

Peptide and Protein Interaction Prediction and Intervention with Computational Methods

Newsha Fallah^a, Saeed Siavashy^b, Nasim Ghaemian^a, Farshad Bahramian^b, Fatemeh Ghorbani-Bidkorbeh^{a*}

^a Department of Pharmaceutics, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

^b Faculty of Mechanical Engineering, K. N. Toosi University of Technology, Tehran, Iran.

Article history:

Received: 18 September 2017

Accepted: 31 October 2017

HIGHLIGHTS

- Highlights the importance of peptides and proteins interactions.
- Summarizes the computational methods which are applicable in peptide and protein interaction prediction.
- Highlights the applications of computational methods in peptides and proteins interactions.

Keywords:

Computational methods
Interaction prediction
Simulation
Peptides and proteins

ABSTRACT

Proteins are the most fascinating multifaceted biomacromolecules in living systems and play various important roles such as structural, sensory, catalytic, and regulatory function. Protein and peptide interactions have emerged as an important and challenging topic in biochemistry and medicinal chemistry. Computational methods as promising tools have been utilized to predict protein and peptide interactions in order to intervene in the biochemical processes and facilitate pharmaceutical peptide design and clarify the complications. This review will introduce the computational methods which are applicable in protein and peptide interaction prediction and summarizes the most successful examples of computational methods described in the literature.

Introduction

The interactions between peptides and proteins are one of the most effective factors on so many cellular processes such as signal transduction, transport of proteins, and some of important immunity system reactions (Rodrigues and Bonvin, 2014). There are multiple diversity and changes in these chain conformations because of ability of bindings between amino acids which can lead to unpredictable interactions (Bhattacharjee and Wallin, 2013). The small peptides called motifs which bind to specific segments of proteins are the main key of connection between peptide and protein in order to transduce a signal (Kilburg and

Gallicchio, 2016). Motifs mostly interact with low affinity; therefore they might interact with many different sites of a protein chain that results in producing different cell products, unwanted reactions, protein degradation, and post translational modification (Bardwell and Treisman, 1994). As an example, domains containing SH₃ groups can bind to a motif with a sequence of proline – X – X – proline (X could be any amino acid) (Bhattacharjee and Wallin, 2013). We divide the peptide ligands based on the place of positive charge into two categories, class 1 which binds from N-terminal and class 2 which binds from C-terminal (Hou et al., 2006). Indeed, these low specific peptides may connect to multiple sites in a protein sequence or even to an unrelated protein; hence designing a precise peptide could result in having a determined and controlled outcome (Grigoryan et al., 2009).

Nowadays, peptide and protein therapeutics are used

* Corresponding Author:

Email: f.ghorbani@sbm.ac.ir (F. Ghorbani-Bidkorbeh)

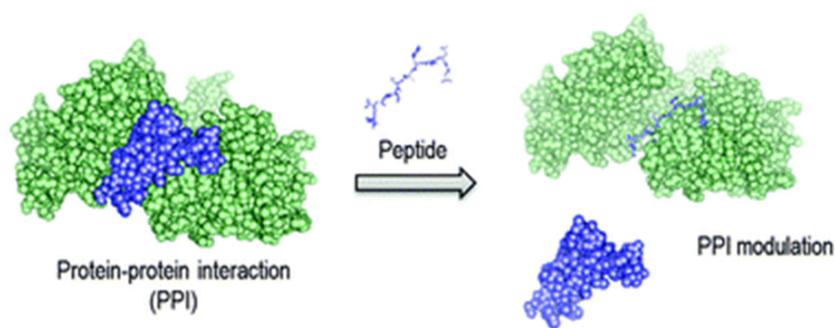


Figure 1. Modulating Protein-protein interactions: the potential of peptides. The whole figure is reproduced from (Nevola and Giralt, 2012) with permission of The Royal Society of Chemistry.

in treating a wide range of diseases. Therefore, designing them and predicting the possible reactions is significant in order to manipulate the cells in a right way (Chen and Keating, 2012). For this purpose, computational simulation has been used to design a specific and high-affinity ligand and predetermine the protein-peptide reaction and decreasing the undesirable effects (Błaszczuk et al., 2016). Many methods are employed for this process such as Monte Carlo-based and molecular dynamic, but it is difficult to occupy a certain method in order to devise various ligands (Chen and Keating, 2012). Figure 1 simply shows how designing a unique peptide can make a correct interaction in a biologic system. On the left side a connective part (blue) has been chosen in order to role as a pattern for designing an exogenous peptide. On the right side, the designed peptide totally mimics the connective part and makes a similar interaction between two parts. Conclusively, designing a peptide or motif based on the natural sequences could results in having the same interaction (Nevola and Giralt, 2012).

