

RESEARCH ARTICLE

# Large-scale docking predicts that sORF-encoded peptides may function through protein-peptide interactions in *Arabidopsis thaliana*

Rashmi R. Hazarika<sup>1</sup>, Nikolina Sostaric<sup>1</sup>, Yifeng Sun<sup>1,2</sup>, Vera van Noort<sup>1,3\*</sup>

**1** Department of Microbial and Molecular Systems, KU Leuven, Leuven, Belgium, **2** Faculty of Engineering Technology, Campus Group T, KU Leuven, Leuven, Belgium, **3** Institute of Biology Leiden, Leiden University, Leiden, The Netherlands

\* [vera.vannoort@kuleuven.be](mailto:vera.vannoort@kuleuven.be)



**OPEN ACCESS**

**Citation:** Hazarika RR, Sostaric N, Sun Y, van Noort V (2018) Large-scale docking predicts that sORF-encoded peptides may function through protein-peptide interactions in *Arabidopsis thaliana*. PLoS ONE 13(10): e0205179. <https://doi.org/10.1371/journal.pone.0205179>

**Editor:** Manuela Helmer-Citterich, Universita degli Studi di Roma Tor Vergata, ITALY

**Received:** June 8, 2018

**Accepted:** September 20, 2018

**Published:** October 15, 2018

**Copyright:** © 2018 Hazarika et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Relevant data are within the paper and its Supporting Information Files. In addition, annotations are made available through our portal ARA-PEPs <https://www.biw.kuleuven.be/CSB/ARA-PEPs/>.

**Funding:** This work has been supported by the KU Leuven Research Fund. NS is a doctoral fellow (1112318N) of the Research Foundation – Flanders (FWO). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the

## Abstract

Several recent studies indicate that small Open Reading Frames (sORFs) embedded within multiple eukaryotic non-coding RNAs can be translated into bioactive peptides of up to 100 amino acids in size. However, the functional roles of the 607 Stress Induced Peptides (SIPs) previously identified from 189 Transcriptionally Active Regions (TARs) in *Arabidopsis thaliana* remain unclear. To provide a starting point for functional annotation of these plant-derived peptides, we performed a large-scale prediction of peptide binding sites on protein surfaces using coarse-grained peptide docking. The docked models were subjected to further atomistic refinement and binding energy calculations. A total of 530 peptide-protein pairs were successfully docked. In cases where a peptide encoded by a TAR is predicted to bind at a known ligand or cofactor-binding site within the protein, it can be assumed that the peptide modulates the ligand or cofactor-binding. Moreover, we predict that several peptides bind at protein-protein interfaces, which could therefore regulate the formation of the respective complexes. Protein-peptide binding analysis further revealed that peptides employ both their backbone and side chain atoms when binding to the protein, forming predominantly hydrophobic interactions and hydrogen bonds. In this study, we have generated novel predictions on the potential protein-peptide interactions in *A. thaliana*, which will help in further experimental validation.

## Introduction

Over the years, the functional importance of short plant signaling peptides has been overshadowed by other groups of molecules. For instance, the phytohormone auxin was shown to be involved in bidirectional polar transport across tissues, controlling plant growth-related processes [1,2]. Furthermore, microRNAs are considered to be important signaling molecules, regulating developmental processes in plants by moving from one cell to another over long distances [3]. It was only within the last decade that the roles of plant peptides in a wide variety

Research Foundation – Flanders (FWO) and the Flemish Government – department EWI.

**Competing interests:** The authors have declared that no competing interests exist.

of cellular functions were established by multiple studies. For instance, plant peptides may participate in cell-to-cell communications, directly interact with pathogens, or interfere with signaling cascades. The plant peptidome is broadly comprised of three main types of peptides *viz.*, derived from non-functional precursors, derived from functional precursors, and not derived from any precursor protein [4,5]. Most of the plant peptides identified so far are products of non-functional precursors, and a nice example is the CLAVATA3 (CLV3)/EMBRYO SURROUNDING REGION (CLE) family comprising of post-translationally modified peptides which are involved in the regulation of transcription factor activity and/or other downstream events, leading to altered transcription factor expression patterns by participating in the MAPKs cascade [6]. Some peptides however are not derived from any precursor protein and may instead be directly encoded by short Open Reading Frames (sORFs) that were earlier assumed to be non-coding [7–12]. Several recent studies now clearly demonstrate that sORFs embedded within non-coding RNAs (ncRNAs), intergenic regions and pseudogenes can indeed be translated into bioactive peptides. So far only a handful of sORF-encoded peptides have been studied in detail. One such example is the early nodulin gene *enod40* that encodes bioactive oligopeptides involved in root nodule organogenesis [13–15].

In our previous work, we have identified several Transcriptionally Active Regions (TARs) induced upon the application of biotic (*Botrytis cinerea*) and abiotic stress (Paraquat) in *Arabidopsis thaliana*. These TARs could be translated into Stress-Induced Peptides (SIPs), which can be specifically categorized depending on the applied stress condition into *Botrytis cinerea* Induced Peptides (BIPs) and Oxidative Stress Induced Peptides (OSIPs), catalogued in a database ARA-PEPs [16,17]. Although some physiological effects of sORF-encoded peptides have been discovered, the molecular mechanisms by which they exert their function through interaction with other molecules is largely unknown. We postulate that they could work through interactions with proteins, as protein-peptide interactions have previously been well established as important mediators of protein-protein interactions, partaking in signal transduction, cell-to-cell communication, protein trafficking and other regulatory pathways [18–24].

Peptide-mediated interactions constitute 15–40% of all protein-protein interactions [21,22]. Most of the studies performed so far in order to understand protein-peptide interactions focused on small peptides that may be short linear recognition motifs originating from disordered protein regions [18,20]. Investigating those interactions is experimentally challenging, and this has led to limited progress in the field of protein-peptide interactions validation. On the other side, the successful modeling of such complexes depends on prior structural knowledge of the protein that acts as a receptor. A number of protein-peptide docking methods, such as Rosetta FlexPepDock [25,26], GalaxyPepDock [27–29], MedusaDock [30], DynaDock [31], CABS-dock [32, 33], pepATTRACT [19], HADDOCK [34,35], and tools to predict binding sites on proteins, such as PepSite2 [21,36], have been developed. Moreover, curated data also exists for characterization of protein-peptide interactions, e.g. a non-redundant database of high-resolution peptide-protein complexes called the *peptiDB* [18]. Although docking strategies are the preferred methods for predicting protein-peptide interactions, they are associated with certain limitations, such as difficulty in docking peptides longer than 4 Amino Acids (AA), owing to their high degree of conformational flexibility. In our current study, we opted to combine the peptide-protein docking method pepATTRACT-local with binding site predictions obtained from the PepSite2 server, that uses training data of known protein-peptide complexes from Protein Data Bank (PDB) [37] to define Spatial Position Specific Scoring Matrices (S-PSSMs). Furthermore, as biological systems are not static, we also looked into dynamics of the obtained docked models and calculated the energetics of binding based on multiple conformations that the protein-peptide system can acquire in solution.

We hypothesize that a SIP encoded by a TAR may bind on a protein at one of its pockets, or to a known ligand or cofactor-binding site, and consequently affect the function of the protein as a whole. Moreover, peptides may bind at the interfaces of multi-chain complexes and modulate their activity. Protein-peptide interactions involve smaller interfaces whose affinity is usually weaker and are transient as they can rapidly make and break interactions in response to sudden cellular perturbations, for instance stress conditions [23,38].

In recent years, there has been a growing interest in developing protein-protein interaction inhibitors based on peptides or peptide derivatives. Molecules that can mimic the binding or functional sites of proteins are promising candidates for different types of biological applications. Synthetic peptides are widely choiced molecules for mimicry of protein sites because they can be easily synthesized as exact copies of protein fragments, or they may be generated by introducing diverse chemical modifications to the peptide sequence, and/or by modifying the peptide backbone [39]. Peptide mimics have the potential to be developed as attractive targets for agriculture, especially plant disease control, and for therapeutic interventions [40].

