte Carlo sampling in amino acid sequence space was used to design sequences with specific backbone structures from random sequences. Moreover, a pipeline has been developed to design small proteins that bind to target proteins based on the results of many experiments [33,34]. Similarly, methods have been published to predict the backbone of a specific 3D structure or the sequence of a complex structure of a target protein from a random sequence using all-atom geometry and AlphaFold prediction accuracy indicators, such as the predicted aligned error (PAE) and predicted local-distance difference test (pLDDT) [35,36]. Protein design has become possible using such deep learning networks.

AfDesign implements the “binder hallucination” method, which is based on the AlphaFold peptide docking method and the new protein design method “deep network hallucination” described above. However, our examination of peptide design using AfDesign binder hallucination showed that the sequences to be designed tend to have more aromatic ring and hydrophobic residues on the interaction surface, and the logS, an index of water solubility of the designed peptide sequences, tends to be relatively low.

The development of strategies to improve membrane permeability and facilitate cellular uptake will be essential for successfully targeting PPIs, because peptides have low cell membrane permeability [37]. One method to improve bioavailability is through water solubility. The higher the water solubility, the easier it is to maintain effective serum concentrations; thus, bioavailability is usually greatly enhanced. High solubility is also one of the requirements for successful recombinant protein production and is an important industrial factor [38]. Although water solubility is an optimization issue in the design of therapeutic peptides, experimentally identifying unnecessary hydrophobic amino acids and replacing them with charged or polar residues adjusts the isoelectric point while

maintaining biological activity, which is said to be primarily an empirical process [39,40]. Recently, there have been many attempts to computationally solve the empirical process by using machine learning to predict protein solubility [41].

There are several possible strategies for peptide design to target PPIs considering solubility.

- A library of highly water-soluble peptide sequences is created, and then docking scores and binding affinities are predicted by protein–peptide docking.
- Design peptide sequences that are likely to bind to target proteins using peptide sequence prediction methods such as AfDesign, and then evaluate water solubility and filter out those that exceed water solubility thresholds.

The first strategy is that the number of sequences in the amino acid sequence of N residues has a huge pattern of 20^N , and even if we set some constraints based on the characteristics of the amino acid residues in some way, it is computationally expensive and impractical to search for the target peptide from the large number of candidate peptides by docking calculations. The second strategy is that, as mentioned above, owing to the characteristics of AfDesign, binder hallucination tends to increase the number of aromatic ring residues and hydrophobic residues on the interaction surface; therefore, after design, there is a possibility that few candidate sequences will remain when filtered using water solubility indices such as logS. The reason why the above strategy does not work well may be because peptide design and solubility optimization are not performed at the same time.

Therefore, this research focused on the possibility of controlling solubility while maintaining or improving the binding affinity of the designed peptide by simultaneously optimizing the binder peptide sequence of the target protein and optimizing solubility by designing the generation process of AfDesign with a loss of solubility constraint.

2. Materials and Methods

2.1. AfDesign Settings

We used the AfDesign binder hallucination protocol to calculate the binder of the MDM2 protein. Protein Data Bank (PDB) 1YCR chain A was used as the MDM2 structural template for AfDesign. binder_len is the model of the p53 peptide in 1YCR chain B. The length was set to 13, which is the same as the length of the sequence “ETFSDLWKLLEN”. The length of the sequence was set to 17 when AfDesign was applied with PD-1 as the target protein. The design method used was design_3stage(); soft_iter, temp_iter, and hard_iter were set to 100, 100, and 10, respectively. Unless otherwise noted, other settings were left at default. AfDesign ran 100 times for all three solubility indices with 100 different settings of seed from 1 to 100 for each weight (0.1, 0.3, 0.5, 0.7, and 1.0) when added to the other weights in AfDesign. The same was performed for the conditions without a solubility index. The versions of jax and jaxlib used in this study were 0.3.1 and 0.3.0+cuda11.cudnn805, respectively. To ensure reproducibility, TF_CUDNN_DETERMINISTIC=1 was set so that JAX can behave deterministically. While AfDesign was updated during our study, the commit history we used was that of the following: ColabDesign repository—<https://github.com/sokrypton/ColabDesign/tree/be242491a2517589c96f09b7d15b7bb37f72bafa> (accessed on 14 March 2022); af_backprop repository—https://github.com/sokrypton/af_backprop/tree/7246fe544e9398de8dab848b11b7b634f16858db (accessed on 14 March 2022). The AlphaFold model used in AfDesign was downloaded from https://storage.googleapis.com/alphafold/alphafold_params_2021-07-14.tar (accessed on 14 March 2022).

2.2. Solubility Loss Calculation

Three solubility indices, the Hydrophobicity Index [42], Hydropathy Index [43], and Solubility-Weighted Index (SWI) [41], were used in this study (Supplementary Table S1). The Hydrophobicity Index evaluates hydrophobicity based on the physical characteristics of 20 amino acids to identify regions of a protein’s primary sequence that are likely to be buried in the membrane [42]. The Hydropathy Index is a hydrophilicity scale that considers the hydrophilicity and hydrophobicity of each of the 20 amino acid side chains and was

developed based on experimental observations from the literature. Specifically, values were calculated using both the water vapor transfer free energy and the distribution in and out of the amino acid side chains as determined by Chothia (1976) [43]. The Solubility-Weighted Index is a predictive index of solubility, and prediction programs using it are superior to many existing de novo protein solubility prediction tools [41]. In this study, the weight of this predictive index was used as the solubility index.

The three solubility indices were normalized to have a maximum of 1 and a minimum of 0. Owing to the meanings of the indices, the normalized Hydrophobicity Index and Hydrophobicity Index were used as they are in the solubility loss function. For normalized SWI, values inverted (minus 1) after normalization were used for the solubility loss function. All modified index values are shown in Supplementary Table S2.

The solubility loss is the geometric mean of the designed amino acids and the corresponding normalized solubility index values. A solubility loss was added to the other losses specified in AfDesign:

$$\text{Solubility loss}(seq) = \frac{1}{N} \sum_{i=1}^N f(s_i) \quad (1)$$

where seq represents the designed amino acid sequence and $seq = (s_1s_2 \dots s_N)$, s_i represents an amino acid residue in the i -th position of seq , N is the sequence length of the `binder_len` to be designed, f represents the amino acid index, and $f(s)$ is the index value of the amino acid residue s .

2.3. Calculation of Solubility

As a preprocessing step, SMILES were prepared by acetylation of the N-terminus of the peptide sequence and amidation of the C-terminus using `RDKit.Chem.MolFromHELM()`, and conformers were prepared by `LigPrep` in the Schrödinger software suite (version 2021-1). Finally, $\log S$ values were calculated as solubility using `QikProp` in the Schrödinger software suite (version 2021-1). If multiple conformers generated from a single peptide sequence, the $\log S$ values were averaged for each peptide sequence.

2.4. Calculation of Protein–Peptide Binding Affinity

Using `AutoDock CrankPep` within `AutoDockFR` software suite version 1.0 [44], 1YCR chain A as a receptor, 1YCR chain B, and the designed peptide as a peptide were calculated. The procedure is as described in the official document. First, the PDB files prepared as receptor and peptide were protonated by `reduce` version 3.23.130521 [45], and then the protonated PDB files were converted to PDBQT files by using two scripts, `prepare_ligand` and `prepare_receptor`, of the `AutoDockFR` software suite. In addition, a docking box was placed and sized over the receptor pocket into which the peptide was to be docked and a `trg` file was calculated using `agfr` to compute an affinity map for the list of atom types in `AutoDock 4` [46]. Up to 2.5 million evaluations of the scoring function were performed to increase the likelihood of finding the best docking pose (the global minimum of the scoring function) and 50 independent searches were performed for each. The argument of `AutoDock CrankPep` is `-N 50 -n 2500000`.

