Fragment Based Molecular Dynamics for Drug Design

Lucia Sessa[1(✉)], Luigi Di Biasi[1], Simona Concilio[2], 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 Industrial Engineering, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy

Abstract. Molecular docking is a computationally efficient method used to predict the conformations adopted by the ligand within a target-binding site. A positive aspect of conventional docking is the possibility of easily distributing the calculation on dedicated grid or cluster. The receptor is usually kept rigid, therefore the changes in the binding pocket geometry induced by the ligand is overlooked. Here we present a new docking approach (DynDock) that exploits molecular dynamics to preserve the flexibility of the receptor. To maintain high computational efficiency, DynDock has been developed to be distributed on a grid. The main advantages of this method are the full flexible molecular docking achieved during the simulation and the reduced number of compounds collected.

Keywords: Docking · Drug design · Molecular dynamics

1 Introduction

The molecular design is a computationally demanding task; it is the process of finding new drugs and involves the design of molecules that are complementary to the target in shape and charge. Usually, these compounds interact with a protein activating or inhibiting its function. There are two major methods of molecular design. The first is the Ligand-Based Drug Design (LBDD) that uses the structural characteristics of all molecules that bind the target of interest, to derive a pharmacophore model [1]. The second method is the Structure-Based Drug Design (SBDD), which is based on knowledge of the three-dimensional structure of the target [2]. The aim is to predict the affinity and the selectivity of a drug candidate using the ligand and the target structure.

In details, SBDD is a cyclic process, which starts from a known target structure usually experimentally obtained by X-ray crystallography or NMR spectroscopy [3]. The knowledge of 3D structures permits to run *in silico* studies to identify potential ligands (Fig. 1).

Following the molecular modelling predictions, the most promising compounds can be synthesized and evaluated for their biological properties. Once synthesized and

M. Pelillo et al. (Eds.): WIVACE 2017, CCIS 830, pp. 49–58, 2018.
https://doi.org/10.1007/978-3-319-78658-2_4

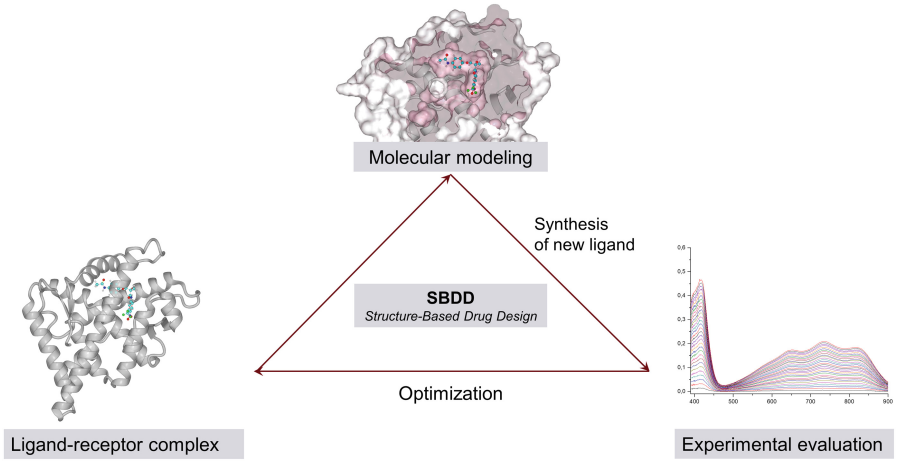

Fig. 1. Structure-based drug design cycle

tested, the new 3D structures can be solved and made ready for a further optimization cycle. This process reasonably permits to increase the affinity of new ligands but it is extremely costly.

Another limit of SBDD is that the experimental structures of the complex receptor/ligand are not always available and, when accessible, we must take into account that ligands can induce conformational changes in proteins and different ligands may stabilize different receptor conformations. Nevertheless, the crystallographic data represents only a successful binding event of a specific protein conformation and a specific ligand [4].

For these reasons, it is clear that the flexibility of the target receptor is an essential aspect that must be considered in the docking studies and it is not recommended to use only one structure of the receptor to perform the analysis.

Molecular dynamics (MD) is also a useful technique to evaluate critical phenomena and conformational changes involved in the molecular interactions [5, 6]. Keeping the proteins flexible in the molecular modelling studies has a high computational cost. The most popular docking programs limit the receptor flexibility to side-chain mobility only [7, 8]. In some other cases, they consider several snapshots extracted from a molecular dynamics of the receptor. This approach assumes that the protein flexibility could be encoded in an arbitrary set of MD conformations, but the molecular dynamics is strongly dependent on the ligand nature [4, 9].