As detailed above, when no information about the peptide-binding site on protein receptors is available, there is need for computational approaches to predict peptide-binding sites on protein surfaces, as these models can serve as starting points for experimental characterization of novel protein-peptide interactions. This will be especially beneficial in studying the model plant *A. thaliana*, in which peptides have multiple important roles, but have been understudied till now. Molecular docking studies can be used effectively to explore the binding mode of putative peptides onto proteins, serving as an excellent approach for *de novo* design of peptides targeting various other biosynthetic pathways in major eukaryotes. In the current study, we investigate the potential roles of sORF-encoded stress induced peptides in targeting the key regulatory enzymes, as this could further indicate their roles in mediating stress-response mechanisms.

## Materials and methods

### Screening of peptide binding pockets on protein surfaces

We generated a peptide fragment library consisting of 23,113 k-mers ranging in size from 4 to 10 amino acids, using sequences of SIPs [16,17] and following a sliding window approach. We extracted 2,561 structures corresponding to 1,009 *A. thaliana* proteins from Protein Data Bank, PDB ([www.rcsb.org](http://www.rcsb.org)) [37]. 996 structures were retained after filtering out redundant ones. We carried out an all-vs-all screening of potential peptide binding sites on *A. thaliana* proteins using PepSite2 [21,36]. Screened pairs were adjusted for multiple testing using the Benjamini and Hochberg procedure [41] and significance was determined using a False Discovery Rate (FDR) cutoff set to 0.25. Multiple testing was performed using the R Bioconductor package “multtest” [42].

We used the default settings of BLASTP2.2.28+ algorithm [43] with E-value  $\leq 10$  to screen out peptides that show sequence similarity with a protein chain, as entire or a part of a SIP may mimic a specific binding motif on the protein, or resemble a loop from a large structured protein, a disordered region in protein termini or interfaces between defined domains [18,20].

### Building peptide models, docking and structure refinement

We shortlisted 576 protein-peptide pairs from PepSite2 output to build atomistic models and perform docking studies, using the protein-peptide coarse-grained *ab initio* docking protocol pepATTRACT [19]. For each peptide, three idealized peptide conformations (extended,  $\alpha$ -helical and polyproline) were built using the Python library PeptideBuilder [44]. The backbone dihedral angles used to represent the three peptide conformations were  $\alpha$ -helical ( $\Phi = -57^\circ$ ,

$\Psi = -47^\circ$ ), extended ( $\Phi = -139^\circ$ ,  $\Psi = -135^\circ$ ), and polyproline conformations ( $\Phi = -78^\circ$ ,  $\Psi = 149^\circ$ ) [35].

In the current study, the rigid body docking models were ranked by ATTRACT score, and the top-ranked 100 structures were subjected to atomistic refinement using the flexible interface refinement method iATTRACT. We used the distance restraint based local docking protocol of pepATTRACT to restrict the sampling during rigid body sampling stage and flexible refinement stage towards the PepSite2 predicted interface residues. The placement of peptide and protein was optimized during iATTRACT refinement. At this stage, the interface region of the peptide and the protein were treated as fully flexible, while simultaneously optimizing the center of mass position and orientation of the peptide.

## Molecular dynamics

**Preparation and parametrization.** We used *pdb4amber* from Amber16 [45] to make the pdb files of 104 high confidence protein-peptide complexes top models suitable for using this software package. All disulfide bonds detected by *pdb4amber* were retained in the system. Detected protein gaps were treated by addition of *N*-methyl group (NME) to the carbon of the backbone amide group of the C-terminal, and acetyl group (ACE) to the backbone nitrogen of the N-terminal amino acid, using PyMol Molecular Graphics System, Version v1.7.4.4, Schrodinger, LLC [46]. Capping prevents the amino acids that are flanking the gap from being recognized as protein termini, and therefore charged.

Parametrization of the systems was done using *tleap* from Amber16. Counter-charged ions ( $\text{Na}^+$  or  $\text{Cl}^-$ ) were added to the non-neutral systems, and each protein-peptide complex was surrounded by a rectangular box of explicit TIP3P water spanning 10 Å from the system. Force field *ff14SB* was used for parametrization of proteins and peptides, and *tip3p* for parametrization of water. Joung/Cheatham parameters were employed for monovalent ions in the chosen water type.

**Optimization.** Systems were optimized in 25,000 steps divided in five cycles, using *sander* from Amber16. First 1,000 steps of each cycle were performed by steepest descent method, while conjugated gradient was used for the remaining steps. In first three cycles, the constraint was applied to 1. the entire protein, 2. heavy protein:peptide atoms, and 3. backbone atoms, using force constant  $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . Constraint on backbone atoms was reduced to  $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  in the fourth cycle, and no constraints were applied in the fifth.

**Molecular dynamics simulations.** After optimization, each system was equilibrated during the initial 500 ps, using *pmemd* from Amber16 package. In the first 300 ps, the canonical NVT ensemble was simulated, with constraint applied to atoms in the protein:peptide complex using the force constant  $25 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . Temperature was increasing from 0 to 300 K during the first 250 ps. In the last 200 ps of equilibration, isothermal-isobaric ensemble NpT was simulated, with temperature held constant at 300 K and pressure at 1.0 bar, with no constraints applied to the system. Throughout equilibration, the SHAKE algorithm was used to apply constraints on bonds containing hydrogen atoms, and time step of 2 fs was used. The cutoff distance for non-bonded interactions was set to 15 Å, and the neighbor list was updated each 20 steps.

Production phase was done as a 4.5 ns continuation of the 500 ps long equilibration, using Gromacs 5 software [47–54]. Conversion from Amber to Gromacs file formats was performed with the help of ParmEd 2.7 tool (<https://github.com/ParmEd/ParmEd>). Constraint on bonds that contain hydrogen atoms was applied using LINCS algorithm, the time step was 2 fs, and the coordinates were written each picosecond. The temperature and pressure were kept at 300 K and 1.0 bar using modified Berendsen thermostat for temperature, and Parrinello-Rahman

barostat for pressure coupling. Particle mesh Ewald method was used for electrostatic interactions, the cutoff distance for non-bonded interactions was 12 Å, and the neighbor list was updated each 20 steps. Periodic boundary conditions were applied throughout equilibration and production phase.

**Analysis and binding energy calculation.** The obtained trajectories were visualized by Visual Molecular Dynamics VMD program [55], and tools from Gromacs package were used to correct for periodic boundary conditions and calculate root mean square deviation (RMSD) of complexes' backbones. Matplotlib [56] was used to visualize the results of analyses.

Molecular Mechanics energies with Generalized Born and Surface Area continuum solvation (MM/GBSA) method was used to calculate the Gibbs energy of protein:peptide binding in the 104 top docking models. The binding energy is calculated as the following average:

$$\Delta G_{bind} = \langle G_{protein:peptide} - G_{protein} - G_{peptide} \rangle_{protein:peptide} \quad (1)$$

with each Gibbs energy term being the following sum:

$$G = E_{bnd} + E_{el} + E_{vdw} + G_{pol} + G_{np} - TS \quad (2)$$

where the bonded, electrostatic and van der Waals interaction energies terms are obtained by molecular mechanics, the polar solvation term by generalized Born, the non-polar solvation term from linear relation to the solvent accessible surface area, while the entropy term is often omitted [57], as is in this study.

Amber *MMPBSA.py.MPI* was used here to calculate  $\Delta G_{bind}$  for protein:peptide systems by MM/GBSA method, using a single trajectory of the complex. The topology files of dry complexes, as well as ligand (peptide) and receptor (protein), were prepared with Amber *ante-MMPBSA.py*. The Gibbs energy terms in Eq (1) were calculated for 100 conformational snapshots from the last 2.5 ns of the production phase for each system, using salt concentration of 0.15 mol dm<sup>-3</sup>. During the MM/GBSA calculations, per-residue binding energy decomposition was also performed in order to get insight into contributions of specific protein and peptide residues to binding.

**Amino acids contribution to binding.** The output files of the MM/GBSA per-residue energy decomposition were used to analyze the characteristics of protein:peptide binding. In each of the 104 systems, the residue with the largest contribution to binding, either in stabilizing or destabilizing manner, was detected. The threshold was then set to 40% of its binding energy contribution value, and all residues that contributed more than the threshold in a given system were taken for the analysis, with taking into account whether amino acid belongs to protein or peptide. The number of appearances of individual amino acid was then calculated, as well as average binding contribution of different amino acids, separately for proteins and peptides, using Python.

