A New Flexible Protocol for Docking Studies

Lucia Sessa[1(✉)], Luigi Di Biasi[1,2], Simona Concilio[3], Giuseppe Cattaneo[2], Alfredo De Santis[2], Pio Iannelli[1], and Stefano Piotto[1]

[1] Department of Pharmacy, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
lucsessa@unisa.it
[2] Department of Informatics, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
[3] Department of Industrial Engineering, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy

Abstract. A significant prerequisite for computational structure-based drug design is the estimation of the structures of ligand-receptor complexes. For this task, the flexibility of both ligand and receptor backbone is required, but it requires the exploration of an extremely vast conformational space. Here we present a protocol to address the receptor flexibility using complementary strategies and the use of receptor sequence conservation. The method aims to increase the accuracy of predicted ligand orientation in the binding pocket and the receptor-ligand binding affinity. The precision in affinity prediction permits to distinguish between binders and non-binders and to identify binding sites and ligand poses necessary for lead optimization.

1 Introduction

It is very important to estimate the in silico potential toxicity of existing or hypothetical compounds. This can be achieved simulating the interactions of a ligand towards target proteins suspected to trigger adverse effect. Docking methods can be useful to evaluate potential toxicity of small molecule [1], and to distinguish between binders and non-binders. Calculation of free energy of binding is a good way to estimate the binding affinity between two molecules. We used the enzyme Androgen Receptor (AR), a nuclear receptor activated by binding androgenic hormones, testosterone, or dihydrotestosterone to test the flexible protocol. AR plays an essential role in the growth of normal prostate, and it is involved in the development of prostate cancer [2]. When the experimental structure of the complex ligand-receptor is known, the ligand can be docked directly in the binding site. However, experimental and theoretical studies show that proteins can fluctuate between different conformations in the absence of ligand [3–9]. In aqueous solution, proteins domains are in constant motion exhibiting a conformational heterogeneity [10–15]. Molecular recognition involves non-covalent association of ligands to protein target with high affinity and specificity. According to the Koshland's 'induced-fit' model, the interaction between a protein and a ligand induces a conformational change in the protein [16]. Among the conformations of the dynamically fluctuating protein, the receptor selects the one that better accommodates the

F. Rossi et al. (Eds.): WIVACE 2015, CCIS 587, pp. 117–126, 2016.
DOI: 10.1007/978-3-319-32695-5_11

ligand. In case of modeling protein-ligand complexes, it is necessary to consider backbone movement. The major flaw in docking methods is to consider only a single representative structure for the receptor. Introducing receptor flexibility in a standard docking protocol is a way to study conformational changes. Flexibility is particularly important when the binding pocket is inside the receptor and the ligand interactions induce structural movements of the backbone of the receptor. We explored the conformation variability by addressing the receptor flexibility and taking into account the receptor sequence conservation.

Keeping proteins flexible during the docking has a high computational cost in virtue of the high number of degrees of freedom. To overcome this limit, our docking strategy is to represent receptor flexibility by utilizing an inexpensive method that offers more target structures.

First, we considered several side-chain conformations of the receptors and we calculated, for each rotamer, the values of binding energy of a ligand during a molecular dynamics simulation. In a second approach, we create a docking ensemble of structures derived by steered molecular dynamics (SMD). In this method, we collected the trajectory when the ligand is pulled away from its binding pocket. SMD simulation is an inexpensive computational method and offers more target structures to perform docking. In the third approach, we consider the low frequency vibrations of a protein. Protein low-frequency vibrations retain important biological functions [17]. For each vibrational mode, we generated a set of snapshots in a similar fashion to the first approach. Finally, we present a protocol for docking of ligands into flexible protein binding sites. The protocol was tested on a list of therapeutically relevant targets with available crystallographic data. The free energy was calculated as the difference between the energy of separated compounds and the energy of the complex. However, the computational complexity of the procedure grows quickly with the numbers of conformers considered. Consequently, to reduce the computational time and cost we have used a specialized grid (GRIMD) to distribute all jobs [18].