The computational complexity of the procedure grows quickly with the numbers of atoms in the ligands. An exhaustive analysis of all possible ligands is far impossible even when a molecule is simplified in groups of atoms or residues. For example, the investigation of a very short peptide of only 10 amino acids, having as a starting point only 3 conformations (alpha, beta and coil), leads to more than $6 \cdot 10^7$ possible sequences, a number beyond the current computational possibilities.

DynDock employs an *in silico* combinatorial molecular dynamics to optimize ligands inside the target protein. This procedure combines the advantages of the SBDD

method with the accuracy of MD, reducing drastically the number of the possible sequences to analyse.

2 Methods

Figure 2 describes the DynDock approach.

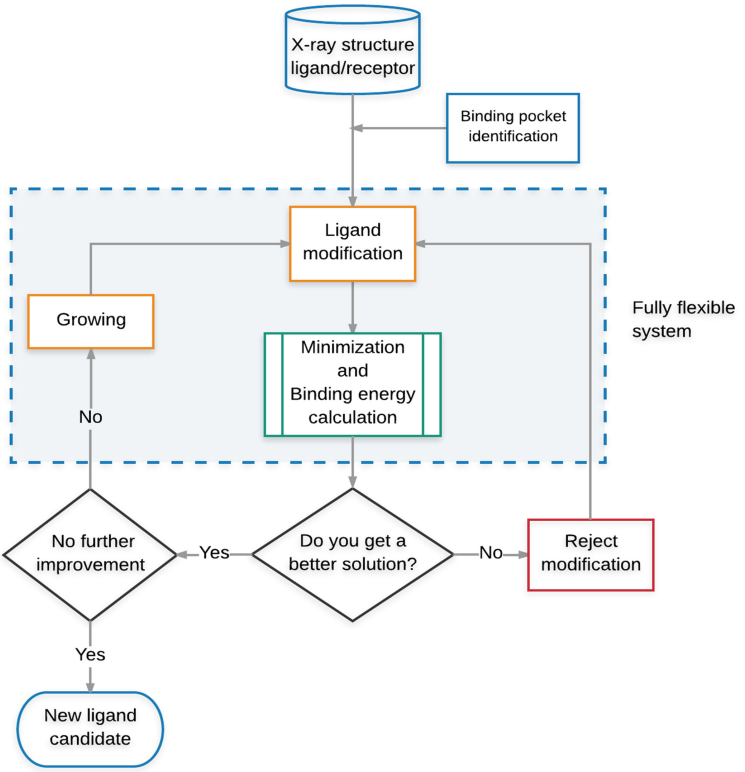

Fig. 2. DynDock flowchart

The DynDock is a hierarchical method to design a ligand candidate. We start from a fragment of a known ligand and we proceed with a cycle in which we evaluate the effect of a set of possible moves. The single move can be an addition, a deletion, an atom or residue swap, or a cyclization. After the move, we perform a molecular dynamics run of 1 ns and then we anneal the system. After the relaxation, we calculate the binding energy of the ligand and the distortion energy of the receptor. At the end of the cycle, the ligand having the highest binding energy is chosen for a further cycle. The procedure ends when no further energy improvement is observed. DynDock ensures to find always a better ligand, though it cannot guarantee to find the best one.

2.1 Fragment-Based Molecular Dynamics

The starting point for DynDock is the structure of a target protein bound to a ligand. In this work, we used as initial structures the microbial enzyme *Streptomyces griseus* Protease A (SGPA) [10] in complex with the peptide Pro-Ala-Pro-Tyr (5SGA PDB). It is a proteolytic enzyme with a serine residue (Ser 195) in its active site. The proximity of a histidine and an aspartate is essential for its activity (see the ligand diagram interaction in Fig. 2a) [11, 12].

The first step of DynDock is the preparation of the receptor structure for the docking, removing all water molecules and adding missing hydrogens. From the experimental structure of the ligand/receptor complex, it is possible to identify the binding site in which the ligand is grown. From literature data, we know that the active site of SGPA is an external portion of the receptor and the residue of tyrosine of the ligand interacts with the catalytic triad. In Fig. 2b, it is possible to see the position of the ligand PAPY on the external surface of the enzyme. The surface is colored by atom types. To put in evidence the active site, we colored only the residue with distance 4 Å from the ligand (Fig. 3).

