

## Detection of Peptide-Binding Sites on Protein Surfaces Using the Peptimap Server

Tanggis Bohnuud, George Jones, Ora Schueler-Furman,  
and Dima Kozakov

### Abstract

Peptide-mediated interactions are of primordial importance to the cell, and the structure of such interaction provides an important starting point for their further characterization. In many cases, the structure of the peptide-protein complex has not been solved by experiment, and modeling tools need to be applied to generate structural models of the interaction. PeptiMap is a protocol that identifies the peptide-binding site when only the structure of the receptor is known, but no information about where the peptide binds is available. This is achieved by mapping the surface for solvents to identify ligand-binding sites, similar in approach to ANCHORMAP in which amino acids are mapped. Peptimap is a free open access web-based server. It can be accessed at <http://peptimap.cluspro.org>.

**Key words** Peptide-protein interactions, Binding site prediction, Solvent mapping, Peptide mapping

---

### 1 Introduction

Peptide-mediated interactions are of primordial importance to the cell. It has been estimated that up to half of known protein-protein reactions are mediated by peptides in higher Eukaryotes (e.g., [1]). It is believed that peptides play an essential role in the modulation of these protein-protein interactions. As such, they are important drug targets, and it is therefore crucial to improve our understanding of these important players. The structures of such interactions provide good starting points for their further characterization. However, in many cases, such a structure has not been solved by experiment, and *in silico* modeling tools need to be developed and applied to fill the gap. These methods can be broken down into two main steps: the identification of possible binding spots on the protein surface, and the prediction of the peptide sequence and peptide pose in the binding spots. If the binding site is known, modeling protocols such as FlexPepDock [2, 3], peptide Haddock [4], pepCrawler [5], and others can be used to generate the full

atom model of the interaction (reviewed in [6], and covered by other chapters in this volume).

PeptiMap is a protocol that identifies peptide-binding sites when only the structure of the receptor is known. The approach is based on the experimental observation, in both crystal soaking and NMR experiments, that small organic molecules of varying size and polarity tend to accumulate on the protein surface in regions that bind larger ligands and proteins [7, 8]. FTmap is a Fast Fourier Transform (FFT)-based method for computational mapping of the surface for solvents [9] to identify ligand-binding sites [10, 11]. The approach is similar to ANCHORMAP in which amino acids are mapped [12]. In PeptiMap, the FTmap is modified to specify the search for peptide-binding sites. This chapter will introduce the reader to PeptiMap's functionality and the means to use the freely accessible PeptiMap server at <http://peptimap.cluspro.org>.

---

## 2 Materials

The following are needed to run PeptiMap on the server (at *peptimap.cluspro.org*; Fig. 1a):

1. An input structure of the protein receptor on which we would like to identify peptide-binding site(s) (in Protein Data Bank format, *see* [www.pdb.org](http://www.pdb.org) [13]). It can be uploaded, or provided as PDB id.
2. Chain(s) of the protein receptor to be mapped.

The output of the mapping will be available for visual inspection on the server (Fig. 1c), as well as a PyMOL session file for download and inspection using the molecular viewer PyMOL (PyMOL can be obtained at [www.pymol.org](http://www.pymol.org); Schrodinger LLC).