**Characteristics of SIPs and *A. thaliana* proteins.** We scanned SIPs for hydrophobicity using the grand average of hydropathy (GRAVY) number, which is a measure of the hydrophobicity/hydrophilicity of a protein based on Kyte and Doolittle equation. The hydropathy values range from -2 to +2 for most proteins, with the positively ranked proteins being more hydrophobic.

Gene ontology (GO) analysis was performed using the ClueGo Cytoscape plugin [58]. Lists of 835 unique proteins from the initial screening analysis were mapped to corresponding UniProt IDs using mappings from SIFTS database [59] ([www.ebi.ac.uk/pdbe/docs/sifts/index.html](http://www.ebi.ac.uk/pdbe/docs/sifts/index.html)). The list of proteins was used to query REACTOME\_Pathways and GO\_BiologicalProcess ontology and the type of evidence set was All\_experimental. Pathways with *p*-values ≤ 0.05 were displayed, the minimum GO tree interval was set as 3 and the maximum level was set as 8,

the GO term/pathway selection was set as a threshold of 4% of genes per pathway and the kappa score was set as 0.4.

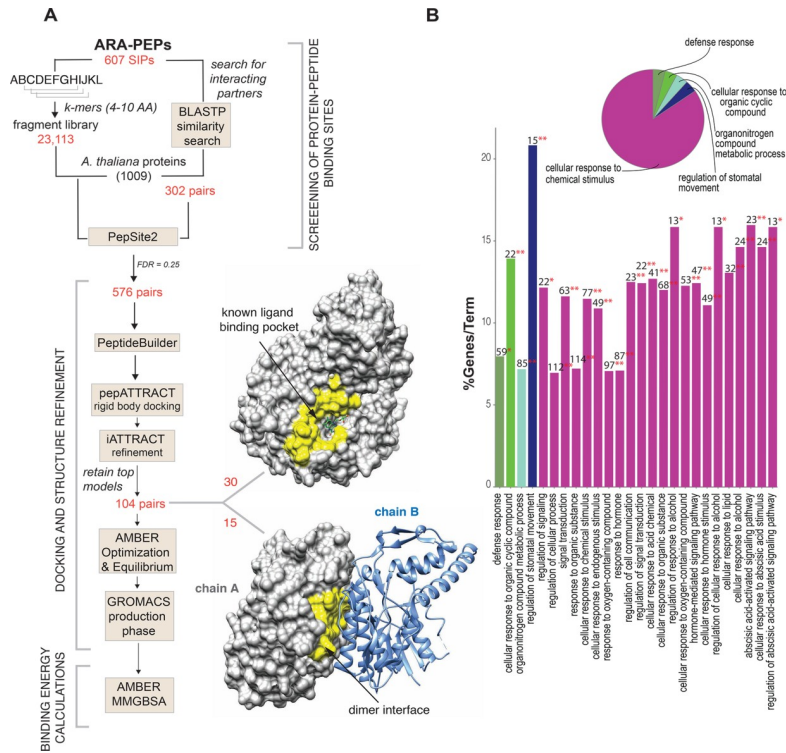
**Analysis of interactions at the protein-peptide interface.** We manually inspected the top 10 models for each docked protein-peptide pair predicted by pepATTRACT using molecular visualization softwares UCSF Chimera [60] and PyMOL Molecular Graphics System. PDBeMotif, a web server for checking the PDB structure for ligands and binding sites [61] and Catalytic Site Atlas, a database of enzyme active sites and catalytic residues on enzymes [62] was used for finding ligand/cofactor binding sites and enzyme active sites respectively. We analyzed if a specific peptide binding site lies at the interface of multi-chain proteins and assumed that residues on the 2 chains less than 6.0 Å apart were interacting residues. The distance between C $\alpha$  atoms located in chains A and B, with coordinates  $A(x_1, y_1, z_1)$  and  $B(x_2, y_2, z_2)$ , was calculated according to the Euclidean distance equation  $D(A, B) = \sqrt{\{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\}}$ . All calculations were performed using the Biopython package from Python.

Protein-peptide bindings were characterized using BINDing ANALyzer (BINANA) [63], HBPLUS [64] and Protein-Ligand Interaction Profiler (PLIP) [65] tools. BINANA was used to characterize important protein-ligand interactions such as close contacts (any receptor atom within 4.0 Å of the ligand atoms), hydrogen bonds (distance cutoff = 4.0 Å and angle cutoff  $\leq 40^\circ$ ), hydrophobic contacts (ligand carbon atom within 4.0 Å of a receptor carbon atom), salt bridges and pi-pi interactions.

## Results

### Short peptides may exert their function by interacting with proteins

In a previous study, we identified 189 TARs in response to plant oxidative stress by the herbicide Paraquat and the fungus *Botrytis cinerea*, which could be translated into 607 SIPs [16,17]. A peptide fragment library consisting of 23,113 k-mers, ranging from 4 to 10 AA residues, was generated and searched for potential binding sites on *A. thaliana* proteins from the PDB repository, using PepSite2. 12,540,140 protein-peptide pairs were screened, and multiple testing correction was performed using Benjamini-Hochberg correction method [41] on all  $p$ -values with FDR set as 0.25. We have used a fairly high FDR of 25% because our study is an initial exploratory step to identify potential protein-peptide bindings and using a rather relaxed FDR would help us to not leave out any interesting candidates. We additionally screened for short peptide motif matches on *A. thaliana* proteins using BLASTP with default E-value  $\leq 10$  and found 302 matches with sequence identity ranging from 20% to 100%. The reasoning for finding peptides that show sequence similarity with a protein chain is that, an entire or a part of a SIP may mimic a specific binding motif on the protein, or resemble a loop from a large structured protein, a disordered region in protein termini or interfaces between defined domains and modulate the protein's activity [18,20]. A feasible subset for initial docking analysis was devised by pooling together 576 pairs comprising of the above 302 matches and randomly selecting 274 other pairs significant below the FDR threshold (Figs 1A and S1A). The list of docked pairs can be accessed through the url ([https://www.biw.kuleuven.be/CSB/ARA-PEPs/SIP\\_PDB\\_interactions.php](https://www.biw.kuleuven.be/CSB/ARA-PEPs/SIP_PDB_interactions.php)) (S1 Table). In our study, 46 protein-peptide complexes failed to dock, and the reason could be the large conformational changes of the protein caused by binding of a flexible peptide; this indeed remains a big problem of docking based methods [35]. From among the docked models, we selected 104 top protein-peptide pairs by manually inspecting models with most negative ATTRACT force field energy followed by calculation of free energy of binding using molecular dynamics simulations (S2 Table). Our results show that there exists a huge repertoire of potential peptide binding sites on several *A. thaliana* proteins which remain to be experimentally validated. Large-scale



**Fig 1. Overview of screening pipeline for protein-peptide pairs and enrichment analysis of proteins.** (A) Workflow showing the screening of 23,113 k-mers for binding sites on 1009 *A. thaliana* proteins. In total 12,540,140 protein-peptide pairs were screened, and multiple testing was performed on all *p*-values with False Discovery Rate (FDR) set as 0.25. A subset of 576 pairs were docked, and 104 pairs were further studied in detail. 30 peptides may bind to a ligand binding/catalytic site on a protein and 15 peptides may bind at the dimer interface between 2 chains of a protein complex. The peptide binding pocket is highlighted in yellow. (B) Histogram showing specific GO terms related to the associated proteins from protein-peptide screening analysis. The bars represent the number of proteins from the analyzed cluster associated with the term, and the label displayed on the bars is the percentage of proteins compared to all proteins associated with the term. The overview pie-chart presents functional groups for the proteins where the name of the group is given by the most significant term in the group. GO enrichment analysis revealed 5 main groups and each group section in the pie-chart correlates with the number of terms in each group.

<https://doi.org/10.1371/journal.pone.0205179.g001>

predictions of potential protein-peptide pairs can aid in future experimental studies for understanding cell-to-cell communication during plant development or stress-tolerance mechanisms.