The present review explores to summarize the computational methods which are mostly applicable in peptide and protein interaction prediction and furthermore we focus on some applications through introducing examples.

Molecular modeling

There are many computational methods which are used for visualization, calculation and analysis of the molecular properties. One of these methods is Molecular Modeling (MM) (Neumaier, 1997). In this method, potential energy is the most important parameter which is used for calculation of potential energy in different atoms configurations. The stability of different complexes can be determined by calculating atomic arrangement potential energy which is related to the atoms coordination in atomic structure. These forces can be estimated by solving

motion equations numerically.

The other parameters of the potential energy are constant because of harmonic forces. These constant parameters are used for vibrational spectra predictions, analysis of normal and collective modes of motion (Kuczera et al., 1998).

Molecular dynamic

Molecular Dynamic (MD) is a computational method for analyzing the physical movements which solves the motion equations in a system of moving particles at a specific amount of potential energy. The equations which are solved in MD have conformance with the Newton's equations of classical mechanics but some extra equations are added for considering surrounding environment parameters (e.g. Langevin, Brownian dynamics). One of the advantages of Newton's equations is the presence of a systematic way to generate a group of configurations of simulated system that correspond with classical time evolution. These configurations are needed for providing data about structures. They also describe the energy flow and specify the details of changes in structural dynamics in a model which is used for calculating the time correlation functions and rates of conformational transition (Levitt, 1983).

Monte Carlo

Monte Carlo (MC) method is a computational algorithm based on calculation of random sampling (Valleau and Torrie, 1977). In molecular modeling, MC method is used for generating samples of system structures and corresponding distribution of system temperature. After calculation of average energy and structural parameters, the results should be validated by statistical mechanics, theoretical, and experimental results (Valleau and Torrie, 1977).

Table 1. Summary of local docking programs.

Method	Description	Reference
HADDOCK	Related to experimental data of biochemical and biophysical interaction	(Dominguez et al., 2003)
GalaxyPepDock	Generating protein-peptide structures by PepBind database. As inputs, It needs PDB format of protein structure and FASTA peptide sequence.	(Das et al., 2013; Lee et al., 2015)
MedusaDock	An approach uses where there is difference in conformational states of the binding partners.	(Ding et al., 2010)

Protein and peptide docking

In recent years, peptides have played an important role in new drug development. Peptide unique properties such as their ability to react and bind with small proteins increase research on these macromolecular structures. Estimation of protein-peptide complex structure and molding the interaction between peptide and protein is one of the important advantages of numerical molding. On the other hand, the structure determination may be possible by experimental investigation such as crystallography and nuclear magnetic resonance spectroscopy of proteins (NMR). However, determination of peptide-protein complexes structures for all kinds of peptides and proteins is impossible because of wide variety of peptides and proteins. Molecular docking is a useful tool which uses databases like Protein Data Bank (PDB). Today, there are too many types of software of protein-drug modeling or rigid protein-protein docking. Actually because of more degrees of freedom in protein-peptide docking, this kind of interaction is more complicated than protein-drug or protein-protein docking (Meng et al., 2011; Rose et al., 2011)

Local docking: Peptide docking to a known binding site

In fact, good interaction estimation and good knowledge of interaction details have a great effect on success of protein-peptide complexes modeling. In local docking

method the prior structural data are exploited for structural predictions. These techniques are based on evaluation of many peptides that are derived from a known structure for a single receptor (Dominguez et al., 2003).

Some current local docking programs are introduced in Table 1.

Global docking: Peptide docking to an unknown binding site

The application of this method is when the binding site is unclear. Usually the first product is peptide form ensembles, then peptide fits to the external surface of desired protein by rigid-body docking. Search area is the major parameter that has big effect on accuracy and efficiency of these methods. In modeling, most protocols start with using coarse-grained protein model to find rough estimation of viable binding area. Then, higher resolutions up to atomic scale is used for increasing the method precision (Błaszczuk et al., 2013; Kurcinski et al., 2015). Summary of some methods is presented in Table 2.

Pathway free energy

In pathway free energy method, free energy that exists between the unbound or bound states of the ligand-protein complex is computed by thermodynamic rules and equations. In other words, this method is based on

Table 2. Summary of global docking programs.