---

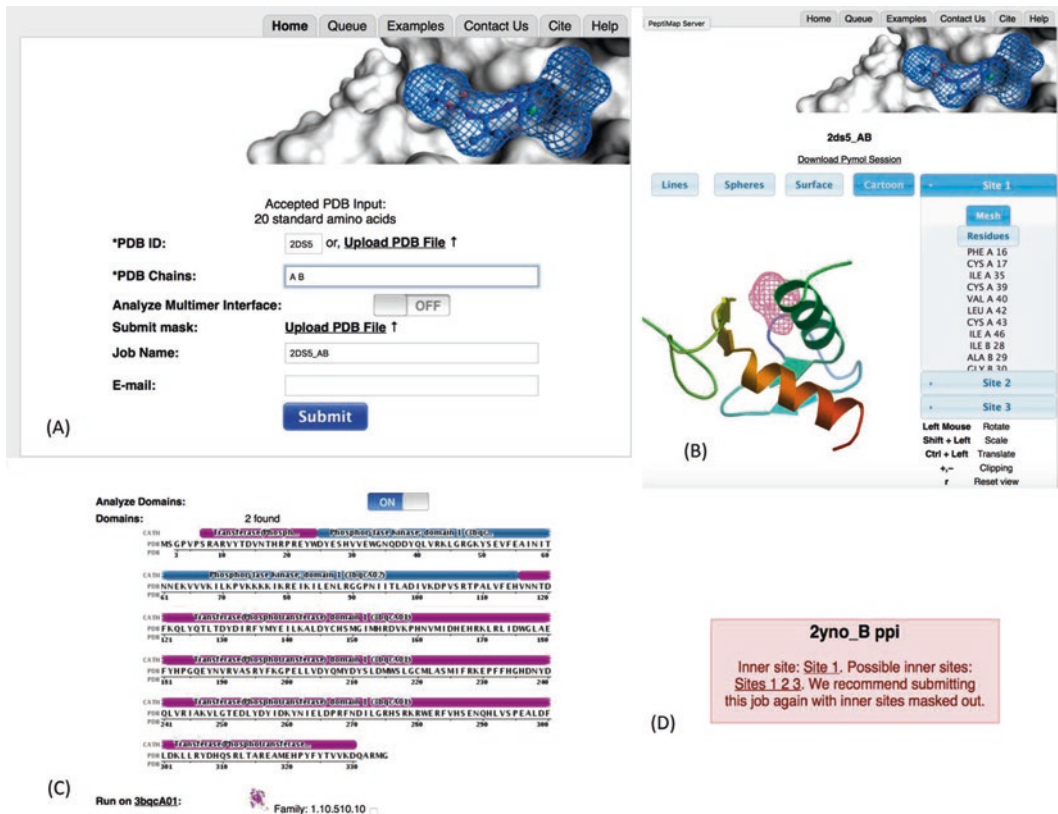
## 3 Methods

We first describe the setup of the PeptiMap Server, and detail the input options. We then present case studies to demonstrate the utility of PeptiMap. Finally, we provide more details about the underlying algorithms, as implemented in the server. Further details are also available in a previous publication on PeptiMap [14].

### 3.1 Input Interface

PeptiMap provides a simple interface with which to search for peptide-binding sites on a given receptor protein surface (*see* Fig. 1a). The following information is required/accepted as input:

1. PDB file to map, provided either in the form of a PDB id, or uploaded onto the server.
2. Chains to be investigated: indicate simply the chain(s) name, separated by spaces in case of multiple chains.



**Fig. 1** Overview of the PeptiMap server pages. (a) The main page where the job is submitted. (b) Results page for 2DS5 (see also Fig. 2): For each site, buttons allow to toggle the display of the predicted site (as mesh), as well as the binding residues on the protein receptor (shown as *sticks* and displayed as *list*). A Java applet allows changing the view of the protein, by scaling, rotating, and translating. In addition, the receptor can be shown as *cartoon* (default option), *sticks*, *spheres*, or *surface* representation. (c) Output of domain mapping option for 3BQC (see also Fig. 3). (d) Example for Output that detects inaccessible pockets for 2YNO, and suggests repeating the mapping using the provided PDB file for masking (download the PDB file and upload it in your new submission)

- For multi-chain structures: Map single or multiple chains (see **Note 1** for guidelines on what unit to map, and Subheading 3.3.2 on how these guidelines are implemented).
- For single-chain structures: Map specific domains or full receptor chain (currently the PDB must be provided as PDB ID for invoking this option). You can search the provided protein to identify (multiple) domains in the provided structure (see Subheading 3.3.3 and Fig. 1c). If the structure contains multiple domains, Peptimap provides the option to search the entire structure, or alternatively to focus on specified domains (see **Note 2** for guidelines/suggestions of what part of the structure to map).
- Mask option—removes from the search on the protein surface the atoms present in a provided mask PDB file. Such a file can

be provided, or alternatively, Peptimap will identify sites that are too small for peptides to bind, and suggest filtering them out, by automatically providing the mask file (*see* Subheading 3.2.3 and Fig. 1d). The criteria used to define such sites inaccessible to peptides are defined in Subheading 3.3.4.