2 Materials and Methods

2.1 Data Set – Selection of Complexes

The structures for the docking simulations were taken from the X-ray structures in the PDB database [19]. We performed a search using the following query parameters: human protein Androgen receptor (Uniprot ID P10275), structure complexed with a ligand containing experimental binding affinity and the X-ray resolution up to 2.0 Å to ensure crystallographic structures with sufficient structural quality.

The receptor structures were prepared by removing all water molecules and substrates including the ligand molecule. All structures were prepared adding missing hydrogens and optimizing the hydrogen-bonding network. The internal cavity volume inside the macromolecules was calculated with the software YASARA Structure 15.6.21 [20].

2.2 Classical MD Simulations

We represented receptor flexibility by utilizing several snapshot from molecular dynamics (MD). The MD simulations were performed with the software YASARA Structure 15.6.21. We used AMBER03 as force field with long-ranged PME potential and a cutoff of 8.0 Å. The starting structures for the simulations of the ARs were extracted from the X-ray structures from the PDB database. A cubic periodic simulation cell of 512000 $Å^3$ was defined around all atoms of the receptor structures. The MD simulation was then initiated at 298 K and integration time steps for intramolecular forces every 1.25 fs.

Ten structures for each receptor were selected from the MD simulation at regular time intervals (each 500 ps). A simulation cell was centered on the binding pocket sides and dimensions of the box were adapted for each structure to cover the entirety of the active site.

2.3 Steered Molecular Dynamics Simulation

We represented receptor flexibility utilizing structures collected by steered molecular dynamics simulation (SMD). The SMD was carried out using the software YASARA Structure 15.6.21. We collected the trajectory traced when the ligand is pulled from its binding pocket to the outside. The initial structures were retrieved from the X-ray structures from the PDB database and were cleaned by erasing all water molecules and substrates different from the ligand molecule. A cubic periodic simulation cell of 512000 $Å^3$ was built around the entire complex. The charges were assigned at physiological conditions (pH 7.4). The simulation box was filled with water choosing a density of 0.997 g/mL. The simulation cell was neutralized with NaCl with a final concentration of 0.9 %. We minimized the energy of the system using first a steepest descent minimization followed by a simulated annealing minimization. The pulling acceleration of the ligand was 3 $Å/ps^2$. The simulation was stopped when the distance between the centers of mass of receptor and ligand was > 30 Å.

2.4 Low Vibrational Modes

We considered the low frequency vibrations that a protein may undergo. The lowest vibrational modes of a protein were determined using the Schrodinger's biologics suite [15]. Before running the calculation, the proteins were properly prepared, using the Protein Preparation Wizard and all waters and solvent molecules were deleted. The input structures were minimized and the vibrational modes were generated as a set of structures sampled at regular intervals along a full cycle of the vibrational mode. We set the following parameters to 5 vibrational modes to view, and 20 frames per mode.

For each vibrational mode, we generated a set of snapshots of the structure at a particular point in the vibration. On these structures, we performed docking simulation similarly to the MD approach.

2.5 Docking Method

Ligand structures were extracted from the X-ray pose of the complex receptor-ligands and were minimized into YASARA by using AMBER03 force field.

The molecular docking simulations were performed using VINA provided in the YASARA package. The force field selected was AMBER03. The ligands were independently docked 250 times against 5 receptor ensembles with alternative high-scoring solutions of the side-chain rotamer network each. The results were clustered using a tolerance of 5.0 Å. We calculated the values of binding energy and corresponding binding constants of each ligand respect to the receptor.