Method	Description	Reference
The CABS-dock	The base of this procedure is coarse-grained CABS protein model.	(Jamroz et al., 2013, Kurcinski et al., 2015)
PepATTRACT	This procedure uses combining coarse-grained peptide for investigating the possible binding sites.	(May and Zacharias, 2008; de Vries et al., 2015)

free energy computation as ratios of partition functions (Guvench and MacKerell, 2009; Gallicchio and Levy, 2011).

The superiority of these methods to other methods is modeling the absolute free energy. The path can be parameterized as λ , ranging from 0 to 1. These methods can be classified by their alchemical or physical pathways (Deng and Roux, 2009).

Physical pathway

In physical pathway approach, the ligand transferring from bulk solvent to the binding site occurs. In fact, binding pathway along each one of the six intermolecular freedom degrees can be parameterized (Woo and Roux, 2005).

Alchemical approaches

In alchemical approach, ligand is limited to the binding site area and unlike the physical pathway methods, it is not transferred. Actually, the transformation between the bound and unbound states performed by dialing in ligand–receptor interactions in a space is called alchemical space. This is achieved by computational calculations which leads to construct a λ -dependent potential energy function and shows ligand is fully interacting with the receptor at $\lambda=0$ and ligand–receptor interactions are turned off at $\lambda=1$ (Gilson et al., 1997; Boyce et al., 2009).

Molecular mechanics

The two most common end-point methods are Molecular Mechanics Poisson Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) which are applied to protein–ligand/protein–peptide interactions (de Ruyter and Oostenbrink, 2011; Genheden and Ryde, 2015). Furthermore, the mining minima method is another example of these methods. However, it is not applicable to protein–peptide binding (Chia-en et al., 2007; Chen et al., 2010). The main advantage of these methods is simplicity of parameters which are the free protein and peptide states, and then the simulation of the protein–peptide complex can be done.

Artificial intelligence

Artificial intelligence is an evolutionary algorithm that was created for predicting, designing, and optimizing peptide structures. Generally, this method consists of five steps: (i) first step is about identification of a peptide called "seed peptide" which has desired activity. (ii) Generation of variants from physicochemical environment that surrounds the seed peptide. (iii) The next step after synthesis in which biased library will be tested. (iv) A

relationship between quantitative sequence-activity will be modeled utilizing an artificial neural network, and finally, (v) effective computational search based on neural network algorithm done for de novo design. This strategy has been successfully utilized to identify novel peptides (MacKerell et al., 1998; Radhika and Rao, 2015; Paladino et al., 2017).

Applications

Due to wide spread use of computational methods in recent decades, there has been many studies and experiments in this field. As mentioned before, proteins seem to have a flexible conformation which makes the structure and bonds forecasting even more complicated. Some issues like inaccuracies, time consumption, and requirement of multiple CPUs lead the process towards more efficient methods, for instance Monte Carlo, Rosetta docking, ATTRACT, and Swarm dock. Here we review some of these studies as instances (Kmieciak et al., 2016).

According to significant role of signal transduction in human body immunity system, Cho et al. designed a case study experiment on TNF α -Mediated NF- κ B signal transduction pathway in order to identify the key factors in this process. In order to illustrate the accuracy of quantitative hypothesis of this signaling, mathematical models have been used and simulated then a wide range of parameter values has been compared in order to consider the method sensitivity (Cho et al., 2003).

Nowadays, there are multiple sources presenting the protein structures in order to ease the prediction of reactions. GalaxyPepDock is one of online servers which perform docking based on an experimental database. This server first overview the structures then optimize an interaction based on energy and flexibility of bonds. Here we review an instance; the input information are protein structure and peptide sequence. Then, based on the similarities of these structures, the server chooses and simulates a template from a database. Consequently, the server optimizes the bonds based on their energy and finally suggests the most preferred templates (Lee et al., 2015).

Another aspect of using computational methods is about a serious problem in today's human health complication; cancer. In this subject, Debasree Sarkar et al. used some of protein structure databases and MEME, a tool for finding ungapped and repeating sequences in order to find new motifs and predict the role of these peptide sequences in three cancer hub proteins and recreate novel motifs, and exert a scoring system called overlapping linear peptide (OLP) in order to estimate the overlapping of novel motifs with proteins which might result in producing a better known process of protein-peptide interaction effects in cancer cells (Sarkar et al., 2016).