Once the appropriate options for the desired search have been selected, an email address and job name can be entered. The email address allows Peptimap to inform the submitting party that the search has been completed. The results of the search will be displayed on the Peptimap site (*see* Fig. 1b), and can also be downloaded as a Pymol session.

## 3.2 PeptiMap Case Studies

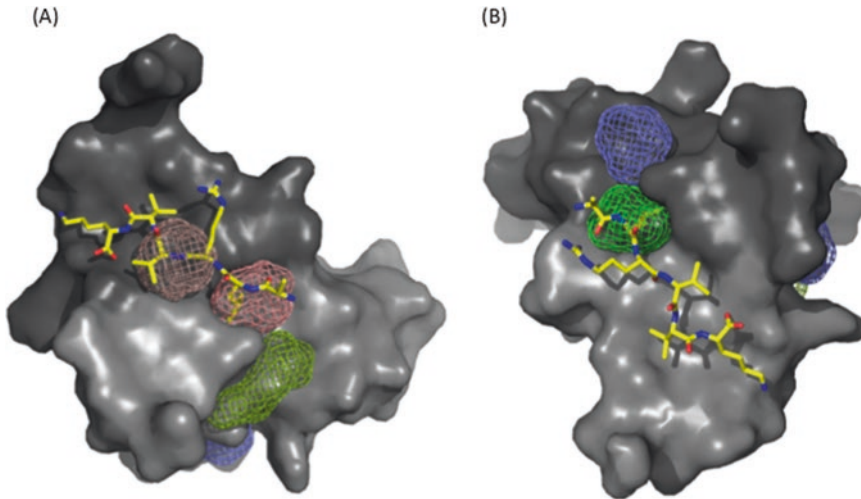
### 3.2.1 The Analyze Multimer Interface Option, Applied to 2DS5

In the following three examples, we demonstrate how the *Analyze Multimer Interface*, *Analyze Domain*, and *Submit Mask* options implemented in the PeptiMap server can be applied to focus and improve binding site prediction.

We will use the *Analyze Multimer Interface* option to determine whether or not the structure is a homo-multimer and if so, how many monomers should be included in the mapping unit. The example used is the Zinc Binding Domain of ClpX (PDB id 2DS5 [15]). In the following, we describe each step involved for the prediction of peptide-binding sites in detail.

1. Go to the site [peptimap.bu.edu](http://peptimap.bu.edu).
2. Locate the **PDB ID** field and type 2DS5.
3. In the **PDB Chains** field type A B.
4. Now click on the **Analyze Multimer Interface** button function. After a few seconds, a message should appear specifying the buried surface area between the individual chains, in this case:  
*A B are tight multimers (Buried SA: 42.6 % of monomer surface, >25 %). We recommend submitting the multimer.*
5. Once you have finished selecting options, specify an email where the notifications will be sent, and submit the job. This job took approximately 30 min to finish.

The result of this search revealed six binding hotspots on the protein surface (the top site is shown on the server page, and two additional sites can be displayed using the buttons to the right, Fig. 1b, while details of the top-6 sites are included in the PyMOL session). Figure 2 shows the corresponding protein bound to two copies of an SSPB tail peptide (2DS8 [15]), superimposed over the results generated by Peptimap. One peptide interacts with hotspots ranked 1 and 4, while a second peptide interacts with the predicted hotspot ranked 2. Thus, Peptimap was able to identify the binding sites of the two peptides.