Specific inhibitors may mimic portions of protein interfaces and can bind to a peptide binding pocket located at the interface between two monomers. In our study we found that 15 peptides bind at the interface between subunits of protein complexes around 6.0 Å apart, indicating the ability to modulate complex’s activity (Fig 1A and S3 Table). Among the 15 models, in three of the cases the peptides bind in a similar way to known characterized short peptides or portion of a full-length protein (S4 Fig). We also observed that 30 peptides bind to a known ligand/cofactor binding site (Fig 1A and S4 Table). Ligand and protein binding sites may often overlap within protein families as it has been shown that a peptide may compete with the ligand for the binding site or non-competitively bind to the pocket along with the ligand molecule. The possibility that several peptides may simply interact with known ligand/cofactor binding pockets could also be a result of PepSite2 having bias towards known larger pockets in a way similar to many other data driven methods which use existing, characterized proteins as templates and often are not well-suited to predict features that are missing in their template library [66].

While we observe that both pepATTRACT-local and pepATTRACT-blind produce similar results for 56% of the docked pairs, the other 44% of the pairs were not docked at the same site of the protein using the two methods (S2A Fig). These results show that the use of restraints in docking protocols can help in concentrating the search around relevant regions of the protein-peptide interaction space especially when experimental restraints are not available. Other reports have previously shown that restraint-based dockings yield better results as compared to blind docking methods [67,68].

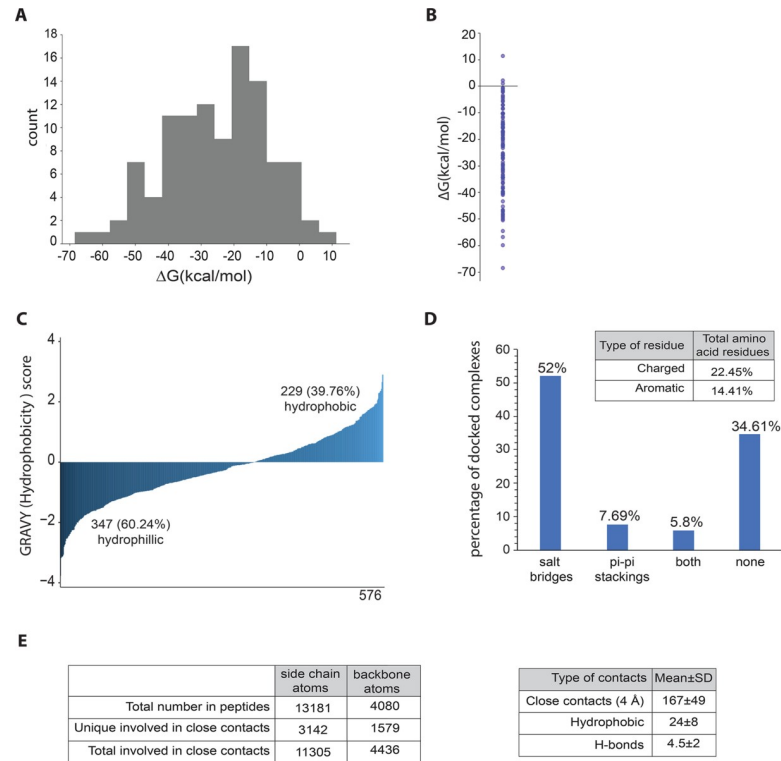
In addition, we investigated if protein pockets could bind multiple peptides, or in other words, whether a peptide prefers to bind to one specific pocket on a protein. To test this, we generated a randomly shuffled list of peptides while keeping the list of PDB structures intact, followed by scanning for binding sites using PepSite2. Among the protein structures that did bind a peptide, in 95.15% of the cases, peptides bound to the same protein pocket while in 4.85% cases peptides bound to a different pocket on the same protein (S2B Fig). It is possible that the S-PSSMs capture the binding modes of amino acids in such a way that amino acids in a peptide sequence may prefer to bind to chemically similar binding sites on proteins, e.g. hydrophobic amino acids from the peptide tend to bind hydrophobic protein regions [21] of appropriate sizes. While some reports suggest that peptides often look for a large enough pocket to bind followed by latching onto it with the help of a few hotspot residues [18,69], other reports suggest that several different peptides are able to bind to the same protein domain by exhibiting special properties such as promiscuity [70]. Furthermore, the seemingly more important role of peptide backbone compared to side chain atoms (detailed below) in protein binding provides another explanation for the observed promiscuity, as backbone atoms are the same independent of the amino acid sequence of the peptide.

We mapped 835 unique *A. thaliana* protein chains from the initial screening to Uniprot IDs using annotations from SIFTS database [59] and performed Gene Ontology (GO) enrichment analysis using REACTOME\_Pathways and GO\_BiologicalProcess ontology. GO analysis of the *A. thaliana* proteins with significant scores revealed that they may be categorized into 5 main groups *viz.* defense response, cellular response to organic cyclic compound, organonitrogen compound metabolic process, regulation of stomatal movement and cellular response to chemical stimulus (Fig 1B).

## Characterization of protein-peptide binding interactions

We carried out a general characterization of protein-peptide interactions in the 104 docked pairs. We extracted top 10 models from the docking results after iATTRACT refinement and calculated the average number of interactions within each docked pair. Each receptor atom that comes within 4 Å of any ligand atom is considered as a close contact. We determined the mean number of close contacts per docked pair to be  $167 \pm 49$  (S1B Fig). All protein-peptide pairs interacted with each other using hydrogen bonds and hydrophobic interactions. The mean number of hydrophobic interactions per system is  $24 \pm 8$ , and the mean number of hydrogen bonds per docked pair is  $4.5 \pm 2$ , where donors of hydrogen bonds are localized on protein in 51% of the cases (Fig 2E). While hydrogen bonds and hydrophobic interactions are omnipresent, not all pairs formed salt bridges and pi-pi stackings (Fig 2D and 2E). Within the peptides in our top 104 models, 22.5% of the total amino acid residues are charged (Arg, Lys, His, Asp, Glu) and 14.4% residues are aromatic (His, Phe, Tyr, Trp). In agreement with the amino acid composition, we also found salt bridges to be more prevalent than pi-pi stackings in the docked systems: 52% of protein-peptide pairs contain salt bridges with the mean number per system  $1.72 \pm 1$ , and 7.7% form pi-pi stackings with the mean of  $1.63 \pm 0.7$ . An additional 5.8% of pairs have





**Fig 2. Characterization of peptides and protein-peptide binding interactions.** (A) Histogram showing  $\Delta G_{\text{bind}}$  values for 104 top models (B)  $\Delta G_{\text{bind}}$  values as individual data points (C) Hydrophobicity index for all the 576 peptides predicted to interact with *A. thaliana* proteins (D) Total number of charged and aromatic residues in the peptides that interact with proteins. The plot also shows the total number of salt-bridges and pi-pi stackings formed in the top models. (E) Different types of interactions formed by the protein-peptide pairs.

<https://doi.org/10.1371/journal.pone.0205179.g002>

both salt bridges and pi-pi stackings, while the remaining 34.6% do not form any salt bridges or pi-pi stackings (Fig 2D).

In total, 23.8% of peptide side chain and 38.7% of backbone atoms participate in close contacts, with an average number of contacts per interacting atom being 3.6 and 2.8, respectively (Fig 2E). In average, the ratio of the unique peptide side chain:backbone atoms involved in close contacts is 2:1 for the top model in 104 docked systems, while the overall side chain:backbone ratio of atoms is 3.2:1. If side chain and backbone atoms of the peptide were equally important in protein-binding, we would expect the ratio of atoms involved in close contacts to also be 3.2:1. Instead, its lower value suggests that peptide backbone atoms might be more important in protein-peptide interactions than the side chain ones.

We determined the overall hydrophobicity index for all the 576 peptide fragments and found that 60.2% of the peptides are hydrophilic, while the remaining 39.8% are hydrophobic in nature (Fig 2C). A larger fraction of SIPs in our dataset have high hydrophilicity or a lower GRAVY index score, suggesting that they may mainly interact with globular proteins rather than with hydrophobic regions that spans membranes. Peptides with fewer ionic/charged groups are generally less soluble in water and are therefore prone to aggregation and interacting with hydrophobic pockets of larger proteins.