2.5. Creation of Sequence Logos

The sequence with the lowest final loss in AfDesign in the latter 10 iterations (`hard_iter`) `design_3stage()` was selected as the best sequence. Sequence logos were created on the `WebLogo 3` server (<http://weblogo.threepiusone.com/create.cgi> (accessed on 26 April 2022)) [47] with the 100 best sequences per solubility index and per weight as input.

2.6. Competitive Peptide Binding Predictions Using AlphaFold

This method was based on previously published paper [48]. Modeling was performed in the publicly available `localcolabfold` (<https://github.com/YoshitakaMo/localcolabfold>

(accessed on 29 June 2022)). To predict competitive peptide binding of MDM2 to p53 and the sequence with the highest affinity among the 90 sequences remaining after filtering (see Section 4), we first used DEVYYWYYHLEND:...:ETFSDLWKLLEN (“...” was an MDM2 sequence but is omitted here) as input to colab_batch. The MSA search option was specified as MMseqs2 (UniRef + Environmental). The resulting MSA file (a3m file) generated only MDM2 MSA and only queried the sequence itself for both peptide sequences. The a3m file thus obtained was used as input for 20 competitive peptide binding predictions. For these predictions, AlphaFold2-ptm was specified in the model-type option and no template was used. For each competitive peptide binding modeling, predictions were run 20 times (five predictions each) with the random-seed option changed from 1 to 20. Thus, 100 predictions were accumulated.

As described in the referenced paper, one of the peptides matches the experimental structure whereas the other is far from the binding site; therefore, from all the predictions obtained, the RMSD of each peptide was measured using PyMOL as follows. For the complex of DEVYYWYYHLEND and MDM2, as no experimental structure was available in this case, we used the predicted top rank results as the pseudo-native structure using localcolabfold. AlphaFold2-ptm was specified as the model-type option for prediction.

First, the MDM2-p53 complex (PDB ID: 1YCR) (this is native MDM2 and native p53) was fetched. Next, MDM2 of the complex of DEVYYWYYHLEND and MDM2 predicted using localcolabfold was aligned with native MDM2 (this determines the native position of each peptide.) MDM2 from all competitive peptide binding prediction results were then aligned with native MDM2. In this condition, the all atom (heavy atom) RMSD between native p53 and the p53 for competitive peptide binding predictions, the all atom (heavy atom) RMSD between pseudo-native competitor peptide and the competitor peptide for competitive peptide binding predictions, and the C_{α} RMSD between native p53 and the competitor peptide for competitive peptide binding predictions were calculated using rms_cur.

2.7. Interatomic Interactions of MDM2 and Peptide

The online web tool PLIP [49] (<https://plip-tool.biotec.tu-dresden.de> (accessed on 30 June 2022)) was used to analyze the atomic interactions between the designed peptide and MDM2. To detect interactions between MDM2 and peptide, select the inter/peptide mode and enter the chain ID of the peptide in advanced options. The structure of the designed sequence is the top ranked one predicted by localcolabfold, which was used in the competitive peptide binding prediction.

3. Results

3.1. Binder Design Targeting PPI Using AfDesign

Our approach to solubility-aware protein-binding peptide design was performed using AfDesign binder hallucination. AfDesign is a public module created by Dr. Sergey Ovchinnikov [36], and it is a tool for designing protein-binding peptides based on Dr. Sergey's previous work with trRosetta [33,50]. AlphaFold is used as the structure prediction oracle, the optimization target is the binding sequence, and the loss function is defined as a flexible function of AlphaFold confidences. It returns pLDDT, PAE, and distogram, among which pLDDT measures the quality of the local model for each residue, and PAE indicates the confidence level of each amino acid residue pair, and distogram indicates the distance prediction probability of the C_{β} of an amino acid residue pair. The objective of this study was to investigate whether a new loss function with a solubility measure could be added to the original AfDesign loss function to optimize the solubility of the sequence that is designed. Binding affinity was measured using AutoDock CrankPep, which was highly rated in a benchmarking evaluation of protein-peptide docking tools [51]. and further validated by measuring logS by QikProp as a measure of solubility. A schematic diagram of the entire process is shown in Figure 1.

The principle of the AfDesign binder design is that the PDB file of the target protein is used as a template, and the PDB file is matched to the AlphaFold input. In addition, the “residue index” of the last residue of the target protein and the starting residue of the complex peptide are sufficiently spaced to create a “chain break” state [52]. This is a trick to predict complexes using AlphaFold and is also used in ColabFold [53]. The “chain break” peptide is a random sequence with a specified length and an initial sequence. Finally, a loss function is created using the confidence index output from the AlphaFold model, which is then used as a backprop to optimize the “chain break” peptide sequence as a complex [36]. We used the MDM2-p53 complex (PDB ID: 1YCR) in this research as representative example of a PPI protein complex. The target protein was MDM2, and the original binder was p53 peptide.

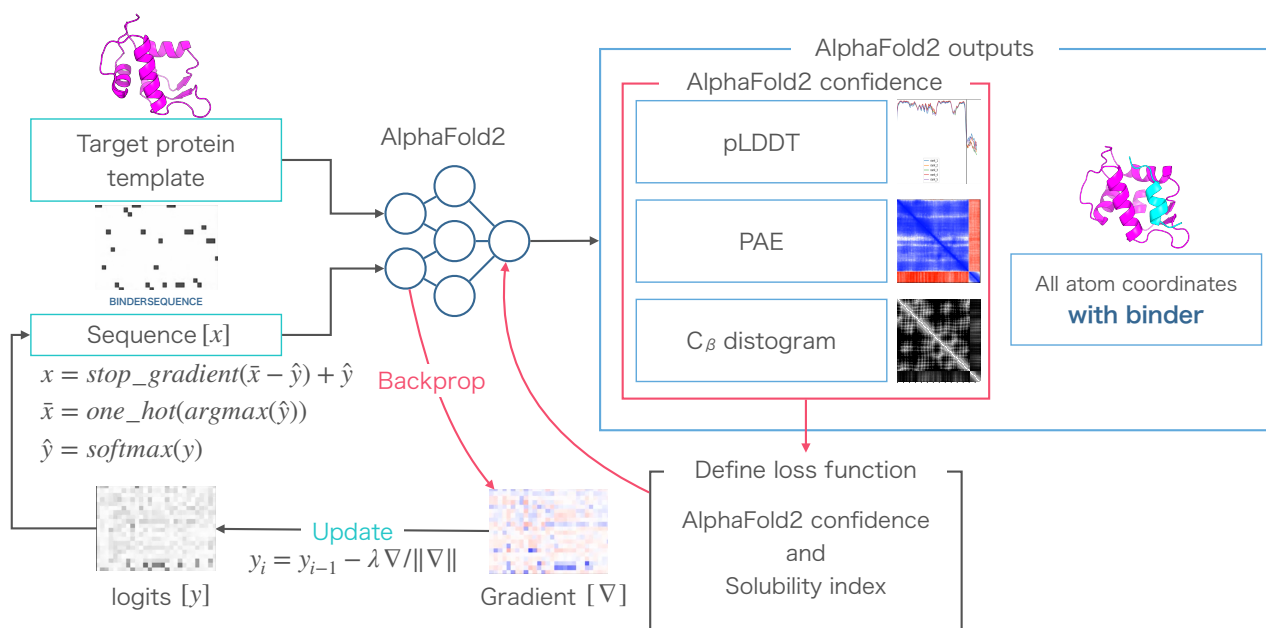


Figure 1. Schematic of the optimization method of the AfDesign binder hallucination. The output of AlphaFold is an all-atom coordinate, two reliability indices (predicted aligned error (PAE) and predicted local-distance difference test (pLDDT)), and a distogram. These are used to define the loss function, which is back-propagated to compute the gradient for the design sequence and then updated and predicted in a loop for optimization. To evaluate the optimized sequence, the binding affinity is calculated by docking and logS is calculated as a peptide solubility index.