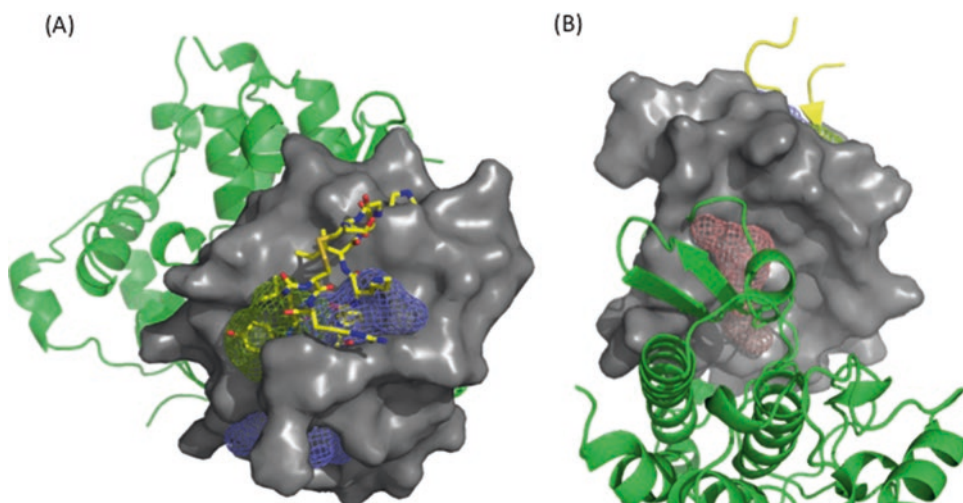
**Fig. 2** Mapping on a homodimer: the predicted binding site on homodimeric Zinc Binding Domain of ClpX (2DS5), overlapped with the structure of ClpX complexed with SspB-tail peptides ALRVVK (2DS8). The two domains are shown as surface, in different *shades of gray*. Two views are shown, to capture the two peptides (*yellow stick* representation): (a) In the first peptide, residues L2 and V4 overlap with the predicted binding hotspots ranked 1 and 4 (*red and pink*), respectively. (b) In the second peptide, residue L2 overlaps with the predicted site ranked 2 (*green*). This and the following figures were generated from the PyMOL sessions provided by the server. The binding sites are colored according to their ranking (ranks 1–5 *red-green-blue-pink-yellow* meshes)

### 3.2.2 The Analyze Domain Function, Applied to 3BQC

We will analyze the domain composition of the catalytic subunit of protein kinase CK2 to identify the domain to map for peptide-binding sites. The structure of this protein is available as PDB ID 3BQC, chain A [16]. Go to the site [peptimap.bu.edu](http://peptimap.bu.edu).

1. Locate the PDB ID field and type 3BQC.
2. In the PDB Chain field type A. This automatically defines the job name to 3bqc\_a.
3. Now click on the Analyze Domain button function. The domain composition according to CATH [17] will be displayed (Fig. 1c). For our case, this includes (a) the Transferase Phosphorylase family (CATH id 1.10.510.10) and (b) the Phosphorylase Kinase Family (CATH id 3.30.200.20).
4. In this example, we will look at one of the domains, the Phosphorylase Kinase Family domain. Click on the box next to this family: 3.30.200.20.
5. Once you have finished selecting options, specify an email where the notifications will be sent, and submit the job. This job took approximately 40 min to finish.

The result of this search revealed six binding hotspots on the protein surface. Figure 3 shows the bound protein-peptide complex (PDB ID 4IB5, human protein kinase CK2 catalytic subunit in a



**Fig. 3** Single domain mapping: shown are the predicted binding sites on protein kinase CK2 (3BQC), overlapped with the peptide-bound structure (to CK2 $\beta$ -competitive cyclic peptide GCRLYGFKIHGC; 4IB5). The mapped domain is shown as *gray* surface, the second domain as green cartoon. (a) View of the peptide with residues F7 and Y5 overlapping sites ranked 3 and 5 (*blue* and *yellow* meshes, respectively). (b) View of the domain interface, covering the top-ranking site (*red*), that could be masked