After analysis of the docked models, we took a further look into the dynamics of the top model from each of the 104 protein-peptide docked pairs and calculated the free energy of binding based on 100 conformational snapshots from molecular dynamics for each system

(Fig 2A and 2B). Per-residue decomposition of protein-peptide binding energies also allowed identification of amino acid types that frequently (in multiple systems) have significant binding contribution (S3A Fig). For instance, arginine residue, located in proteins at the peptide binding interface, stands out as a recurring amino acid with significant stabilizing effect on the binding (large value of x-axis and negative value of the  $\Delta G_{\text{bind}}$  contribution in S3A Fig). Interestingly, an equivalent prevalent contribution of negatively charged peptide amino acids is seemingly lacking (seen as a lack of green data points with large x values belonging to Glu and Asp in S3A Fig). Visual investigation of trajectories obtained by molecular dynamics shows that arginine charge is instead stabilized by salt bridges with negatively charged carboxyl groups of peptide C-terminal amino acids in 70% of cases (31 out of 44 prominent protein Arg residues), making this interaction independent of amino acid type present in the peptide. Other prominent protein residues that predominantly stabilize interactions with the peptides are the charged (Glu, Asp) and aromatic ones (Trp, Phe, Tyr).

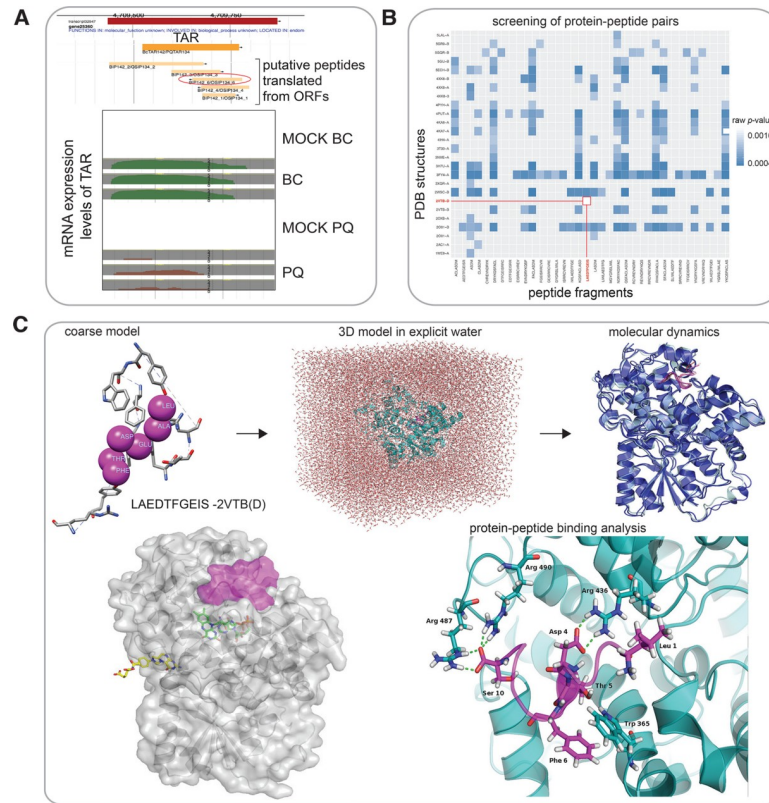
Local destabilizing effect on binding is shown by different peptide amino acids, containing side chains of largely different properties (S3A Fig). However, a more detailed view reveals that this is a consequence of amino acid location within a peptide, rather than its chemical composition (S3B Fig). In average, non-terminally located amino acids contribute to the binding in a stabilizing manner, N-termini destabilize protein-peptide interaction, while C-terminal amino acids have different average effect depending on amino acid type, and rarely have significant contribution (S3B and S3C Fig). Positively charged arginine residue in peptide is an interesting example: its overall contribution is stabilizing (S3A Fig) but depending on its position within the peptide it shows effects that range from stabilizing to destabilizing (S3B Fig). The same holds true for several other amino acids.

Overall, the largest destabilizing factor in binding across the 104 top protein-peptide models is the inability of protein to stabilize the N-terminal positively charged amino group of the peptide. However, the negative  $\Delta G_{\text{bind}}$  values for almost all systems (101 out of 104; Fig 2A and 2B) show that this factor is insufficient to destabilize the overall binding of the peptide to the predicted part of the protein.

### Peptide BIP142\_3/OSIP134\_3 may bind to CRYD protein

The 10-mer peptide fragment LAEDTFGEIS from BIP142\_3/OSIP134\_3 can be translated from BcTAR142/PQTAR134, expressed under stress conditions involving either *B. cinerea* or Paraquat. The above TAR is solely expressed under stress conditions and shows no expression under mock treatments (Fig 3A). The full-length peptide sequence of BIP142\_3/OSIP134\_3 (41 AA) was split into smaller fragments (<10-mers) and searched for protein partners. Bindings with adjusted *p*-values  $\leq 0.05$  were retained (Fig 3B). Additionally, BLASTP screenings revealed that the stretch of BIP142\_3/OSIP134\_3 (AA position: 9 to 32) show sequence similarity (46%) to cryptochrome DASH (CRYD) protein (chain D in 2VTB PDB structure), hence this peptide fragment may potentially act as a peptide mimic or affect the activity of the protein by binding to one of its pockets (Fig 3B). Members of the cryptochrome DASH subclade are involved in the DNA repair of cyclobutane pyrimidine dimers in single stranded DNA [71]. Cryptochromes in general are photolyase-like flavoproteins that mediate blue-light regulation of gene expression and photomorphogenic responses, including abiotic stress responses in *Arabidopsis*, as well as in all kingdoms of life [72].

The peptide LAEDTFGEIS bound to *A. thaliana* CRYD shows one of the strongest bindings among the 104 top models, with  $\Delta G_{\text{bind}}$  value of  $-48.02 \text{ kcal mol}^{-1}$ . Several residues in this complex have large contribution to this binding (represented as sticks in Fig 3C), with all except N-terminal leucine of the peptide contributing in a stabilizing manner. Protein and peptide are



**Fig 3. Overview of short peptide LAEDTFGEIS from stress induced BcTAR142/PQTAR134.** (A) BcTAR142/PQTAR134 shows mRNA expression levels under treated (PQ and BC) and mock conditions (mock\_BC and mock\_PQ). (B) Screening of all possible peptide fragments from BIP142\_3/OSIP134\_3 against all *A. thaliana* proteins in PDB. The LAEDTFGEIS-2VTB (Chain D) model indicated with a red line in the Figure. (C) A coarse model of peptide LAEDTFGEIS bound to CRYD protein (chain D of 2VTB). Restraint based docking was performed, followed by surrounding of the 3D model with explicit water. The solvated structure was optimized and then used for MD simulation. The conformational snapshots from the MD were used to calculate  $\Delta G_{\text{bind}}$  value for protein-peptide binding, and for visual inspection of the mode of binding (details in the text). Finally, superposition of the docked model (CRYD as grey, and docked LAEDTFGEIS as magenta surface) and the original PDB structure, which has FAD (green) and MTHF (yellow) co-factors bound, suggests that peptide might have an effect on FAD binding, and consequently on CRYD's function.

<https://doi.org/10.1371/journal.pone.0205179.g003>

bound via various types of interactions: salt bridges (Arg 436 and Asp 4; Arg 487, Arg 490 and Ser 10), stacking interactions (Trp 365 and Phe 6) and hydrogen bonds (Trp 365 and Thr 5). Destabilizing effect of N-terminal peptide residue, found in multiple other systems as well (S3B Fig), is likely the consequence of the lack of a negatively charged residue at the corresponding position in the protein, which would make favourable interactions with N-terminal amine group.

Two cofactors can bind to CRYD: flavin adenine dinucleotide (FAD) and 5,10-methenyltetrahydrofolate (MTHF), out of which the first one is necessary for catalytic activity. According to UniProt database [73], amino acids Arg 436 and Asp 485, which coincide with LAEDTFGEIS binding site, are involved in ATP binding. If the peptide (represented as magenta surface in Fig 3C) indeed binds CRYD (represented as grey surface in Fig 3C) in a way predicted in this study, it could block FAD binding or even bind simultaneously with it, therefore having an effect on the activity of this enzyme, and consequently on the aforementioned type of DNA repair (Fig 3C).