In addition, we used a new tool for molecular docking, named mod-VINA, developed from a modification of AutoDock VINA [21]. This method permits to increase the accuracy of docking and it is useful to predict the best ligand pose correlating the docking analysis to the receptor sequence conservation. Contrarily to VINA, this tool does not create the ligand poses in a random manner in all the space available around the receptor, but it generates a pose in local spheres created close to conserved residues. With this approach, mod-VINA not only improves its accuracy in pose generation, but also reduces considerably the computing time elapsed for blind docking. mod-VINA is convenient when the binding pocket is unknown and the only strategy is to perform a blind docking.

3 Results and Discussion

We collected several snapshots of receptor structure by MD, SMD and vibrational low modes.

To validate the accuracy of our flexible protocol in docking, we have used a set of 10 crystallographic conformations of AR with the same target but with a different ligand. In Table 1 are reported the PDB codes of the selected receptors and the corresponding ligands.

Table 1. The AR targets and the corresponding ligand

PDB code for AR	Ligand
2AM9	TES
2AX6	HFT
2AX8	FHM
2AX9	BHM
3B5R	B5R
3B65	3B6
3B66	B66
3B67	B67
3B68	B68
3L3X	DHT

The collected proteins showed the 100 % of sequence identity and the same position of the buried binding pocket.

We performed two parallel experiments of docking, one with a rigid protein target and one considering flexible receptor structures. In addition, we compared the results for both experiments in the re-docking and in the cross-docking analysis.

3.1 Rigid Receptor Docking Simulations

The predicted structure of a ligand-receptor complex with the better binding energy was superposed on the crystallographic complex conformation to calculate the root mean square deviation (RMSD).

The RMSD values are reported in Fig. 1. The cell holding the minimum value of RMSD (0.5 Å) is colored in green and the cell with the maximum value of RMSD (20 Å) is colored in white. All other cells are colored proportionally.

Diagonals values are the RMSD values for the re-docking analysis. As expected, the values are very low and accurate. These results were not surprising because the receptor structure used was exactly the conformation adopted in the bound state from the receptor with the specific ligand.

In the cross-docking studies, we docked the ligand from one complex into the receptor of the other complex and *vice versa*. The RMSD values between the predicted poses and the experimental data are reported in Fig. 1 and are color mapped as described before for the re-docking.

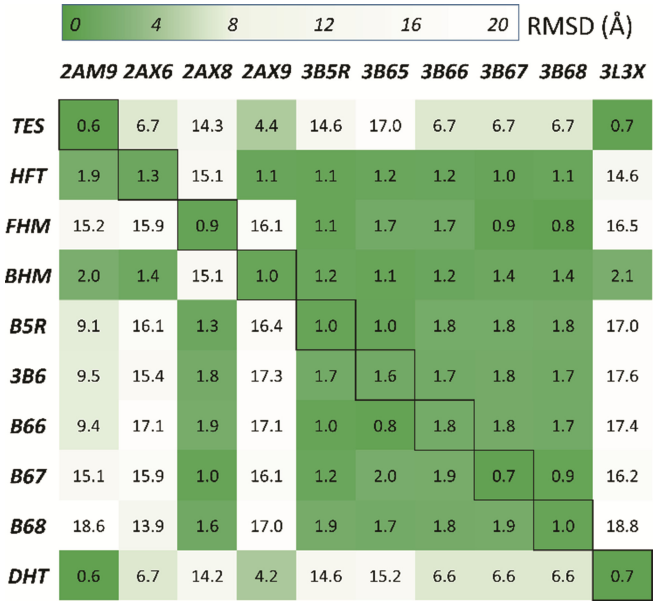
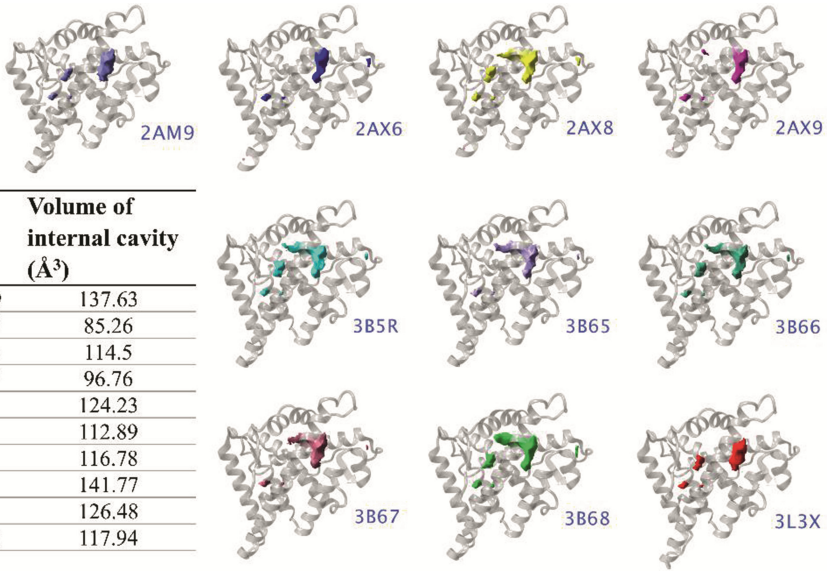