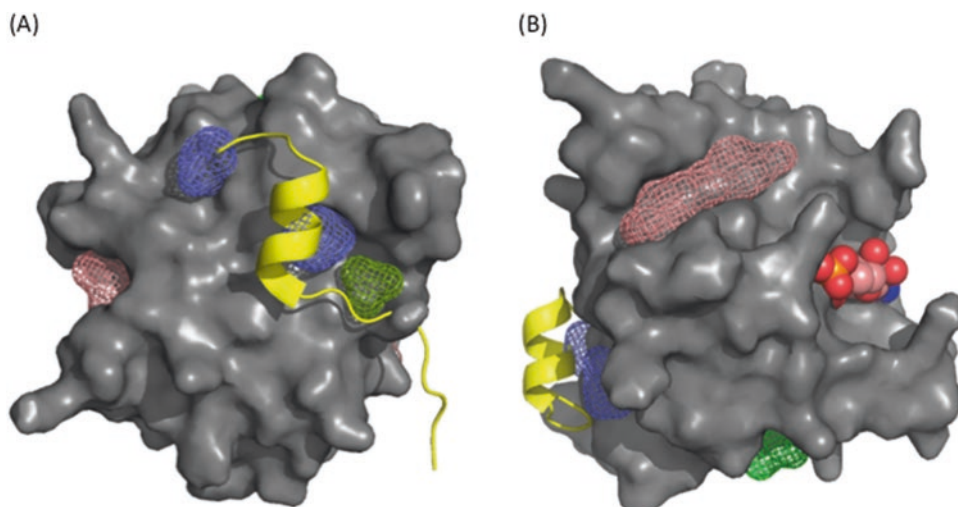
complex with a CK2 $\beta$ -competitive cyclic peptide [16]), superimposed onto the results generated by Peptimap. The bound peptide is clearly located within the predicted binding hotspots ranked 3 and 5 (Fig. 3a). Examination of the structure reveals the best-ranking site is inaccessible within the context of the second domain (Fig. 3b). (Note that to identify only sites outside this domain interface, a second search could be performed using the *Submit mask* option, see next Subheading).

### 3.2.3 The *Submit mask* Option, Applied to 1EJ1 and 2YN0

We will use the *Submit mask* function to remove from the search parts of the protein that will be inaccessible due to a ligand bound, on the example of the mRNA Cap-Binding interface of eIF4E bound to M7G (PDB id 1EJ1 [18]).

1. Go to the site [peptimap.bu.edu](http://peptimap.bu.edu).
2. Locate the *PDB ID* field and type 1EJ1.
3. In the *specify chains* field type A.
4. Using a text processor, remove from the PDB file all residues that do not contact the ligand M7G (i.e., keep only residues with at least one atom within 10 Å of M7G; remove also non-amino acid entities). Save the resulting PDB file.
5. Click on the *Submit mask* option and select the file that you just have generated.
6. Once you have finished selecting options, specify an email where the notifications will be sent, and submit the job. This job took approximately 32 min to finish.





**Fig. 4** Using a mask to exclude a ligand binding site: the predicted peptide-binding site on Eukaryotic Translation Initiation Factor 4E, eIF4E (1EJ1), overlapped with the structure of eIF4E complexed with the 4E-BP1 peptide (1WKW). **(a)** The peptide overlaps sites ranked 2, 3, and 5 (green, blue, and yellow, respectively). **(b)** Rotated view, showing the masked ligand binding site

The result of this search revealed six binding hotspots on the protein surface. Figure 4 shows the bound protein (PDB id 1WKW [19]) superimposed over the results generated by Peptimap. The peptide 4EBP1 interacts with hot spots ranked 2, 3, and 5.

Masking may also be used to skip deep binding pockets inaccessible to peptides. In the example of 2YNO, the propeller WD40 repeat domain of COP1 [20], such sites are automatically identified by the server, and a PDB file is provided for submission as masking file (Fig. 1d; see Subheading 3.3.4 for details how these sites are identified).