## Discussion

Steroids, peptides and other small bioactive compounds mainly regulate cellular communication in eukaryotes, including plants. Over the last decade, an increasing number of secreted peptides have been shown to influence a variety of developmental processes in plants, such as meristem size, root growth, stomatal differentiation, and organ abscission [74]. sORFs that might encode peptides have been overlooked in gene prediction programs owing to their small size. Moreover, there exist only a handful of publicly available T-DNA insertion collections of peptide encoding genes [75,76]. In this study, we predict that a large number of SIPs in *A. thaliana* may exert their function through protein-peptide interactions, by binding on protein surfaces. We found 30 peptides that may bind at known ligand/cofactor binding sites on proteins. The identification of ligand/cofactor binding sites in protein structures can aid in determination of peptide ligand types and experimental validation of the function of the receptor [77]. A peptide may compete for the ligand-binding site on the receptor, or it may non-competitively bind to the pocket together with the ligand molecule and play a role in modulating the receptor. Additionally, we screened 15 peptides that may bind to a pocket at the interface between two monomers of a multi-chain complex. The design of peptides and peptidomimetics that mimic portions of dimeric/multimeric protein interfaces have been shown to be a useful approach for the discovery of inhibitors that bind at protein-protein interfaces [78]. Currently, there is a lot of interest in drugs that can inhibit dimerization of a functionally obligate homodimeric enzyme. However, design of peptides that may disrupt protein-protein interactions is far more challenging than designing enzyme active site inhibitors, due to factors such as the large interfacial areas involved, and flat and featureless topologies that these binding surfaces may exhibit [79].

The characterization of protein-peptide interactions can be used to evaluate the binding affinity of the model. One of the major factors determining the binding of a peptide to a protein is the size of the pocket [80]. In our study, binding analysis of the predicted protein-peptide pairs revealed that different peptides tend to bind in the same pocket of one protein. We also observed that the peptides mostly interact with proteins with the help of side chains, but this is due to the reason that we have 3.2 times more of side chain atoms than backbone atoms. However, the peptide backbone atoms participate in more unique protein-peptide interactions as compared to the side chains. This is in agreement with another finding where peptides use more H-bonds in binding to their protein partner involving the peptide backbone. In the *PepTiDB* dataset comprising of 103 protein-peptide complexes, 19 peptides bind as  $\beta$ -strands, which use far more H-bonds on average, while 18 peptides were bound as  $\alpha$ -helices, which form less H-bonds with proteins and contain more nonpolar atoms at the interface [18].

The pepATTRACT-local docking method has advantages over other protein-peptide docking methods. First, pepATTRACT-local docking outperforms blind docking whose performance is similar with other local docking methods [19]. Second, this approach completes a run in about one hour for each pair, which is beneficial for large-scale prediction of protein-peptide interactions. In our study, 46 protein-peptide pairs failed to dock. The reason may be due to large conformational changes upon peptide binding onto the receptor, which still remains a huge problem while trying to accurately predict interactions [35]. Some failed cases reveal that the peptide is deeply buried into the protein surface. These failed pairs can be docked using other local docking methods. However, for most of protein-peptide interactions, only very small conformational changes upon peptide binding have been observed on the protein surface.

The *A. thaliana* genome may encode thousands of small proteins that could function as peptide signals and more than 600 plasma membrane-bound receptor-type proteins that could

act as receptors for peptide ligands [81]. Several sORF-encoded peptides may target regulatory enzymes involved in metabolic pathways by downregulating or upregulating the activity of these key enzymes. Predicting potential protein-peptide pairs and confirmation of physical interaction between these pairs is crucial to advance our understanding of cell-to-cell communication during plant development or stress-tolerance mechanisms [2]. Apart from the CLE family peptides which have been quite comprehensively studied, there may be many more peptides that remain to be identified. In general, our study aims at predicting potential protein-peptide interactions on protein surfaces which can be experimentally validated by researchers in the future.

## Supporting information

**S1 Fig. Plots showing  $p$ -values and binding site predictions of screened protein-peptide pairs.** (A) Plot of adjusted  $p$ -values using Benjamini-Hochberg correction method vs. raw  $p$ -values for the 576 docked models. (B) Histogram showing distribution of the number of close contacts across the 104 top models.

(TIF)

**S2 Fig. Additional tests for understanding protein-peptide bindings.** (A) Comparison of pepATTRACT-local and blind docking protocols (B) Effect of random shuffling on the binding of peptides to pockets

(TIF)

**S3 Fig. Plots showing per residue decomposition of protein-peptide binding energies.** (A) Average contributions to the binding energy for each amino acid type, for peptide and protein amino acids separately, and (B) for peptide amino acids at different locations within the peptides. (C) Individual data points for all amino acids, from which the averages were made, with red lines representing the average values. The represented data includes only amino acids whose binding contribution is at least 40% of the maximal contribution value within the respective system.

(TIF)

**S4 Fig. Examples of protein-peptide models showing binding modes of characterized peptides and SIPs.** The peptide binding pocket is highlighted in yellow.

(TIF)

**S1 Table. List of 576 docked protein-peptide pairs.**

(PDF)

**S2 Table. List of top 104 protein-peptide pairs from among the 576 docked models.** The table includes the PepSite2  $p$ -values and  $\Delta G_{\text{bind}}$  values of the models.

(PDF)

**S3 Table. List of 15 protein-peptide models where the peptide binding pocket lies at the subunits interface of a protein complex.** For fields that are indicated as monomers in Protein stoichiometry, the other chain in the structure is either a characterized peptide or the monomers may biologically aggregate to form dimers.

(PDF)

**S4 Table. List of 30 protein-peptide models where the peptide-binding pocket overlaps with known ligand binding or catalytic sites.** Empty fields in the ligand column indicate only catalytic sites and no known ligand is known to bind at the respective sites.

(PDF)

## Acknowledgments

RRH, NS and YS performed the analysis. RRH and VvN guided the work. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation–Flanders (FWO) and the Flemish Government–department EWI.

## Author Contributions

**Conceptualization:** Rashmi R. Hazarika, Vera van Noort.

**Formal analysis:** Rashmi R. Hazarika, Nikolina Sostaric, Yifeng Sun.

**Funding acquisition:** Vera van Noort.

**Investigation:** Rashmi R. Hazarika, Nikolina Sostaric.

**Methodology:** Rashmi R. Hazarika, Nikolina Sostaric, Vera van Noort.

**Project administration:** Vera van Noort.

**Resources:** Vera van Noort.

**Software:** Rashmi R. Hazarika, Nikolina Sostaric.

**Supervision:** Vera van Noort.

**Validation:** Nikolina Sostaric.

**Visualization:** Rashmi R. Hazarika, Yifeng Sun, Vera van Noort.

**Writing – original draft:** Rashmi R. Hazarika, Nikolina Sostaric.

**Writing – review & editing:** Rashmi R. Hazarika, Vera van Noort.

## References

1. Grunewald W, Friml J. The march of the PINs: developmental plasticity by dynamic polar targeting in plant cells. *EMBO J.* 2010; 29: 2700–2714. <https://doi.org/10.1038/emboj.2010.181> PMID: 20717140
2. Murphy E, Smith S, De Smet I. Small Signaling Peptides in Arabidopsis Development: How Cells Communicate Over a Short Distance. *Plant Cell.* 2012; 24: 3198–3217. <https://doi.org/10.1105/tpc.112.099010> PMID: 22932676
3. Marín-González E, Suárez-López P. “And yet it moves”: Cell-to-cell and long-distance signaling by plant microRNAs. *Plant Sci. Elsevier Ireland Ltd;* 2012; 196: 18–30. <https://doi.org/10.1016/j.plantsci.2012.07.009> PMID: 23017896
4. Matsubayashi Y. Post-translational modifications in secreted peptide hormones in plants. *Plant Cell Physiol.* 2011; 52: 5–13. <https://doi.org/10.1093/pcp/pcq169> PMID: 21071428
5. Tavormina P, De Coninck B, Nikonorova N, De Smet I, Cammue BPA. The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions. *Plant Cell.* 2015; 27: 2095–2118. <https://doi.org/10.1105/tpc.15.00440> PMID: 26276833
6. Yamaguchi YL, Ishida T, Sawa S. CLE peptides and their signaling pathways in plant development. *J Exp Bot.* 2016; 67: 4813–4826. <https://doi.org/10.1093/jxb/erw208> PMID: 27229733
7. Amor B Ben, Wirth S, Merchan F, Laporte P, D’Aubenton-Carafa Y, Hirsch J, et al. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res.* 2009; 19: 57–69. <https://doi.org/10.1101/gr.080275.108> PMID: 18997003
8. Chen M, Chen J, Zhang D. Exploring the secrets of long noncoding RNAs. *Int J Mol Sci.* 2015; 16: 5467–5496. <https://doi.org/10.3390/ijms16035467> PMID: 25764159
9. Ladoukakis E, Pereira V, Magny EG, Eyre-Walker A, Couso JP. Hundreds of putatively functional small open reading frames in *Drosophila*. *Genome Biol.* 2011; 12: R118. <https://doi.org/10.1186/gb-2011-12-11-r118> PMID: 22118156