Fig. 1. Re-docking and cross-docking results for 10 different conformations of AR. The color mapped RMSD values are calculated between the predicted pose with highest energy of binding and the experimental pose (Color figure online).



PDB code	Volume of internal cavity (Å^3)
2AM9	137.63
2AX6	85.26
2AX8	114.5
2AX9	96.76
3B5R	124.23
3B65	112.89
3B66	116.78
3B67	141.77
3B68	126.48
3L3X	117.94

Fig. 2. Internal cavities of 10 different X-ray poses for AR. The secondary structures of the receptors are colored in gray to put in evidence the shape of their internal cavities (Color figure online).

We found that rigid cross-docking fails with several types of complexes with values of RMSD up to 18 Å. Analyzing the docked poses we observed that in those cases, the most favorite position for the ligand is on the receptor molecular surface and not in the empty space inside the receptor.

To understand this docking mistake, we measured the binding pocket buried inside the receptor calculating the volume of the internal cavities.

Figure 2 shows the internal cavity of the ten receptors and the volume for each cavity. It is possible to observe that AR displays different internal cavities shapes and volumes.

Slight changes in the structure of the binding pocket between different crystallographic structures can radically affect the outcome and alter the docking results. As showed for the receptors 3L3X or 2AX9 the internal cavity of the buried binding pocket does not have space enough to accommodate all ligands. In these cases, the cross-docking results are very bad. The changes of the cavity volume might explain the failure of traditional docking method and support the hypothesis that a single representative structure for the receptor is not enough.

3.2 Flexibility Receptor Docking Simulations

The comparison among the different approaches is summarized in Fig. 3.

We show the results for two different X-ray poses of AR (3B68 and 3L3X) re-docked with their crystallographic ligand such as B68 and DHT respectively (A and D graphs) and with a second substrate (B and C graphs). We reported the RMSD calculated