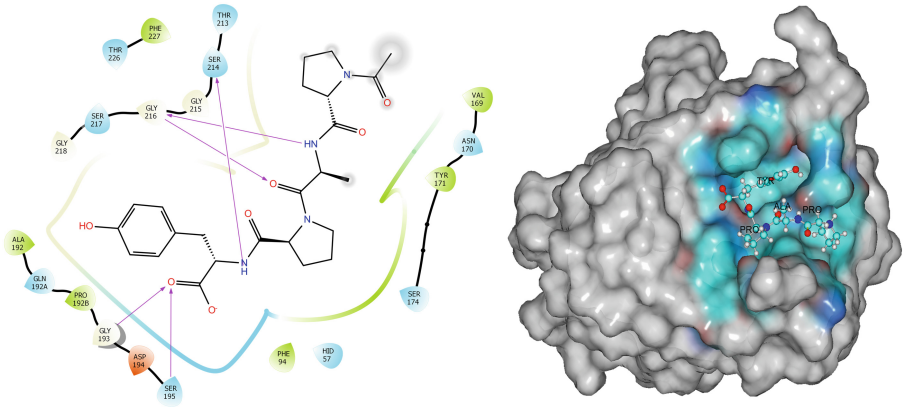

Fig. 3. (a) 2D interaction between the ligand and the enzyme. (b) Molecular surface of Serine protease in complex with PAPY inhibitor (5SGA PDB). (Color figure online)

The tyrosine residue is critical for the enzyme inhibition, and therefore, we have chosen the terminal Tyr as starting point for peptide elongation.

2.2 Preliminary Screening

The elongation procedure consists in adding an amino acid chosen among the 20 natural residues. Each amino acid, when binds a peptide, forms an amide bond characterized by two torsion angles. They describe the rotations of the polypeptide backbone around the bonds between N-Cα (φ) and Cα-C (ψ). It is well known that three regions of φ, ψ correspond to the most stable conformations namely α-helices, β-sheet

and coil. Therefore, for each residue added by DynDock, we must take into account the different geometries, and the number of possible structures becomes:

$$peptides = (conf * aa)^{elongation} \quad (1)$$

Where *conf* is the number of conformational regions, and *aa* is the number of amino acids. That for 3 geometries only and for an elongation of 10 residues, gives:

$$peptides = (3 \cdot 20)^{10} \cong 6.05 \cdot 10^{17} \quad (2)$$

Though the molecular dynamics approach promises a big progress in the docking field, it is evident that the astronomic number of possible peptides renders this way unmanageable. The DynDock approach greatly reduces the number of possible peptides and only the best *n*-mers can be kept for further elongation. If only the peptide with the highest binding energy is kept for each iteration, the number of peptides to analyze may collapse to:

$$peptides = elongation \cdot conf \cdot aa = 10 \cdot 3 \cdot 20 = 600 \quad (3)$$

Unfortunately, an important aspect of peptide folding was neglected in the above consideration. The interaction of a peptide with a receptor is not a simple cumulative process because the peptide residues can interact with the peptide itself changing conformation and then altering the binding with the receptor. Consequently, we cannot keep the best residue only for each elongation step, but we can safely choose to keep the a few number of peptides (*bestResults*) for each step. The number of possible peptides for *BestResults* = 5 becomes:

$$peptides = bestResults \cdot elongation \cdot conf \cdot aa = 5 \cdot 10 \cdot 3 \cdot 20 = 3000 \quad (4)$$

Considering the last 2 or 3 residues (*nmer*) during an elongation should give more realistic results, but the number of possible peptides would reach soon extremely large numbers (see Table 1) according to Eq. (5).

$$peptides = bestResults \cdot elongation \cdot (conf \cdot aa)^{nmer} \quad (5)$$

We have faced this problem in two different ways. The first is the distribution of the calculation on a dedicated grid. We have used GRIMD [13], an info structure that permits easily the delivery of molecular dynamics calculation on available PCs. The second solution is more chemistry-oriented. A preliminary screening is made before the

Table 1. Number of sequences to analyze based on the number of residues kept in memory

Elongation	Sequences to analyze
1 amino acid	3000
2 amino acids	180000
3 amino acids	$1.08 \cdot 10^7$

ligand fragment building. This step is essential to reduce the time of the entire drug design. The user can set the elongation (e.g. 6 amino acids) and the solubility of the ligand candidate. The solubility is an important parameter to be considered, in order to ensure the possibility of chemical synthesis and the biological screening. Water solubility can be predict as a function of the surface hydrophobicity of the ligand. The tendency of a protein to aggregate and so to decrease its water solubility can be related to the hydrophobic surface [14]. Ligand candidates with potential low water solubility are not considered.

Finally, in order to avoid too exotic peptides, we introduced a phylogenetic control of sequences. We downloaded the human proteome from the UniProt database [15] (Proteome ID UP000005640) and we calculated the dimer abundance. We assigning a different weight based on dimer probabilities.

The DynDock method favors the building of dimers with high frequency.

2.3 Ligand Fragment Modification

The ligand fragment modification starts considering a swap, deletion and addition of amino acid residues.