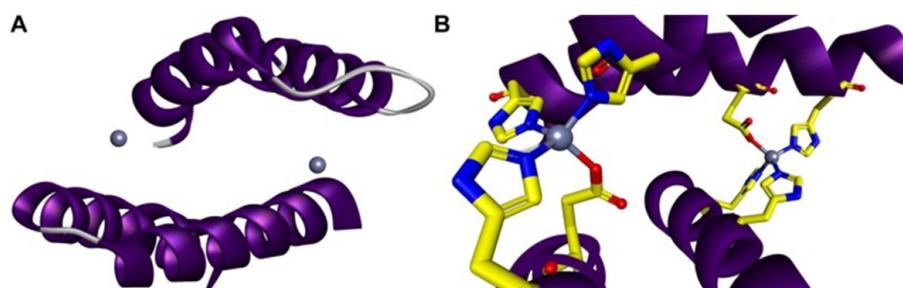


Figure 2. Computationally designed protein-protein interactions with high affinity and desired orientation. (A) The symmetric homodimer design with two interface zinc sites each coordinated by four histidines at $i, i + 4$ positions on each helix. A Rosetta-based approach for the rational design of a protein monomer to form a zinc-mediated, symmetric homodimer. Incorporating metal-binding sites at the target interface may be one approach for increasing affinity and specifying the binding mode. (B) Metal interface design, named MID1, forms a tight dimer in the presence of zinc (MID1-zinc) with a dissociation constant <30 nM. (Tiwari et al., 2012).

Designing an upgraded protein which is more reactive is also possible by utilizing these computational methods. Amy E. Palmer et al reengineered the interface of Calmodulin with docking in order to be more sensitive to calcium ions therefore eases the monitoring of calcium dependent reactions in in-vitro studies. In this process they calculated the whole energy by Monte Carlo simulating progress (Palmer et al., 2006).

Toxic superoxide dismutase1 (SOD1) may be one of the factors responsible in the pathophysiology of familial amyotrophic lateral sclerosis (FALS). Designing and finding molecular agents that bind to mutant and misfolded SOD1 could be a potential approach for this disease treatment.

The obstacle in designing such agents is high structural homology between the mutant and the normal SOD1 proteins. Using a computational method would make the process of designing agents and predicting their interactions with SOD1 enzyme easier and faster (Banerjee et al., 2017).

Cyclophilins are a group of proteins which catalyze the isomeration of peptidyl-prolyl peptide bond. These proteins exist in almost every known cell. Cyclophilin A, the first member of the group that got discovered, mediates the action of cyclosporine, the immunosuppressant drug. The interaction between Cyclophilin A and HAGPIA peptide from the HIV-1 capsid protein was illustrated using pepATTRACT web server. It plots 50 poses of the HAGPIA peptide and tells how the peptide is inclined to interact with the receptor (Banerjee et al., 2017; de Vries et al., 2017).

Another approach is based on metal mediated interactions. Tiwari et al. as illustrated in figure 2 designed a helix bundle which connects to metal sites specifically using computational methods such as Rosetta Match algorithm and Monte Carlo method (Tiwari et al., 2012).

Conclusion

One of the difficult challenges in the field of protein and peptide interaction is the prediction of this procedure in order to build new motifs and gain the ability of managing the process along the required field. Intricacy of the interactions between amino acids has become less complicated due to development of powerful computers in recent decades and vast application of computer simulation and modeling tools in different fields. Application of these tools by researchers causes some features such as peptide design, prediction of peptide-protein complex structures during interactions, effectiveness of interactions, and reducing cost and energy. We hope in future utilizing these methods, researches can develop new motifs with unique features and furthermore protein-peptide interaction will be more figured out.

Competing Interests

The authors declared no conflict of interest.

References

- Banerjee, V., Oren, O., Ben-Zeev, E., Taube, R., Engel, S. and N. Papo, (2017). "A computational combinatorial approach identifies a protein inhibitor of superoxide dismutase 1 misfolding, aggregation, and cytotoxicity." *Journal of Biological Chemistry*, **292**(38): 15777-15788.
- Bardwell, V. J. and R. Treisman, (1994). "The POZ domain: a conserved protein-protein interaction motif." *Genes & Development*, **8**(14): 1664-1677.
- Bhattacharjee, A. and S. Wallin, (2013). "Exploring protein-peptide binding specificity through computational peptide screening." *PLoS Computational Biology*, **9**(10): e1003277.
- Błaszczczyk, M., Jamroz, M., Kmiecik, S. and A. Kolinski, (2013). "CABS-fold: server for the de novo and consensus-based prediction of protein structure." *Nucleic Acids Research*, **41**(W1), pp.W406-W411.