10. Crappé J, Van Crielinge W, Trooskens G, Hayakawa E, Luyten W, Baggerman G, et al. Combining in silico prediction and ribosome profiling in a genome-wide search for novel putatively coding sORFs. *BMC Genomics*. 2013; 14: 648. <https://doi.org/10.1186/1471-2164-14-648> PMID: 24059539
11. Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. Long non-coding RNAs as a source of new peptides. Tautz D, editor. *Elife*. eLife Sciences Publications, Ltd; 2014; 3: e03523. <https://doi.org/10.7554/eLife.03523> PMID: 25233276
12. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet*. 2014; 15. <https://doi.org/10.1038/nrg3520> PMID: 24514441
13. Campalans A, Kondorosi A, Crespi M. Enod40, a Short Open Reading Frame-Containing mRNA, Induces Cytoplasmic Localization of a Nuclear RNA Binding Protein in *Medicago truncatula*. *Plant Cell*. 2004; 16: 1047–1059. <https://doi.org/10.1105/tpc.019406> PMID: 15037734
14. Hashimoto Y, Kondo T, Kageyama Y. Lilliputians get into the limelight: Novel class of small peptide genes in morphogenesis. *Dev Growth Differ*. 2008; 50. <https://doi.org/10.1111/j.1440-169X.2008.00994.x> PMID: 18459982
15. Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, et al. enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J*. 1994; 13: 5099–5112. PMID: 7957074
16. Hazarika RR, De Coninck B, Yamamoto LR, Martin LR, Cammue BPA, van Noort V. ARA-PEPs: a repository of putative sORF-encoded peptides in *Arabidopsis thaliana*. *BMC Bioinformatics*. *BMC Bioinformatics*; 2017; 18: 37. <https://doi.org/10.1186/s12859-016-1458-y> PMID: 28095775
17. De Coninck B, Carron D, Tavormina P, Willem L, Craik DJ, Vos C, et al. Mining the genome of *Arabidopsis thaliana* as a basis for the identification of novel bioactive peptides involved in oxidative stress tolerance. *J Exp Bot*. 2013; 64: 5297–5307. <https://doi.org/10.1093/jxb/ert295> PMID: 24043855
18. London N, Movshovitz-attias D, Schueler-furman O. The Structural Basis of Peptide-Protein Binding Strategies. *Struct Des*. Elsevier Ltd; 2010; 18: 188–199. <https://doi.org/10.1016/j.str.2009.11.012> PMID: 20159464
19. Schindler CEM, De Vries SJ, Zacharias M. Fully Blind Peptide-Protein Docking with pepATTRACT. *Structure*. Elsevier; 2015; 23: 1507–1515. <https://doi.org/10.1016/j.str.2015.05.021> PMID: 26146186
20. Kilburg D, Gallicchio E. Recent Advances in Computational Models for the Study of Protein-Peptide Interactions. *Adv Protein Chem Struct Biol*. Elsevier Inc.; 2016; 105: 27–57. <https://doi.org/10.1016/bs.apcsb.2016.06.002> PMID: 27567483
21. Petsalaki E, Stark A, García-Urdiales E, Russell RB. Accurate prediction of peptide binding sites on protein surfaces. *PLoS Comput Biol*. 2009; 5. <https://doi.org/10.1371/journal.pcbi.1000335> PMID: 19325869
22. Neduva V, Russell RB. Linear motifs: Evolutionary interaction switches. *FEBS Lett*. 2005; 579: 3342–3345. <https://doi.org/10.1016/j.febslet.2005.04.005> PMID: 15943979
23. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. *Structure*. Elsevier Ltd; 2010; 18: 1233–1243. <https://doi.org/10.1016/j.str.2010.08.007> PMID: 20947012
24. Pawson T, Nash P. Assembly of Cell Regulatory Systems Through Protein Interaction Domains. *Science* (80-). 2003; 300: 445 LP–452.
25. Raveh B, London N, Schueler-Furman O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins Struct Funct Bioinforma*. 2010; 78: 2029–2040. <https://doi.org/10.1002/prot.22716> PMID: 20455260
26. Raveh B, London N, Zimmerman L, Schueler-Furman O. Rosetta FlexPepDock ab-initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors. *PLoS One*. Public Library of Science; 2011; 6: e18934. <https://doi.org/10.1371/journal.pone.0018934> PMID: 21572516
27. Ko J, Park H, Seok C. GalaxyTBM: template-based modeling by building a reliable core and refining unreliable local regions. *BMC Bioinformatics*. 2012; 13: 198. <https://doi.org/10.1186/1471-2105-13-198> PMID: 22883815
28. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res*. 2013; 41: 384–388. <https://doi.org/10.1093/nar/gkt458> PMID: 23737448
29. Lee H, Heo L, Lee MS, Seok C. GalaxyPepDock: A protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res*. 2015; 43: W431–W435. <https://doi.org/10.1093/nar/gkv495> PMID: 25969449
30. Ding F, Yin S, Dokholyan N V. Rapid Flexible Docking Using a Stochastic Rotamer Library of Ligands. *J Chem Inf Model*. American Chemical Society; 2010; 50: 1623–1632. <https://doi.org/10.1021/ci100218t> PMID: 20712341