### 3.3 Details of Implementation

#### 3.3.1 Mapping of a Receptor Surface for Peptide-Binding Sites

We provide here a short overview of the protocol as implemented in the PeptiMap server. More details can be found in our previous publication of the PeptiMap algorithm [14].

Peptimap uses Fast Fourier Transform (FFT)-based grid-sampling to search the receptor surface for binding sites [14]. Water molecules and other ligands are removed prior to calculations. Sixteen small molecules are used as probes, as done previously for ligand mapping (see, e.g., Fig. 1b in [11]). These small molecules are varied enough to cover all the functional groups found in amino acids. The probes are allowed to adopt different conformations, which are all subjected to an FFT search to select favorable positions and conformations on the receptor surface. The 2000 best poses are then clustered using a greedy algorithm. This selects the lowest energy pose and incorporates all poses within 4 Å into the cluster (see [14]). The poses in the cluster are removed from the general

bin and the procedure is repeated with the lowest energy pose remaining. If a cluster contains less than ten probes it is removed. The remaining clusters are then minimized with the CHARMM potential, using the Analytic Continuum Electrostatic (ACE) model [21]. The minimized results are re-clustered using 4 Å full atom RMSD as the clustering restraint, and clusters with less than ten members are discarded. The remaining clusters are re-ranked based upon Boltzmann-averaged energy, and the six lowest energy poses for each type of probe are retained. Hotspots are generated by grouping neighboring clusters together. If there are multiple neighbors to a cluster (neighbor being defined as a cluster falling within 4 Å), then they are considered a better hotspot.

### 3.3.2 Definition of Tight Homo-Multimer Interaction

Homo-multimers are identified based on their sequence similarity—two monomers are considered homomers only if they are exactly the same. To define if monomers in these homo-multimers are tightly or loosely associated, the ratio of buried surface area is calculated: if >25% of the surface area is buried in the homo-multimer, the multimer is considered to be obligatory and it is suggested to consider it as one unit to map (*see* also **Note 1**).

### 3.3.3 Definition of Domains in Receptor Protein

Peptimap uses the CATH classification [17] to define domains on the receptor structure. If no classification is available, sequence alignment is used to find the CATH domain closest to the provided structure.

### 3.3.4 Identification of Binding Pockets Inaccessible to Peptides

The use of small molecules as a search method allows for identification of sites that are not necessarily accessible to peptides, such as those in the core of a protein. In order to remove these sites, a sphere is generated at the center of the site and one hundred rays are projected outward. If 80% or more of the rays contact the protein (i.e., they pass within 2 Å of an atom), the site is considered internal and it is suggested to the reader to discard it.

---

## 4 Notes

1. Criteria for deciding on mapping monomers or homo-multimers. Detection of peptide-binding sites depends on defining the structure that the peptide will encounter on its search. For obligate homo-multimers, it can be assumed that the individual monomer structure is never encountered (if it exists as stable, separate structure at all). In contrast, transient homo-multimers might also expose single monomers. Therefore, the following rule is suggested:

*Rule #1:* If the interaction between the individual monomers is strong, map the homo-multimer, else map individual monomer structures. Strength of interaction is estimated based



on the buried surface area of the interface (*see* Subheading 3.3.2 for details on buried surface area criteria).

2. Criteria for deciding on which region to map in a multi-domain protein. Multi-domain protein structures solved by x-ray crystallography might display pockets at the domain interface that could fit a peptide well. However, these pockets are not necessarily stable, as the two domains might not be oriented in a fixed orientation shown in the crystal, but rather show a more flexible orientation, which will affect the stability of the binding pocket. In calibrating PeptiMap, we have therefore devised simple rules to distinguish true binding sites at domain interfaces from unstable false-positive hits, for protein chains that contain more than one structured domain (*see* Subheading 3.3.3 for details of how domains are mapped).

*Rule #2:* If the mapped domains are identical, it is recommended to map the full structure.