To evaluate the distribution of solubility of the binder sequences designed in the AfDesign binder hallucination protocol, 100 designs were performed using MDM2 as the target protein and different seed settings in AfDesign. The logS of the p53 peptide sequence, a known binding peptide of MDM2, and the binder sequence designed in the AfDesign binder hallucination protocol were calculated by Schrödinger’s QikProp as an index of solubility. As shown in Figure 2a, the distribution of logS of the binder sequence designed in the AfDesign binder hallucination protocol was much lower than that of the p53 peptide sequence, indicating that the binder tends to be designed with lower solubility. To confirm the amino acid composition of the designed sequence, the sequence logos were generated from 100 designed sequences using WebLogo [47]. Figure 2b shows that the binder sequences designed in the AfDesign binder hallucination protocol contained more hydrophobic amino acid leucine and aromatic amino acids tyrosine, phenylalanine, and tryptophan in residues 4 to 11. In general, the binding interface of PPI was consistent with high hydrophobicity [54]. Comparing the designed sequence with the sequence of the p53 peptide, “ETFSDLWKLLPEN”, the composition ratio of hydrophobic amino acids tended to be higher and solubility tended to be lower.

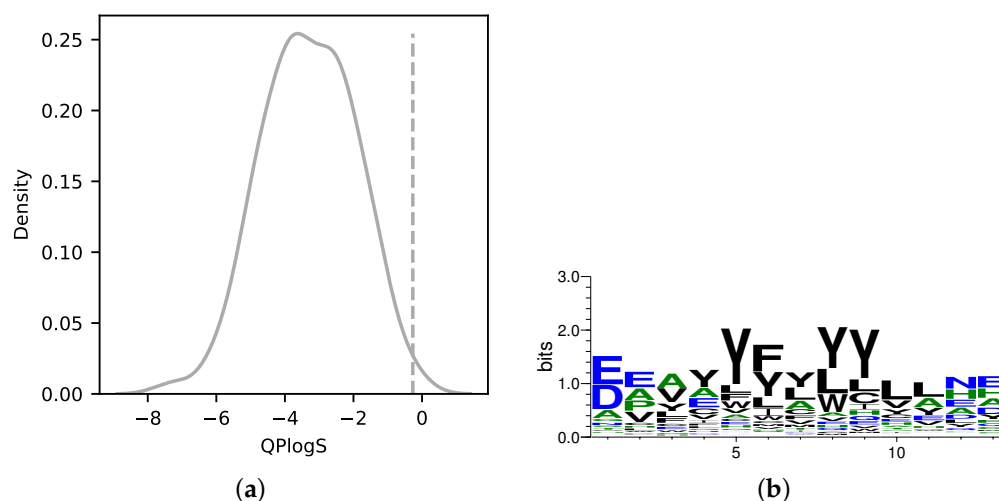


Figure 2. Distribution of $\log S$ of MDM2 binder sequences designed by AfDesign and sequence logo of the binder sequences. Distribution of $\log S$ calculated by QikProp. Dashed lines are $\log S$ values for the p53 peptide (a). Sequence logo for sequences designed with the AfDesign binder hallucination protocol (b). Hydrophilic amino acids (R, K, D, E, N, and Q) are shown in blue, neutral amino acids (S, G, H, T, A, and P) are shown in green, and hydrophobic amino acids (Y, V, M, C, L, F, I, and W) are shown in black.

From the results in Figure 2b, the $\log S$ distribution of the binder sequence designed with the default AfDesign binder hallucination protocol was relatively low compared with that of the p53 peptide. Therefore, in order to test whether it is possible to control water solubility from the bioavailability and industrial perspective of therapeutic peptides by adding solubility constraint as a new loss to the design process of the AfDesign binder sequence, we simultaneously optimized the binder peptide sequence of the target protein, MDM2, and optimized the solubility. Based on previous studies on amino acid and protein solubility, three solubility indices were used to verify the results:

- Hydrophobicity Index [42];
- Hydropathy Index [43];
- Solubility-Weighted Index [41].

The different index values related to solubility were each normalized and added as solubility loss to the loss function for the AfDesign binder hallucination protocol, and the weights of the loss were weighted in the range of 0.0–1.0 to design the binders. The following is a summary of the results.

3.2. Solubility-Aware Binder Design Using AfDesign

As shown in Figure 3, the sequence designed using SWI for solubility loss had the lowest degree of change in solubility with weight among the three indices, while the sequence designed using the Hydrophobicity Index as the solubility index had the highest degree of change in solubility with weight among the three indices. The sequence logo of the sequence designed with the AfDesign binder hallucination protocol considering solubility loss is shown in Figure 4. As Figure 4 shows, compared with the Hydropathy Index, aromatic amino acids tyrosine and phenylalanine gradually decreased in sequence frequency with weight in the Hydrophobicity Index and SWI. This is consistent with the greater weight of tyrosine and phenylalanine. Moreover, compared with the Hydrophobicity Index and SWI, the Hydropathy Index showed a gradual decrease in sequence frequency of leucine, a hydrophobic amino acid, along with its weight. This is consistent with the greater weight of leucine (see Supplementary Table S1). The addition of solubility loss to the AfDesign binder hallucination protocol resulted in an increase in solubility, while the sequences were designed according to the characteristics of each solubility index.

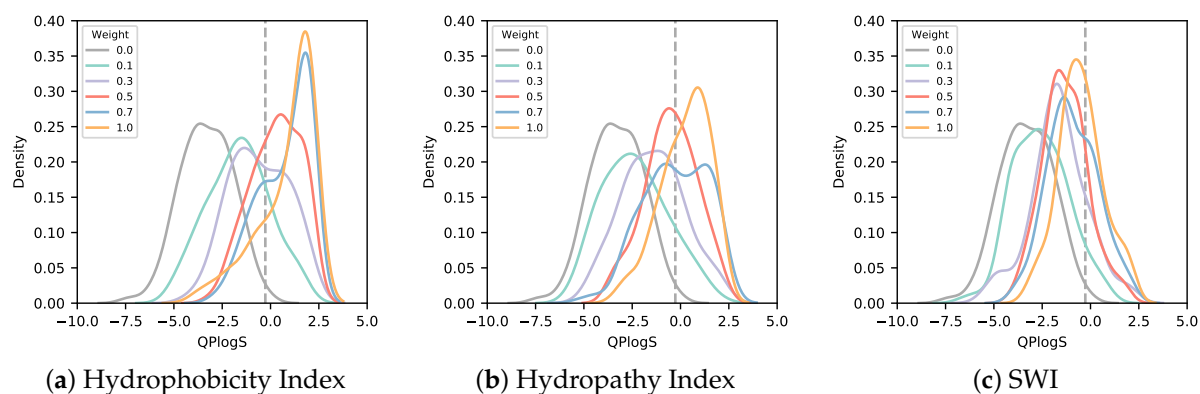


Figure 3. Distribution of $\log S$ for various weights of sequences designed in AfDesign using three solubility indices. The $\log S$ of sequences designed in AfDesign using the Hydrophobicity Index (a), Hydropathy Index (b), and Solubility-Weighted Index (c) as solubility indices in the loss function. Each color indicates the weight of the solubility loss. The gray line with a weight of 0.0 shows the distribution of $\log S$ for the sequence designed by AfDesign without a solubility index. The dashed line shows the $\log S$ of the p53 peptide.