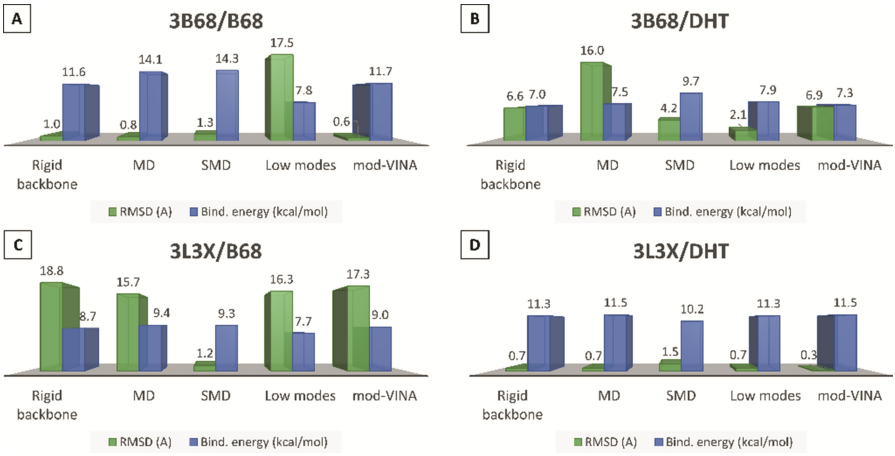


Fig. 3. Flexible protocol results for AR structures 3B68 and 3L3X re-docked (A and D) and cross-docked (B and C) with B68 and DHT. Here we show the docked poses with the highest energy of binding (blue column) predicted using the rigid backbone taken from the X-ray structure and the snapshots from MD, SMD and low-frequency vibrations. In addition we report the results obtained using the tool mod-VINA. To establish the accuracy of the prediction we reported the RMSD values (green column) calculated between the docked pose and the experimental one (Color figure online).

between the experimental and the predicted pose with the highest energy of binding. Low RMSD values indicate the correct superpose between two structures, also the more positive values of binding energy are correlated to the more favorable receptor-ligand interaction.

The re-docking experiments (A and D graphs) are very accurate with RMSD < 1 Å. As expected, in these cases the flexible docking protocol does not improve the results. This was not true for the cross-docking simulations (B and C graphs) where a flexible docking protocol can improve the overall quality of the simulations. In the receptor conformation in 3L3X, the internal cavity does not have space enough to accommodate the ligand B68, so the docking results are poor. However, we improved the docking performances using the SMD approach. We observed an RMSD value of 1.2 Å that means a complete overlapping between the cross-docked pose of B68 docked in the AR conformation taken from the 3L3X PDB and the experimental complex structure reported in the 3B68 PDB.

The other system analyzed was formed by the receptor structure frozen in 3B68 and the ligand DHT. This receptor shows a volume of the internal cavity big enough to host the substrate DHT. However, the overall quality (Fig. 3) is not satisfactory. In fact, the shape of the cavity may not be compatible with the geometry of the ligand. We obtained good results using the snapshot collected by the low-frequency vibrations in macromolecules. The different conformations collected from the low motions included a compatible internal cavity in term of shape and volume resulting in a good cross-docking performance (RMSD 2.1 Å).

In contrast, the dynamics of the receptor backbone, an approach conceptually very simple, did not improve the results of rigid docking. The performance of cross-docking was very low in all considered conformations. This was attributed to the nature of the binding pocket that is buried inside the receptor. Therefore the dynamics of the backbone deforms and reduces the binding pocket volume. The consequent slight changes in the structure of the binding pocket radically affected the docking results.

4 Conclusion

A major lack in standard docking protocol is the use of only one structure to represent the receptor. The crystallographic data corresponds, in the best case, only to the energy minimum of a specific pair receptor-ligand. When the binding pocket is buried inside the receptor, the protein deforms the internal cavity and the entire backbone to accommodate the ligands. We observed that different conformations of the same target show different volume and shape of the internal cavity. Consequently, the cross-docking analysis are typically very poor. To increase the accuracy of cross-docking we suggest a flexible protocol to deform the internal cavity in a "natural" way. We increased the receptor flexibility by enhancing target information and improving the precision of ligand position. We represented receptor flexibility by utilizing several snapshot from molecular dynamics simulations (MD), by steered molecular dynamics simulation (SMD) and using the structures extracted from the low frequency vibrations that a protein may undergo. Supposing that the binding pocket in a receptor corresponds to a conserved receptor region, we used a new tool (mod-VINA) for the docking developed from VINA to correlate the cross-docking simulation to the receptor sequence conservation.