- Blaszczyk, M., Kurcinski, M., Kouza, M., Wieteska, L., Debinski, A., Kolinski, A. and S. Kmiecik, (2016). "Modeling of protein-peptide interactions using the CABS-dock web server for binding site search and flexible docking." *Methods*, **93**: 72-83.
- Boyce, S. E., Mobley, D. L., Rocklin, G. J., Graves, A. P., Dill, K. A. and B. K. Shoichet, (2009). "Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site." *Journal of Molecular Biology*, **394**(4): 747-763.
- Chen, T. S. and A. E. Keating, (2012). "Designing specific protein-protein interactions using computation, experimental library screening, or integrated methods." *Protein Science*, **21**(7): 949-963.
- Chen, W., Gilson, M. K., Webb, S. P. and M. J. Potter, (2010). "Modeling protein-ligand binding by mining minima." *Journal of Chemical Theory and Computation*, **6**(11): 3540-3557.
- Chia-en, A. C., Chen, W. and M. K. Gilson, (2007). "Ligand configurational entropy and protein binding." *Proceedings of the National Academy of Sciences*, **104**(5): 1534-1539.
- Cho, K. H., Shin, S. Y., Kolch, W. and O. Wolkenhauer, (2003). "Experimental design in systems biology, based on parameter sensitivity analysis using a monte carlo method: A case study for the tnfa-mediated nf- κ b signal transduction pathway." *Simulation*, **79**(12): 726-739.
- Das, A. A., Sharma, O. P., Kumar, M. S., Krishna, R. and P. P. Mathur, (2013). "PepBind: a comprehensive database and computational tool for analysis of protein-peptide interactions." *Genomics, Proteomics & Bioinformatics*, **11**(4): 241-246.
- de Ruiter, A. and C. Oostenbrink, (2011). "Free energy calculations of protein-ligand interactions." *Current Opinion in Chemical Biology*, **15**(4): 547-552.
- de Vries, S. J., Rey, J., Schindler, C. E., Zacharias, M. and P. Tuffery, (2017). "The pepATTRACT web server for blind, large-scale peptide-protein docking." *Nucleic Acids Research*, **2017**.
- de Vries, S. J., Schindler, C. E., de Beauchêne, I. C. and M. Zacharias, (2015). "A web interface for easy flexible protein-protein docking with ATTRACT." *Biophysical Journal*, **108**(3): 462-465.
- Deng, Y. and B. Roux, (2009). "Computations of standard binding free energies with molecular dynamics simulations." *The Journal of Physical Chemistry B*, **113**(8): 2234-2246.
- Ding, F., Yin, S. and N. V. Dokholyan, (2010). "Rapid flexible docking using a stochastic rotamer library of ligands." *Journal of Chemical Information and Modeling*, **50**(9): 1623-1632.
- Dominguez, C., Boelens, R. and A. M. Bonvin, (2003). "HADDOCK: a protein-protein docking approach based on biochemical or biophysical information." *Journal of the American Chemical Society*, **125**(7): 1731-1737.
- Gallicchio, E. and R. M. Levy, (2011). "Recent theoretical and computational advances for modeling protein-ligand binding affinities." *Advances in Protein Chemistry and Structural Biology*, **85**: 27.
- Genheden, S. and U. Ryde, (2015). "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities." *Expert Opinion on Drug Discovery*, **10**(5): 449-461.
- Gilson, M. K., Given, J. A., Bush, B. L. and J. A. McCammon, (1997). "The statistical-thermodynamic basis for computation of binding affinities: a critical review." *Biophysical Journal*, **72**(3): 1047-1069.
- Grigoryan, G., Reinke, A. W. and A. E. Keating, (2009). "Design of protein-interaction specificity gives selective bZIP-binding peptides." *Nature*, **458**(7240): 859-864.
- Guvench, O. and A. D. MacKerell Jr, (2009). "Computational fragment-based binding site identification by ligand competitive saturation." *PLoS Computational Biology*, **5**(7): e1000435.
- Hou, T., Chen, K., McLaughlin, W. A., Lu, B. and W. Wang, (2006). "Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain." *PLoS Computational Biology*, **2**(1): e1.
- Jamroz, M., Kolinski, A. and S. Kmiecik, (2013). "CABS-flex: server for fast simulation of protein structure fluctuations." *Nucleic Acids Research*, **41**(W1): W427-W431.
- Kilburg, D. and E. Gallicchio, (2016). "Chapter Two-Recent Advances in Computational Models for the Study of Protein-Peptide Interactions." *Advances in Protein Chemistry and Structural Biology*, **105**: 27-57.
- Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E. and A. Kolinski, (2016). "Coarse-grained protein models and their applications." *Chemical Reviews*, **116**(14): 7898-7936.
- Kuczera, K. (2011). "Molecular Modeling in Peptide and Protein Analysis." In: *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Ltd.
- Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A. and S. Kmiecik, (2015). "CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site." *Nucleic Acids Research*, **43**(W1): W419-W424.
- Lee, H., Heo, L., Lee, M. S. and C. Seok, (2015). "GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization." *Nucleic Acids Research*, **43**(W1): W431-W435.
- Levitt, M. (1983). "Molecular dynamics of native protein: I. computer simulation of trajectories." *Journal of Molecular Biology*, **168**(3): 595-617.
- MacKerell Jr, A. D., Bashford, D., Bellott, M. L. D. R., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S. and D. Joseph-McCarthy, (1998). "All-atom empirical potential for molecular modeling and dynamics studies of proteins." *The Journal of Physical Chemistry B*, **102**(18): 3586-3616.
- May, A. and M. Zacharias, (2008). "Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: Evaluation on kinase inhibitor cross docking." *Journal of Medicinal Chemistry*, **51**(12): 3499-3506.
- Meng, X. Y., Zhang, H. X., Mezei, M. and M. Cui, (2011). "Molecular docking: a powerful approach for structure-based drug discovery." *Current Computer-Aided Drug Design*, **7**(2): 146-157.
- Neumaier, A. (1997). "Molecular modeling of proteins and mathematical prediction of protein structure." *SIAM Review*, **39**(3): 407-460.
- Nevola, L. and E. Giralt, (2015). "Modulating protein-protein interactions: the potential of peptides." *Chemical Communications*, **51**(16): 3302-3315.
- Paladino, A., Marchetti, F., Rinaldi, S. and G. Colombo, (2017). "Protein design: from computer models to artificial intelligence." *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2017**.
- Palmer, A. E., Giacomello, M., Kortemme, T., Hires, S. A., Lev-Ram, V., Baker, D. and R. Y. Tsien, (2006). "Ca²⁺ indicators based on computationally redesigned calmodulin-peptide pairs." *Chemistry &*