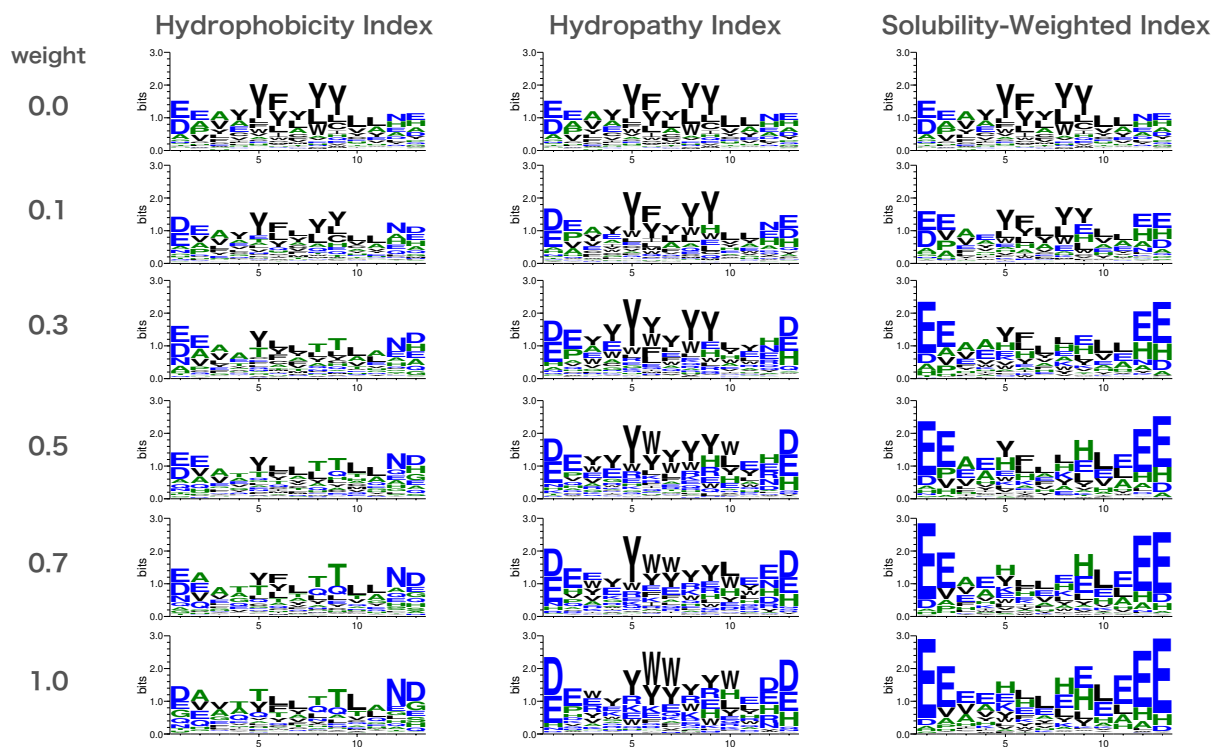


Figure 4. Sequence logos designed by AfDesign binder hallucination with solubility loss for each weight parameter. Hydrophilic amino acids (R, K, D, E, N, and Q) are shown in blue, neutral amino acids (S, G, H, T, A, and P) are shown in green, and hydrophobic amino acids (Y, V, M, C, L, F, I, and W) are shown in black.

To assess the binding affinity of sequences designed with the AfDesign binder hallucination protocol with loss of solubility to the target protein, MDM2, we calculated the binding affinity of the designed peptide to the target protein using AutoDock CrankPep. As shown in Figure 5, the only sequence designed using the Hydrophobicity Index had a lower binding affinity to the target protein than the sequence designed without solubility loss. The binding free energy tended to be higher (weaker as a value of binding). Sequences designed using the Hydropathy Index tended to maintain or increase binding affinity compared with sequences designed without solubility loss.

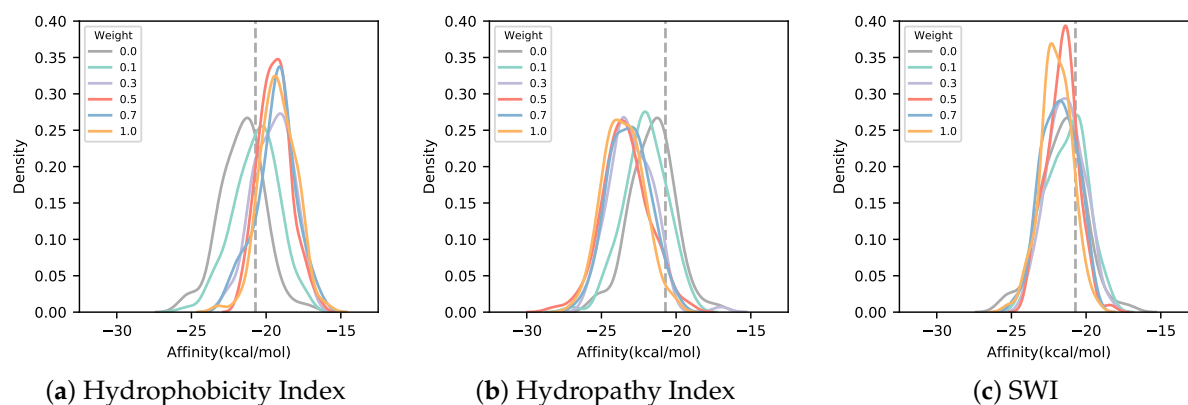


Figure 5. Distribution of binding affinity for various weights of sequences designed in AfDesign using three solubility indices. The binding affinity of sequences designed in AfDesign using the Hydrophobicity Index (a), Hydropathy Index (b), and Solubility-Weighted Index (SWI) (c) as solubility indices in the loss function. Each color indicates the weight of the solubility loss. The gray line with a weight of 0.0 shows the distribution of binding affinity for the sequence designed by AfDesign without a solubility index. The dashed line shows the average of binding affinity of the p53 peptide.

In order to examine the possibility of designing sequences with higher binding affinity and logS than existing complex peptides for the target protein, which was the objective of this study, a scatter plot of binding affinity and logS is shown Figure 6. Figure 6 shows that sequences designed using the Hydrophobicity Index for solubility loss tended to decrease in binding affinity as logS increased, while sequences designed using the Hydropathy Index showed higher binding affinity and higher logS than the sequences designed using the other indices or sequences without solubility loss. The sequences designed using the Hydropathy Index tended to have higher binding affinity with increasing logS than sequences designed using the other indices or sequences without solubility loss. This was the opposite trend for sequences designed using the Hydrophobicity Index. Sequences designed using SWI for solubility loss showed little change in binding affinity as logS increased. Thus, the Hydrophobicity Index and SWI were able to control logS while maintaining or increasing the binding affinity of sequences designed by the AfDesign binder hallucination protocol with solubility loss.