The flexible model was used for cross-docking tests of 10 ligands and 10 different conformations of the androgen receptor. Based on the docking results we suggest that the SMD method provides various conformations with different internal cavities that can host ligands of different geometry. The low-vibrational mode dynamics was another good way to collect receptor structures. The protein vibrations were able to deform the internal cavity in a non-invasive way, as demonstrated by the good results in cross-docking experiment. In contrast, the molecular dynamics simulations were too strong in deforming the cavity of the binding pocket. As result, the cavity collapses and the docked ligands were placed outside the binding pocket on the receptor surface. In the fourth approach we taken into account the protein conservation. In the androgen receptor, the binding pocket is completely conserved. The best poses predicted by mod-VINA were completely superposed to the experimental structure for the re-docking results. Anyway, in the cross-docking this approach was not efficient, because of the different size and shape of the binding pockets.

The flexible docking protocol described in this work is a simple way to generate multiple receptor conformations that can be used to easily dock a large number of compounds.

Acknowledgments. This work was partially supported by the "Data-Driven Genomic Computing (GenData 2020)" PRIN project (2013–2015), funded by the Italian Ministry of the University and Research (MIUR).

References

1. Goldsmith, M.-R., Grulke, C.M., Chang, D.T., Transue, T.R., Little, S.B., Rabinowitz, J.R., Tornero-Velez, R.: DockScreen: A Database of In Silico Biomolecular Interactions to Support Computational Toxicology. Dataset Papers in Science **2014** (2014)
2. Pomerantz, M.M., Li, F., Takeda, D.Y., Lenci, R., Chonkar, A., Chabot, M., Cejas, P., Vazquez, F., Cook, J., Shivdasani, R.A., Bowden, M., Lis, R., Hahn, W.C., Kantoff, P.W., Brown, M., Loda, M., Long, H.W., Freedman, M.L.: The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat. Genet. **47**(11), 1346–1351 (2015)
3. Lopez, D.H., Fiol-deRoque, M.A., Noguera-Salvà, M.A., Terés, S., Campana, F., Piotto, S., Castro, J.A., Mohaibes, R.J., Escribá, P.V., Busquets, X.: 2-Hydroxy arachidonic acid: a new non-steroidal anti-inflammatory drug. PLoS ONE **8**(8), e72052 (2013)
4. Piotto, S., Concilio, S., Bianchino, E., Iannelli, P., López, D.J., Terés, S., Ibarguren, M., Barceló-Coblijn, G., Martin, M.L., Guardiola-Serrano, F.: Differential effect of 2-hydroxyoleic acid enantiomers on protein (sphingomyelin synthase) and lipid (membrane) targets. Biochim. Biophys. Acta (BBA)-Biomembr. **1838**(6), 1628–1637 (2014)
5. Piotto, S., Concilio, S., Mavelli, F., Iannelli, P.: Computer simulations of natural and synthetic polymers in confined systems. Macromolecular symposia **286**(1), 25–33 (2009)
6. Piotto, S., Trapani, A., Bianchino, E., Ibarguren, M., López, D.J., Busquets, X., Concilio, S.: The effect of hydroxylated fatty acid-containing phospholipids in the remodeling of lipid membranes. Biochim. Biophys. Acta (BBA)- Biomembr. **1838**(6), 1509–1517 (2014)
7. Piotto, S.P., Sessa, L., Concilio, S., Iannelli, P.: YADAMP: yet another database of antimicrobial peptides. Int. J. Antimicrob. Agents **39**(4), 346–351 (2012)
8. Scrima, M., Di Marino, S., Grimaldi, M., Campana, F., Vitiello, G., Piotto, S.P., D'Errico, G., D'Ursi, A.M.: Structural features of the C8 antiviral peptide in a membrane-mimicking environment. Biochim. Biophys. Acta (BBA)- Biomembr. **1838**(3), 1010–1018 (2014)
9. Piotto, S., Nesper, R.: CURVIS: A program to study and analyse crystallographic structures and phase transitions. J. Appl. Crystallogr. **38**(1), 223–227 (2005)
10. Torok, Z., Crul, T., Maresca, B., Schutz, G.J., Viana, F., Dindia, L., Piotto, S., Brameshuber, M., Balogh, G., Peter, M., Porta, A., Trapani, A., Gombos, I., Glatz, A., Gungor, B., Peksel, B., Vigh Jr., L., Csoboz, B., Horvath, I., Vijayan, M.M., Hooper, P.L., Harwood, J.L., Vigh, L.: Plasma membranes as heat stress sensors: from lipid-controlled molecular switches to therapeutic applications. Biochim. Biophys. Acta **1838**(6), 1594–1618 (2014)
11. Crul, T., Toth, N., Piotto, S., Literati-Nagy, P., Tory, K., Haldimann, P., Kalmar, B., Greensmith, L., Torok, Z., Balogh, G., Gombos, I., Campana, F., Concilio, S., Gallyas, F., Nagy, G., Berente, Z., Gungor, B., Peter, M., Glatz, A., Hunya, A., Literati-Nagy, Z., Vigh Jr., L., Hoogstra-Berends, F., Heeres, A., Kuipers, I., Loen, L., Seerden, J.P., Zhang, D., Meijering, R.A., Henning, R.H., Brundel, B.J., Kampinga, H.H., Koranyi, L., Szilvassy, Z., Mandl, J., Sumegi, B., Febbraio, M.A., Horvath, I., Hooper, P.L., Vigh, L.: Hydroximic acid derivatives: pleiotropic HSP co-inducers restoring homeostasis and robustness. Curr. Pharm. Des. **19**(3), 309–346 (2013)