31. Antes I. DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins Struct Funct Bioinforma.* 2010; 78: 1084–1104. <https://doi.org/10.1002/prot.22629> PMID: 20017216
32. 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. <https://doi.org/10.1093/nar/gkv456> PMID: 25943545
33. Wabik J, Kurcinski M, Kolinski A. Coarse-Grained Modeling of Peptide Docking Associated with Large Conformation Transitions of the Binding Protein: Troponin I Fragment–Troponin C System. *Molecules.* 2015; 20: 10763–10780. <https://doi.org/10.3390/molecules200610763> PMID: 26111167
34. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc.* 2003; 125: 1731–1737. <https://doi.org/10.1021/ja026939x> PMID: 12580598
35. Trellet M, Melquiond ASJ, Bonvin AMJJ. A Unified Conformational Selection and Induced Fit Approach to Protein-Peptide Docking. *PLoS One.* 2013; 8. <https://doi.org/10.1371/journal.pone.0058769> PMID: 23516555
36. Trabuco LG, Lise S, Petsalaki E, Russell RB. PepSite: Prediction of peptide-binding sites from protein surfaces. *Nucleic Acids Res.* 2012; 40: 423–427. <https://doi.org/10.1093/nar/gks398> PMID: 22600738
37. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000; 28: 235–242. <https://doi.org/10.1093/nar/28.1.235> PMID: 10592235
38. Stein A, Aloy P. Contextual specificity in peptide-mediated protein interactions. *PLoS One.* 2008; 3: 1–10. <https://doi.org/10.1371/journal.pone.0002524> PMID: 18596940
39. Groß A, Hashimoto C, Sticht H, Eichler J. Synthetic Peptides as Protein Mimics. *Front Bioeng Biotechnol.* 2016; 3. <https://doi.org/10.3389/fbioe.2015.00211> PMID: 26835447
40. Beekman AM, Howell LA. Small-Molecule and Peptide Inhibitors of the Pro-Survival Protein Mcl-1. *ChemMedChem.* 2016; 11: 802–813. <https://doi.org/10.1002/cmdc.201500497> PMID: 26696548
41. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B. [Royal Statistical Society, Wiley];* 1995; 57: 289–300.
42. Pollard KS, Dudoit S, van der Laan MJ. Multiple Testing Procedures: the multtest Package and Applications to Genomics BT—Bioinformatics and Computational Biology Solutions Using R and Bioconductor. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, editors. New York, NY: Springer New York; 2005. pp. 249–271. [https://doi.org/10.1007/0-387-29362-0\\_15](https://doi.org/10.1007/0-387-29362-0_15)
43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215: 403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
44. Tien MZ, Sydykova DK, Meyer AG, Wilke CO. PeptideBuilder: A simple Python library to generate model peptides. *PeerJ.* 2013; 1: e80. <https://doi.org/10.7717/peerj.80> PMID: 23717802
45. Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, et al. Amber16, University of California, San Francisco. 2017.
46. Delano WL. The PyMOL Molecular Graphics System. 2002.
47. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Mol Model Annu.* 2001; 7: 306–317. <https://doi.org/10.1007/s008940100045>
48. Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput.* American Chemical Society; 2008; 4: 116–122. <https://doi.org/10.1021/ct700200b> PMID: 26619985
49. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput.* American Chemical Society; 2008; 4: 435–447. <https://doi.org/10.1021/ct700301q> PMID: 26620784
50. David VDS, Erik L, Berk H, Gerrit G, E. MA, C. BHJ. GROMACS: Fast, flexible, and free. *J Comput Chem.* Wiley-Blackwell; 2005; 26: 1701–1718. <https://doi.org/10.1002/jcc.20291> PMID: 16211538
51. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun.* 1995; 91: 43–56. [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)
52. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys.* American Institute of Physics; 1995; 103: 8577–8593. <https://doi.org/10.1063/1.470117>
53. Miyamoto Shuichi & Kollman Peter A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. <https://doi.org/10.1002/jcc.540130805>
54. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys.* American Institute of Physics; 2007; 126: 14101. <https://doi.org/10.1063/1.2408420> PMID: 17212484



55. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996; 14: 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
56. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng*. 2007; 9: 90–95. <https://doi.org/10.1109/MCSE.2007.55>
57. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov. Informa Healthcare*; 2015; 10: 449–461. <https://doi.org/10.1517/17460441.2015.1032936> PMID: 25835573
58. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009; 25: 1091–1093. <https://doi.org/10.1093/bioinformatics/btp101> PMID: 19237447
59. Velankar S, Dana JM, Jacobsen J, Van Ginkel G, Gane PJ, Luo J, et al. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res*. 2013; 41: 483–489. <https://doi.org/10.1093/nar/gks1258> PMID: 23203869
60. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25: 1605–1612. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
61. Gutmanas A, Alhroub Y, Battle GM, Berrisford JM, Bochet E, Conroy MJ, et al. PDBe: Protein data bank in Europe. *Nucleic Acids Res*. 2014; 42: 308–317. <https://doi.org/10.1093/nar/gkt1180> PMID: 24288376
62. Porter CT. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res*. 2004; 32: 129D–133. <https://doi.org/10.1093/nar/gkh028> PMID: 14681376
63. Durrant JD, McCammon JA. BINANA: A Novel Algorithm for Ligand-Binding Characterization. *J Mol Graph Model*. 2011; 29: 888–893. <https://doi.org/10.1016/j.jmgm.2011.01.004> PMID: 21310640
64. McDonald IK, Thornton JM. Satisfying Hydrogen Bonding Potential in Proteins. *Journal of Molecular Biology*. 1994. pp. 777–793. <https://doi.org/10.1006/jmbi.1994.1334> PMID: 8182748
65. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Res*. 2015; 43: W443–W447. <https://doi.org/10.1093/nar/gkv315> PMID: 25873628
66. Gao M, Skolnick J. The distribution of ligand-binding pockets around protein-protein interfaces suggests a general mechanism for pocket formation. *Proc Natl Acad Sci*. 2012; 109: 3784–3789. <https://doi.org/10.1073/pnas.1117768109> PMID: 22355140
67. Vajda S, Kozakov D. Convergence and Combination of Methods in Protein-Protein Docking. *Curr Opin Struct Biol*. 2009; 19: 164–170. <https://doi.org/10.1016/j.sbi.2009.02.008> PMID: 19327983
68. de VS J., van DAD J., Mickaël K, Mark van D, Aurelien T, Victor H, et al. HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins Struct Funct Bioinforma*. Wiley-Blackwell; 2007; 69: 726–733. <https://doi.org/10.1002/prot.21723> PMID: 17803234
69. London N, Raveh B, Schueler-Furman O. Druggable protein-protein interactions—from hot spots to hot segments. *Curr Opin Chem Biol*. Elsevier Ltd; 2013; 17: 952–959. <https://doi.org/10.1016/j.cbpa.2013.10.011> PMID: 24183815
70. Bhattacharjee A, Wallin S. Exploring Protein-Peptide Binding Specificity through Computational Peptide Screening. *PLoS Comput Biol*. 2013; 9. <https://doi.org/10.1371/journal.pcbi.1003277> PMID: 24204228
71. Selby CP, Sancar A. A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proc Natl Acad Sci*. 2006; 103: 17696–17700. <https://doi.org/10.1073/pnas.0607993103> PMID: 17062752
72. Yu X, Liu H, Klejnot J, Lin C. The Cryptochrome Blue Light Receptors. *Arabidopsis Book*. American Society of Plant Biologists; 2010; 8: e0135. <https://doi.org/10.1199/tab.0135> PMID: 21841916
73. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, et al. UniProt: The universal protein knowledgebase. *Nucleic Acids Res*. Oxford University Press; 2017; 45: D158–D169. <https://doi.org/10.1093/nar/gkw1099> PMID: 27899622
74. Butenko MA, Aalen RB. Receptor Ligands in Development. In: Tax F, Kemmerling B, editors. *Receptor-like Kinases in Plants: From Development to Defense*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. pp. 195–226. [https://doi.org/10.1007/978-3-642-23044-8\\_11](https://doi.org/10.1007/978-3-642-23044-8_11)
75. Butenko M a, Wildhagen M, Albert M, Jehle A, Kalbacher H, Aalen RB, et al. Tools and Strategies to Match Peptide-Ligand Receptor Pairs. *Plant Cell*. 2014; 26: 1838–1847. <https://doi.org/10.1105/tpc.113.120071> PMID: 24808051
76. Lease K a, Walker JC. The Arabidopsis Unannotated Secreted Peptide Database, a Resource for Plant Peptidomics. *PLANT Physiol*. 2006; 142: 831–838. <https://doi.org/10.1104/pp.106.086041> PMID: 16998087

77. Glaser F, Morris RJ, Najmanovich RJ, Laskowski RA, Thornton JM. A method for localizing ligand binding pockets in protein structures. *Proteins Struct Funct Genet.* 2006; 62: 479–488. <https://doi.org/10.1002/prot.20769> PMID: 16304646
78. Cardinale D, Guaitoli G, Tondi D, Luciani R, Henrich S, Salo-Ahen OMH, et al. Protein–protein interface-binding peptides inhibit the cancer therapy target human thymidylate synthase. *Proc Natl Acad Sci.* 2011; 108: E542–E549. <https://doi.org/10.1073/pnas.1104829108> PMID: 21795601
79. Fletcher S, Hamilton AD. Targeting protein-protein interactions by rational design: mimicry of protein surfaces. *J R Soc Interface.* 2006; 3: 215–33. <https://doi.org/10.1098/rsif.2006.0115> PMID: 16849232
80. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. Protein clefts in molecular recognition and function. *Protein Sci.* 1996; 5: 2438–52. <https://doi.org/10.1002/pro.5560051206> PMID: 8976552
81. Shiu S-H, Bleecker a B. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci.* 2001; 98: 10763–10768. <https://doi.org/10.1073/pnas.181141598> PMID: 11526204