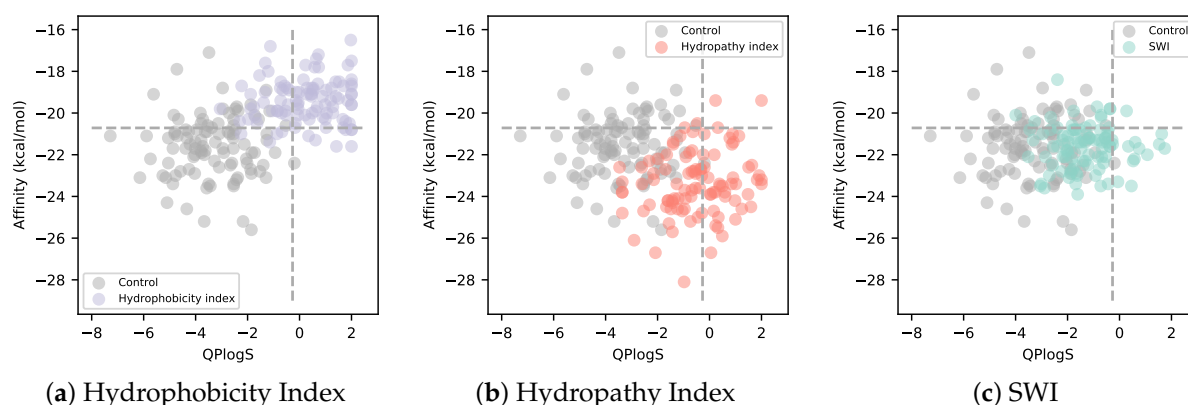


Figure 6. Scatter plots of binding affinity and logS for various weights of sequences designed in AfDesign using three solubility indices. Each point shows the binding affinity and logS of a sequence designed in AfDesign with a weight of 0.5 using the Hydrophobicity Index (a), Hydropathy Index (b), and Solubility-Weighted Index (SWI) (c) as solubility indices in the loss function. Gray dots show the binding affinity and logS of sequences designed in AfDesign without a solubility index. The horizontal dashed line shows the average binding affinity of the p53 peptide. The vertical dashed line shows the logS values for the p53 peptide.

4. Discussion

The results of this study show that the AfDesign binder hallucination protocol with the addition of the Hydropathy Index or SWI for solubility loss can design candidate peptide sequences that may have binding affinity comparable to or greater than that of the p53 peptide. The binding affinities of sequences designed with the AfDesign binder hallucination protocol were higher than those of random sequences. In addition, the binding affinity of the sequence with one residue substitution of p53 peptide was distributed around the binding affinity value of the p53 peptide; its distribution was lower than that of the binding affinity of the sequence designed with the binder hallucination protocol of AfDesign with or without the Hydropathy Index and SWI for solubility loss (Supplementary Figure S1). In summary, it is likely that sequences designed with the AfDesign binder hallucination protocol have higher binding affinities than random sequences as well as the binding affinity of the p53 peptide and its residue substituted sequences.

In addition to the evaluation of solubility and binding affinity, visualization using PyMOL and the evaluation of docking using DockQ were also performed (Supplementary Figure S2). Comparing the 3D structure of the sequence with the highest binding affinity and higher logS than the p53 peptide, which was designed using the solubility loss with SWI in the AfDesign binder hallucination protocol, with the actual 3D structure of MDM2, we find that the coordinates of the sequence are almost identical to those of the p53 peptide. DockQ score was calculated to be 0.500. LRMS, the RMSD of the main chain atoms of the peptide, was calculated to be 5.303 Å; iRMS, the RMSD of the heavy atoms of the residues at the interface between the ligands, was calculated to be 2.294 Å. This is of medium quality in the protein–protein docking criteria of CAPRI and indicates the validity of the sequence designed using solubility loss with SWI in the AfDesign binder hallucination protocol to dock with MDM2.

The impact of adding the solubility index to the loss of the AfDesign binder hallucination protocol is mainly seen in the change in loss from the 101st iteration, when the sequence features are first converted from logits to probabilities by multiplying logits by softmax, and from the 201st iteration, when the probability argmax is first converted to probabilities by multiplying softmax by logits. The main impact on the overall loss of AfDesign is the change in the loss of pLDDT and PAE_inter from the 101st iteration, when the array features are first converted to probabilities by multiplying logits by softmax, and the 201st iteration, when the probabilities are first converted to argmax. This tends to increase the overall loss. Moreover, the variation after the 201st iteration is particularly large compared with the case where the solubility index is not added to the loss of the AfDesign binder hallucination protocol. By varying solubility loss weights, solubility loss decreases. This may be a trade-off between optimizing for sequences that tend to have lower solubility according to the solubility index and optimizing for peptide sequences that are likely to form optimal complexes (Supplementary Figure S3).

In terms of designing PPI inhibitor peptides, the question is whether it is possible to design peptides in which the β -sheet serves as the PPI binding interface as well as those in which the helix domain serves as the PPI binding interface, as in the MDM2p53 complex. Therefore, we tested the possibility of designing a PD-1PD-L1 complex as an example of a complex in which the β -sheets are at the PPI binding interface, and AfDesign is able to design a binder that can form a complex with PD-1. The results show that the default AfDesign binder hallucination protocol is sufficient to design a peptide with high solubility. However, we also found that the distribution of binding affinity was almost the same as that of the random sequence (Supplementary Figure S4).

By creating a solubility loss using three solubility indices for the AfDesign binder hallucination, we were able to design sequences that maintained or increased binding affinity while controlling solubility. The sequences designed by AfDesign could be adversarial, such as those simply preferred by AlphaFold. However, in this study, instead of evaluating the sequences by AfDesign alone, we sampled multiple seeds, calculated the binding affinity to the target protein by the Docking tool, which is not related to AlphaFold,

and calculated logS using the solubility calculation tool. The results show an overall trend of the target protein having a high affinity and solubility being controlled. However, these results are based on computer simulations; true solubility and binding affinity will not be known until experiments are conducted. However, the ability to design sequences with high solubility and high binding affinity that could be candidates for PPI inhibition from among a large number of candidate sequences is meaningful from both the drug discovery and industrial perspectives.

In this study, approximately 1600 sequences were designed by AfDesign binder hallucination with or without three solubility indices and five weights (Supplementary Tables S3–S6). To filter out sequences with high solubility and high binding affinity from a large number of candidate sequences, we calculated the logS and binding affinity of PDIQ peptide, which have been experimentally verified to exhibit extremely higher binding affinity than p53 peptide as a target protein for MDM2 [55]. Since the PDIQ peptide has 12 residues, logS and binding affinity were also calculated for the PDIQ + 'N' peptide, in which asparagine residue was added to 13th residue of the PDIQ peptide. The asparagine was the same as 13th residue of the p53 peptide to allow for comparisons with this study, and they were calculated for the PDIQ + 'N' peptide. The logS of PDIQ was -2.5230 and that of PDIQ + 'N' was 0.4420 , indicating that the logS of PDIQ + 'N' was higher. This may be due to the fact that 'N' is a hydrophilic amino acid. This was also higher than the logS of p53 peptide (-0.2690). Additionally, we calculated the binding affinity of PDIQ + 'N' peptide. The results showed that the binding affinity of PDIQ + 'N' was -22.5 , which is higher than the binding affinity of p53 peptide of -20.8 (Supplementary Table S7). This result is consistent with a previous study that experimentally evaluated the binding affinity of a 12-residue PDIQ peptide and a 12-residue p53 peptide. These results suggest that by filtering sequences with higher logS and higher binding affinity than PDIQ + 'N' peptide, it may be possible to efficiently screen peptides by considering solubility.