12. Gombos, I., Crul, T., Piotto, S., Gungor, B., Torok, Z., Balogh, G., Peter, M., Slotte, J.P., Campana, F., Pilbat, A.M., Hunya, A., Toth, N., Literati-Nagy, Z., Vigh Jr., L., Glatz, A., Brameshuber, M., Schutz, G.J., Hevener, A., Febbraio, M.A., Horvath, I., Vigh, L.: Membrane-lipid therapy in operation: the HSP co-inducer BGP-15 activates stress signal transduction pathways by remodeling plasma membrane rafts. PLoS ONE **6**(12), e28818 (2011)
13. Vigh, L., Torok, Z., Balogh, G., Glatz, A., Piotto, S., Horvath, I.: Membrane-regulated stress response: a theoretical and practical approach. In: Vigh, L., Csermely, L. (eds.) Molecular Aspects of the Stress Response: Chaperones, Membranes and Networks. Advances in experimental medicine and biology, vol. 594, pp. 114–131. Springer Science + Business Media, New York (2007)
14. Lange, O.F., Lakomek, N.-A., Farès, C., Schröder, G.F., Glatz, A., Walter, K.F., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C., De Groot, B.L.: Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science **320**(5882), 1471–1475 (2008)
15. Mittermaier, A., Kay, L.E.: New tools provide new insights in NMR studies of protein dynamics. Science **312**(5771), 224–228 (2006)
16. Tuffery, P., Derreumaux, P.: Flexibility and binding affinity in protein–ligand, protein–protein and multi-component protein interactions: limitations of current computational approaches. J. Roy. Soc. Interface **9**(66), 20–33 (2012)
17. Chou, K.C.: Biological functions of low-frequency vibrations (phonons). III. Helical structures and microenvironment. Biophys. J. **45**(5), 881–889 (1984)
18. Piotto, S., Biasi, L.D., Concilio, S., Castiglione, A., Cattaneo, G.: GRIMD: distributed computing for chemists and biologists. Bioinformation **10**(1), 43–47 (2014)
19. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The protein data bank. Nucleic Acids Res. **28**(1), 235–242 (2000)
20. Krieger, E., Vriend, G.: YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics **30**(20), 2981–2982 (2014)
21. Trott, O., Olson, A.J.: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. **31**(2), 455–461 (2010)