*Rule #3:* If the domains are different (hetero-domain structure), it is suggested to map single specific domains of interest. In that case, we assume that even if a binding site is at a domain interface, one of the domains will be the predominant contributor to binding, and consequently that binding site is also identified on single domains.

## References

1. Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol* 3:e405
2. Raveh B, London N, Schueler-Furman O (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins* 78:2029–2040
3. Raveh B, London N, Zimmerman L, Schueler-Furman O (2011) Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS One* 6, e18934
4. Trellet M, Melquiond AS, Bonvin AM (2013) A unified conformational selection and induced fit approach to protein-peptide docking. *PLoS One* 8, e58769
5. Donsky E, Wolfson HJ (2011) PepCrawler: a fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. *Bioinformatics* 27:2836–2842
6. London N, Raveh B, Schueler-Furman O (2013) Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. *Curr Opin Struct Biol* 23:894–902
7. Mattos C, Ringe D (1996) Locating and characterizing binding sites on proteins. *Nat Biotechnol* 14:595–599
8. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274: 1531–1534
9. Brenke R, Kozakov D, Chuang GY, Beglov D, Hall D, Landon MR, Mattos C, Vajda S (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* 25:621–627
10. Ngan CH, Hall DR, Zerbe B, Grove LE, Kozakov D, Vajda S (2012) FTSite: high accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics* 28:286–287
11. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* 10:733–755
12. Ben-Shimon A, Eisenstein M (2010) Computational mapping of anchoring spots on protein surfaces. *J Mol Biol* 402:259–277

13. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242
14. Lavi A, Ngan CH, Movshovitz-Attias D, Bohnuud T, Yueh C, Beglov D, Schueler-Furman O, Kozakov D (2013) Detection of peptide-binding sites on protein surfaces: the first step toward the modeling and targeting of peptide-mediated interactions. *Proteins* 81:2096–2105
15. Park EY, Lee BG, Hong SB, Kim HW, Jeon H, Song HK (2007) Structural basis of SspB-tail recognition by the zinc binding domain of ClpX. *J Mol Biol* 367:514–526
16. Raaf J, Klopffleisch K, Issinger OG, Niefind K (2008) The catalytic subunit of human protein kinase CK2 structurally deviates from its maize homologue in complex with the nucleotide competitive inhibitor emodin. *J Mol Biol* 377:1–8
17. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG, Lehtinen S, Studer RA, Thornton J, Orengo CA (2015) CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res* 43: D376–D381
18. Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK (1997) Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* 89: 951–961
19. Tomoo K, Matsushita Y, Fujisaki H, Abiko F, Shen X, Taniguchi T, Miyagawa H, Kitamura K, Miura K, Ishida T (2005) Structural basis for mRNA Cap-Binding regulation of eukaryotic initiation factor 4E by 4E-binding protein, studied by spectroscopic, X-ray crystal structural, and molecular dynamics simulation methods. *Biochim Biophys Acta* 1753:191–208
20. Jackson LP, Lewis M, Kent HM, Edeling MA, Evans PR, Duden R, Owen DJ (2012) Molecular basis for recognition of dilysine trafficking motifs by COPI. *Dev Cell* 23: 1255–1262
21. Brooks BR III, Jr CLB, Nilsson ADM, Petrella L, Roux RJ, Won B, Archontis Y, Bartels G, Boresch C, Caflisch S, Caves A, Cui L, Dinner Q, Feig AR, Fischer M, Gao S, Hodoscek J, Im M, Kuczera W, Lazaridis K, Ma T, Ovchinnikov J, Paci V, Pastor E, Post RW, Pu CB, Schaefer JZ, Tidor M, Venable B, Woodcock RM, Wu HL, Yang X, York WDM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614

<http://www.springer.com/978-1-4939-6796-4>

Modeling Peptide-Protein Interactions

Methods and Protocols

Schueler-Furman, O.; London, N. (Eds.)

2017, XI, 311 p. 87 illus., 68 illus. in color. With online files/update., Hardcover

ISBN: 978-1-4939-6796-4

A product of Humana Press