We filtered the sequences designed in this study using the logS and binding affinity of the PDIQ + 'N' peptide as thresholds. Consequently, not a single candidate peptide remained in AfDesign without solubility index. This is mainly due to solubility (Figures 2a and 5). In AfDesign using the solubility index, there were 9 out of 600 sequences using SWI, 90 out of 600 sequences were filtered out using Hydropathy Index, and 2 out of 600 sequences were filtered out using Hydrophobicity Index. Sequences designed by AfDesign using the Hydropathy Index as a solubility index filtered out more sequences than those using other solubility indexes (Supplementary Table S8). A WebLogo was created using these 90 sequences; as with the PDIQ, the sixth and seventh residues are mainly characterized as 'W', and the fifth residue tends to be 'Y', which may enhance binding affinity. In addition, the C- and N-terminal residues tend to be 'E' and 'D', which may enhance the overall solubility (Supplementary Figure S5).

A recent study reported that competitive binding can be simulated in protein–peptide docking using ColabFold [48]. In other words, competition binding experiments of peptides can be performed virtually. We further performed competitive peptide binding prediction using the filtered sequence DEVYYWYYHLEND, which showed the highest binding affinity among the 90 sequences, and visualized the top-ranked results for each of the 20 runs; the designed peptide was found at the MDM2 binding site, while the p53 peptide was far away from the binding site (Supplementary Figure S6). This result is consistent with the results of the reference paper. Furthermore, we calculated the pLDDT and RMSD of all predictions and found that the peptide had a higher pLDDT and lower RMSD than the p53 peptide, except for one prediction among the 100 predictions (Supplementary Figure S7 and Supplementary Table S9). These results indicate that binding affinity may be higher than that of p53, which is consistent with the results of this study.

Furthermore, PLIP was used to compare the interatomic interaction between MDM2/the designed peptide DEVYYWYYHLEND and the MDM2/p53. As shown in Figure 7a, a hydrogen bond between His96 of MDM2 and Asp13 of the designed peptide, a salt bridge between Lys51 of MDM2 and Glu11 of the designed peptide, and between His96 of MDM2 and Asp13

of the designed peptide were identified. In addition, several hydrophobic bonds were identified, such as that between Val75 of MDM2 and Trp6 and that between Val93 of MDM2 and Trp6 of the designed peptide. In the p53 peptide shown in Figure 7b, there is also a salt bridge between Lys51 of MDM2 and Glu28 of p53 peptide (12th residue), and hydrophobic binding between Val75 of MDM2 and Phe19 of p53 peptide (3th residue) and between Val93 of MDM2 and Leu22 of p53 peptide (6th residue), indicating that the major interactions of MDM2/p53 are also reproduced in the model of the designed peptide using ColabFold. The next step will be to develop a design scheme for improving a specific targeted peptide sequence in terms of affinity, biochemical properties, etc.

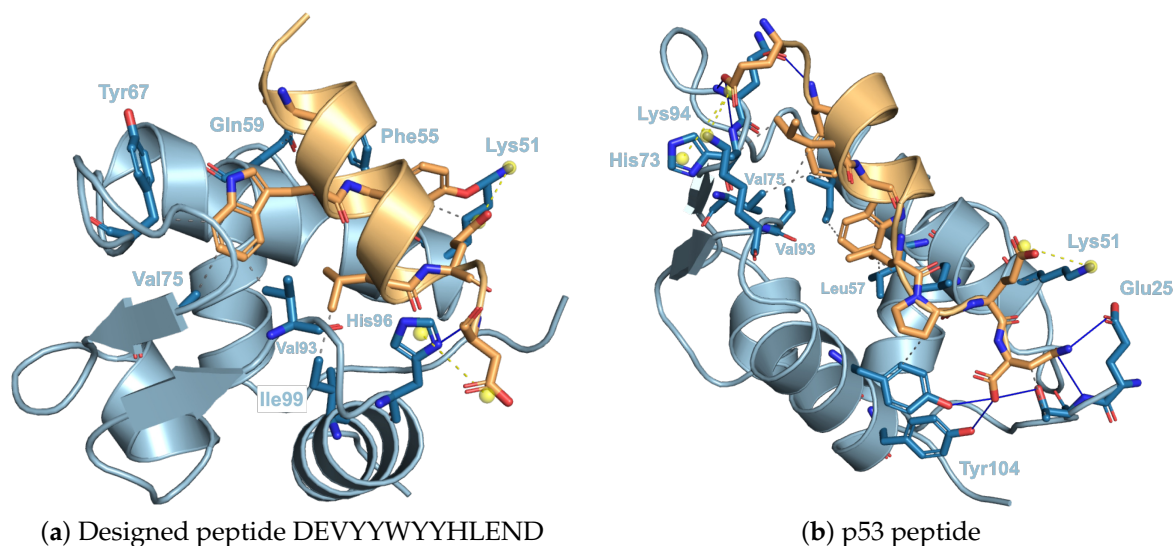


Figure 7. Interatomic interaction analysis for MDM2 with the designed peptide DEVYYWYYHLEND and p53 peptide using PLIP. Interatomic interaction between MDM2 and the designed peptide (a); interatomic interaction between MDM2 and the p53 peptide (b). Light blue color indicates MDM2. Light orange indicates the designed peptide (a) and p53 peptide (b). Hydrophobic bonds, hydrogen bonds, salt bridges, and charge centers are indicated by gray dashed lines, blue solid lines, yellow dashed lines, and the yellow sphere, respectively.

5. Conclusions

In this study, we applied the AfDesign binder hallucination protocol to PPI-targeting peptide design and attempted simultaneous optimization of peptide sequence design and solubility control. Due to the AfDesign flexible loss function concept, we could design peptides likely to be PPI while controlling solubility by adding a solubility index to the loss. Furthermore, the evaluation of binding affinity by docking and competitive peptide binding prediction allowed us to design peptide sequences with higher binding affinity than that of the native peptide with PPI as the target protein. Furthermore, a comparison of interatomic interactions with those of the native MDM2/p53 complex revealed that the designed peptides also have interatomic interactions with common MDM2 amino acid residues. Based on these results, a further optimization of this approach might be useful in drug discovery using PPI-targeting peptide design.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines10071626/s1>, Figure S1: Comparison of MDM2 binding affinities between random sequences and one residue substitution sequences of p53 peptides and sequences designed by AfDesign; Figure S2: Comparison of crystal structures and three-dimensional (3D) structures predicted by AfDesign; Figure S3: Comparison of each solubility index at the median of the AfDesign loss values; Figure S4: Comparison of PD-1 binding affinity between random sequences and sequences designed in AfDesign with each solubility index; Figure S5: Weblog of sequences designed with AfDesign binder hallucination using the Hydropathy Index as

a solubility index; Figure S6: All top rank models from competitive peptide binding prediction of MDM2 with p53 and our highest affinity peptide consistently indicate the highest affinity peptide as the strong binder; Figure S7: Heatmap of all pLDDT values in predictions of competitive peptide binding of the competitor peptide and p53 peptide with MDM2 using ColabFold; Table S1: Values of each solubility index; Table S2: Normalized solubility indices used as solubility loss for AfDesign; Table S3: Sequences designed by AfDesign using Solubility-Weighted Index as a solubility index; Table S4: Sequences designed by AfDesign using Hydropathy Index as a solubility index; Table S5: Sequences designed by AfDesign using Hydrophobicity Index as a solubility index; Table S6: Sequences designed by AfDesign without a solubility index; Table S7: QPlogS and binding affinity for PDIQ and p53 peptide; Table S8: Sequences filtered by logS and Affinity thresholds for PDIQ + 'N'; Table S9: RMSD of p53 peptide and the competitor peptide in the prediction of competitive binding using ColabFold.