Biology, **13**(5): 521-530.

Radhika, V. and V. S. H. Rao, (2015). "Computational approaches for the classification of seed storage proteins." *Journal of Food Science and Technology*, **52**(7): 4246-4255.

Rodrigues, J. P. and A. M. Bonvin, (2014). "Integrative computational modeling of protein interactions." *The FEBS Journal*, **281**(8): 1988-2003.

Rose, P. W., Beran, B., Bi, C., Bluhm, W. F., Dimitropoulos, D., Goodsell, D. S., Prlić, A., Quesada, M., Quinn, G. B., Westbrook, J. D. and J. Young, (2010). "The RCSB Protein Data Bank: redesigned web site and web services." *Nucleic Acids Research*, **39**(suppl_1): D392-D401.

Sarkar, D., Patra, P., Ghosh, A. and S. Saha, (2016). "Computational framework for prediction of peptide sequences that may mediate multiple protein interactions in cancer-associated hub proteins." *PLoS*

One, **11**(5): e0155911.

Tiwari, M. K., Singh, R., Singh, R. K., Kim, I. W. and J. K. Lee, (2012). "Computational approaches for rational design of proteins with novel functionalities." *Computational and Structural Biotechnology Journal*, **2**(3): 1-13.

Valleau J. P. and G. M. Torrie, (1977). "A Guide to Monte Carlo for Statistical Mechanics: 2. Byways." In: Berne B. J. (Eds), *Statistical Mechanics. Modern Theoretical Chemistry*, volume 5, Springer, Boston, MA, pp. 169-194.

Woo, H.J. and B. Roux, (2005). "Calculation of absolute protein-ligand binding free energy from computer simulations." *Proceedings of the National Academy of Sciences of the United States of America*, **102**(19): 6825-6830.

This open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).