Author Contributions: Conceptualization, T.K. and M.O.; methodology, T.K. and M.O.; software, T.K.; validation, T.K.; formal analysis, T.K. and M.O.; investigation, T.K. and M.O.; writing—original draft preparation, T.K. and M.O.; writing—review and editing, T.K. and M.O.; visualization, T.K.; supervision, M.O.; project administration, M.O.; funding acquisition, M.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the Japan Science and Technology Agency (JST) FOREST (Grant No. JPMJFR216J), JST ACT-X (Grant No. JPMJAX20A3), Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant No. 20H04280), and Japan Agency for Medical Research and Development (AMED) Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS) (Grant No. JP22ama121026).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The implementation and experimental data are available on an open-source basis at https://github.com/ohuelab/Solubility_AfDesign (accessed on 13 May 2022).

Acknowledgments: The computational experiments were performed using a TSUBAME 3.0 super-computer at the Tokyo Institute of Technology, for which we thank the institution.

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

References

1. Bonetta, L. Interactome under Construction. *Nature* **2010**, *468*, 851–852. [[CrossRef](#)]
2. Oughtred, R.; Rust, J.; Chang, C.; Breitkreutz, B.-J.; Stark, C.; Willems, A.; Boucher, L.; Leung, G.; Kolas, N.; Zhang, F.; et al. The BioGRID Database: A Comprehensive Biomedical Resource of Curated Protein, Genetic, and Chemical Interactions. *Protein Sci.* **2021**, *30*, 187–200. [[CrossRef](#)]
3. Toogood, P.L. Inhibition of Protein-Protein Association by Small Molecules: Approaches and Progress. *J. Med. Chem.* **2002**, *45*, 1543–1558. [[CrossRef](#)]
4. Arkin, M.R.; Wells, J.A. Small-molecule inhibitors of protein–protein interactions: Progressing towards the dream. *Nat. Rev. Drug Discov.* **2004**, *3*, 301–317. [[CrossRef](#)]
5. Dev, K.K. Making protein interactions druggable: Targeting PDZ domains. *Nat. Rev. Drug Discov.* **2004**, *3*, 1047–1056. [[CrossRef](#)]
6. Jin, L.; Wang, W.; Fang, G. Targeting Protein-Protein Interaction by Small Molecules. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 435–456. [[CrossRef](#)]
7. Ivanov, A.A.; Khuri, F.R.; Fu, H. Targeting protein–protein interactions as an anticancer strategy. *Trends Pharmacol. Sci.* **2013**, *34*, 393–400. [[CrossRef](#)]
8. Souers, A.J.; Levenson, J.D.; Boghaert, E.R.; Ackler, S.L.; Catron, N.D.; Chen, J.; Dayton, B.D.; Ding, H.; Enschede, S.H.; Fairbrother, W.J.; et al. ABT-199, a Potent and Selective BCL-2 Inhibitor, Achieves Antitumor Activity While Sparing Platelets. *Nat. Med.* **2013**, *19*, 202–208. [[CrossRef](#)]
9. Shin, W.-H.; Kumazawa, K.; Imai, K.; Hirokawa, T.; Kihara, D. Current challenges and opportunities in designing protein–protein interaction targeted drugs. *Adv. Appl. Bioinform. Chem.* **2020**, *13*, 11–25.
10. Truoger, J.; George, A.; Holien, J.K. Analysis of Physicochemical Properties of Protein–Protein Interaction Modulators Suggests Stronger Alignment with the “Rule of Five”. *RSC Med. Chem.* **2021**, *12*, 1731–1749. [[CrossRef](#)]
11. Kosugi, T.; Ohue, M. Quantitative Estimate Index for Early-Stage Screening of Compounds Targeting Protein-Protein Interactions. *Int. J. Mol. Sci.* **2021**, *22*, 10925. [[CrossRef](#)]
12. Wodak, S.J.; Janin, J. Structural Basis of Macromolecular Recognition. *Adv. Protein Chem.* **2002**, *61*, 9–73.

13. Conte, L.L.; Chothia, C.; Janin, J. The Atomic Structure of Protein-Protein Recognition Sites¹¹ Edited by A. R. Fersht. *J. Mol. Biol.* **1999**, *285*, 2177–2198. [[CrossRef](#)]
14. Smith, M.C.; Gestwicki, J.E. Features of Protein-Protein Interactions That Translate into Potent Inhibitors: Topology, Surface Area and Affinity. *Expert Rev. Mol. Med.* **2012**, *14*, e16. [[CrossRef](#)]
15. Wells, J.A.; McClendon, C.L. Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces. *Nature* **2007**, *450*, 1001–1009. [[CrossRef](#)]
16. Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. Recent Advances in the Development of Protein-Protein Interactions Modulators: Mechanisms and Clinical Trials. *Signal Transduct. Target. Ther.* **2020**, *5*, 213. [[CrossRef](#)]
17. Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic Peptides: Current Applications and Future Directions. *Signal Transduct. Target. Ther.* **2022**, *7*, 48. [[CrossRef](#)]
18. Cabri, W.; Cantelmi, P.; Corbisiero, D.; Fantoni, T.; Ferrazzano, L.; Martelli, G.; Mattellone, A.; Tolomelli, A. Therapeutic Peptides Targeting PPI in Clinical Development: Overview, Mechanism of Action and Perspectives. *Front. Mol. Biosci.* **2021**, *8*, 565. [[CrossRef](#)]
19. Wu, C.-H.; Liu, I.-J.; Lu, R.-M.; Wu, H.-C. Advancement and Applications of Peptide Phage Display Technology in Biomedical Science. *J. Biomed. Sci.* **2016**, *23*, 8. [[CrossRef](#)]
20. Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. In Vitro Virus: Bonding of mRNA Bearing Puromycin at the 3'-Terminal End to the C-Terminal End of Its Encoded Protein on the Ribosome in Vitro. *FEBS Lett.* **1997**, *414*, 405–408. [[CrossRef](#)]
21. Roberts, R.W.; Szostak, J.W. RNA-Peptide Fusions for the in Vitro Selection of Peptides And Proteins. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297–12302. [[CrossRef](#)]
22. Das, P.; Sercu, T.; Wadhawan, K.; Padhi, I.; Gehrman, S.; Cipcigan, F.; Chenthamarakshan, V.; Strobel, H.; dos Santos, C.; Chen, P.-Y.; et al. Accelerated Antimicrobial Discovery via Deep Generative Models and Molecular Dynamics Simulations. *Nat. Biomed. Eng.* **2021**, *5*, 613–623. [[CrossRef](#)] [[PubMed](#)]
23. Cardoso, M.H.; Orozco, R.Q.; Rezende, S.B.; Rodrigues, G.; Oshiro, K.G.N.; Cândido, E.S.; Franco, O.L. Computer-Aided Design of Antimicrobial Peptides: Are We Generating Effective Drug Candidates? *Front. Microbiol.* **2020**, *10*, 3097. [[CrossRef](#)] [[PubMed](#)]
24. Capecchi, A.; Zhang, A.; Reymond, J.-L. Populating Chemical Space with Peptides Using a Genetic Algorithm. *J. Chem. Inf. Model.* **2020**, *60*, 121–132. [[CrossRef](#)] [[PubMed](#)]
25. Chevalier, A.; Silva, D.-A.; Rocklin, G.J.; Hicks, D.R.; Vergara, R.; Murapa, P.; Bernard, S.M.; Zhang, L.; Lam, K.-H.; Yao, G.; et al. Massively Parallel de Novo Protein Design for Targeted Therapeutics. *Nature* **2017**, *550*, 74–79. [[CrossRef](#)] [[PubMed](#)]
26. Strauch, E.-M.; Bernard, S.M.; La, D.; Bohn, A.J.; Lee, P.S.; Anderson, C.E.; Nieusma, T.; Holstein, C.A.; Garcia, N.K.; Hooper, K.A.; et al. Computational Design of Trimeric Influenza-Neutralizing Proteins Targeting the Hemagglutinin Receptor Binding Site. *Nat. Biotechnol.* **2017**, *35*, 667–671. [[CrossRef](#)]
27. Silva, D.-A.; Yu, S.; Ulge, U.Y.; Spangler, J.B.; Jude, K.M.; Labão-Almeida, C.; Ali, L.R.; Quijano-Rubio, A.; Ruterbusch, M.; Leung, I.; et al. De Novo Design of Potent and Selective Mimics of IL-2 and IL-15. *Nature* **2019**, *565*, 186–191. [[CrossRef](#)]
28. Baran, D.; Pszolla, M.G.; Lapidoth, G.D.; Norn, C.; Dym, O.; Unger, T.; Albeck, S.; Tyka, M.D.; Fleishman, S.J. Principles for Computational Design of Binding Antibodies. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10900–10905. [[CrossRef](#)]
29. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
30. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871–876. [[CrossRef](#)]
31. Gulsevin, A.; Meiler, J. Benchmarking Peptide Structure Prediction with AlphaFold2. *bioRxiv* **2022**. [[CrossRef](#)]
32. Tsaban, T.; Varga, J.K.; Avraham, O.; Ben-Aharon, Z.; Khramushin, A.; Schueler-Furman, O. Harnessing Protein Folding Neural Networks for Peptide-Protein Docking. *Nat. Commun.* **2022**, *13*, 176. [[CrossRef](#)] [[PubMed](#)]
33. Anishchenko, I.; Pellock, S.J.; Chidyausiku, T.M.; Ramelot, T.A.; Ovchinnikov, S.; Hao, J.; Bafna, K.; Norn, C.; Kang, A.; Bera, A.K.; et al. De Novo Protein Design by Deep Network Hallucination. *Nature* **2021**, *600*, 547–552. [[CrossRef](#)]
34. Cao, L.; Coventry, B.; Goreshnik, I.; Huang, B.; Park, J.S.; Jude, K.M.; Marković, I.; Kadam, R.U.; Verschuere, K.H.G.; Verstraete, K.; et al. Design of Protein Binding Proteins from Target Structure Alone. *Nature* **2022**, *605*, 551–560. [[CrossRef](#)] [[PubMed](#)]
35. Jendrusch, M.; Korbel, J.O.; Sadiq, S.K. AlphaDesign: A de Novo Protein Design Framework Based on AlphaFold. *bioRxiv* **2021**. [[CrossRef](#)]
36. *GitHub*, AfDesign (v1.0.2); GitHub: San Francisco, CA, USA, 2022. Available online: <https://github.com/sokrypton/ColabDesign/tree/main/af> (accessed on 14 March 2022).
37. Fosgerau, K.; Hoffmann, T. Peptide Therapeutics: Current Status and Future Directions. *Drug Discov.* **2015**, *20*, 122–128. [[CrossRef](#)] [[PubMed](#)]
38. Esposito, D.; Chatterjee, D.K. Enhancement of Soluble Protein Expression through the Use of Fusion Tags. *Curr. Opin. Biotechnol.* **2006**, *17*, 353–358. [[CrossRef](#)]
39. Wu, S.-J.; Luo, J.; O'Neil, K.T.; Kang, J.; Lacy, E.R.; Canziani, G.; Baker, A.; Huang, M.; Tang, Q.M.; Raju, T.S.; et al. Structure-Based Engineering of a Monoclonal Antibody for Improved Solubility. *Protein Eng. Des. Sel.* **2010**, *23*, 643–651. [[CrossRef](#)]

40. Mant, C.T.; Kovacs, J.M.; Kim, H.-M.; Pollock, D.D.; Hodges, R.S. Intrinsic Amino Acid Side-Chain Hydrophilicity/Hydrophobicity Coefficients Determined by Reversed-Phase High-Performance Liquid Chromatography of Model Peptides: Comparison with Other Hydrophilicity/Hydrophobicity Scales. *Pept. Sci.* **2009**, *92*, 573–595. [[CrossRef](#)]
41. Bhandari, B.K.; Gardner, P.P.; Lim, C.S. Solubility-Weighted Index: Fast and Accurate Prediction of Protein Solubility. *Bioinformatics* **2020**, *36*, 4691–4698. [[CrossRef](#)]
42. Argos, P.; Rao, J.K.M.; Hargrave, P.A. Structural Prediction of Membrane-Bound Proteins. *Eur. J. Biochem.* **1982**, *128*, 565–575. [[CrossRef](#)] [[PubMed](#)]
43. Kyte, J.; Doolittle, R.F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [[CrossRef](#)]
44. Zhang, Y.; Sanner, M.F. AutoDock CrankPep: Combining Folding and Docking to Predict Protein–Peptide Complexes. *Bioinformatics* **2019**, *35*, 5121–5127. [[CrossRef](#)] [[PubMed](#)]
45. Word, J.M.; Lovell, S.C.; Richardson, J.S.; Richardson, D.C. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-Chain Amide Orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747. [[CrossRef](#)]
46. 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](#)]
47. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)]
48. Chang, L.; Perez, A. AlphaFold Encodes the Principles to Identify High Affinity Peptide Binders. *bioRxiv* **2022**. [[CrossRef](#)]
49. Adasme, M.F.; Linnemann, K.L.; Bolz, S.N.; Kaiser, F.; Salentin, S.; Haupt, V.J.; Schroeder, M. PLIP 2021: Expanding the Scope of the Protein-Ligand Interaction Profiler to DNA and RNA. *Nucleic Acids Res.* **2021**, *49*, W530–W534. [[CrossRef](#)]
50. Norn, C.; Wicky, B.I.M.; Juergens, D.; Liu, S.; Kim, D.; Tischer, D.; Koepnick, B.; Anishchenko, I.; Foldit Players; Baker, D.; et al. Protein Sequence Design by Conformational Landscape Optimization. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2017228118. [[CrossRef](#)]
51. Weng, G.; Gao, J.; Wang, Z.; Wang, E.; Hu, X.; Yao, X.; Cao, D.; Hou, T. Comprehensive Evaluation of Fourteen Docking Programs on Protein–Peptide Complexes. *J. Chem. Theory Comput.* **2020**, *16*, 3959–3969. [[CrossRef](#)]
52. Baek, M. Twitter Post: Adding a Big Enough Number for Residue_Index Feature Is Enough to Model Hetero-Complex Using AlphaFold (green&cyan: Crystal Structure/Magenta: Predicted Model w/residue_index Modification). Available online: <https://twitter.com/minkbaek/status/1417538291709071362> (accessed on 20 July 2021).
53. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* **2022**, *19*, 679–682. [[CrossRef](#)] [[PubMed](#)]
54. Shin, W.-H.; Christoffer, C.W.; Kihara, D. In Silico Structure-Based Approaches to Discover Protein-Protein Interaction-Targeting Drugs. *Methods* **2017**, *131*, 22–32. [[CrossRef](#)] [[PubMed](#)]
55. Phan, J.; Li, Z.; Kasprzak, A.; Li, B.; Sebt, S.; Guida, W.; Schönbrunn, E.; Chen, J. Structure-Based Design of High Affinity Peptides Inhibiting the Interaction of P53 with MDM2 and MDMX. *J. Biol. Chem.* **2010**, *285*, 2174–2183. [[CrossRef](#)] [[PubMed](#)]