The ligand modification step occurs during molecular dynamics simulations (MD), which means the ligand fragment and the receptor, are always in contact. A cubic cell of 57 × 57 × 57 Å was built around all atoms under periodic boundary condition. The MD simulation is set at 298 K with 1.25 fs of integration time steps for intramolecular forces. After each modification, the system is left to move for 1 ns to allow the receptor to better accommodate the ligand. The all-atom structure of the complex fragment/target is minimized using the force field AMBER14 [16] and the steepest descent minimization followed by a simulated annealing minimization [17].

2.4 Output Selection

The binding affinity of the ligand was calculated using the function *YaEnergy* already reported in [18]. *YaEnergy* permits to estimate the binding energy taking into account the biological history of the receptor. It has been written after an extensive genetic algorithm including a term that depends on the minimal distance between the ligand barycenter and the nearest conserved residues. The sequence of the enzyme SGPA is highly conserved through species indicating that the sequence has been maintained by evolution despite speciation. As shown in Fig. 4, the residues of the binding pocket are extremely conserved confirming that functional residues are generally more preserved [19]. The conservation string was obtained from the Consurf database [20], a server for identification of structurally important residues in protein sequences (http://conseq.tau.ac.il/).

The binding energy calculated at the end of the molecular dynamics was used to build up a new ranking function for peptide selection. Whereas the choice of high binding energies is straightforward, energy alone is not enough because tends to bias longer peptides. A long peptide, in fact, can interact in more ways than a shorter one. The rank function at denominator has the peptide length and a negative surface area term at numerator. There is also a corrective term based on the receptor distortion.

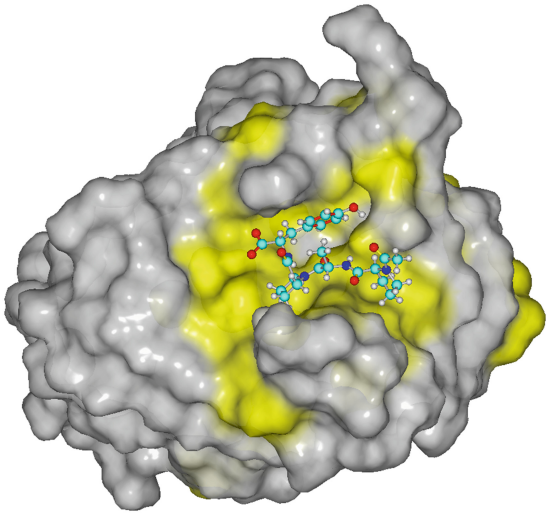

Fig. 4. 5SGA surface. The conserved residues are colour mapped in yellow onto the protein surface. (Color figure online)

The rationale is that a peptide, modifying the 3D structure of the flexible receptor, can drive its geometry far from the experimental data. In the ranking function, we have added a term to award receptor structures not dissimilar from the crystallographic data.

$$DynDock\ rank = \frac{yaEnergy + 20 \cdot \frac{En.end}{En.in} - \frac{SurfArea}{90}}{Length} \tag{6}$$

Whereas *YaEnergy* is the peptide binding energy, *En.in* and *En.end* are the initial and final energies of the receptor calculated with the force field AMBER14, and *SurfArea* is the molecular surface area of the peptide.

DynDock selects the molecules with the highest binding energy and lowest deformation of the receptor. The receptor changes its structure during the fragment growing to better accommodate the ligand. This could damage the 3D structure and lead to an unrealistic structure. For this reason, the DynDock rank function is rescaled on the dimension of the ligands, optimizing the binding energy value with the ligand surface area. This is essential to prevent that the algorithm prefers bigger peptides that, having more atoms, have more chances to interact with the receptor.

Based on the DynDock rank value, the step that involves the ligand modification can be accepted or rejected.

The process ends when there are no further energy improvements. 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 the calculation [13].

3 Results and Discussion

The evaluation of the DynDock procedure can be done in terms of binding energy. Interestingly, following the methods herein described, after binding each peptide can be forced to leave the receptor and the activation energy required to leave the receptor estimated. This calculation permits to evaluate the residence time [21] of the ligand (that is the inverse of the unbinding kinetic constant rate k_{off}) [22–24] that is of fundamental importance in drug discovery.

Starting from the first amino acid (tyrosine), DynDock adds new amino acids and chooses the best dimer among the 20 possibilities. The dimer was then further elongated until the length of 4 residues. All ligands are ordered by length and by ranking value. Based on the rank function value, DynDock selects only two dimers made, in this example, glutamine-tyrosine (QY) and alanine-tyrosine (AY) and proceeds further with the elongation. We decided to fix the elongation to 4 residues to make easier the comparison with the crystallographic structure. The best ligand developed by DynDock protocol (PGAY) shows higher binding affinity that the experimental molecule PAPY and it still maintains the interactions with the catalytic triad (see Fig. 5).

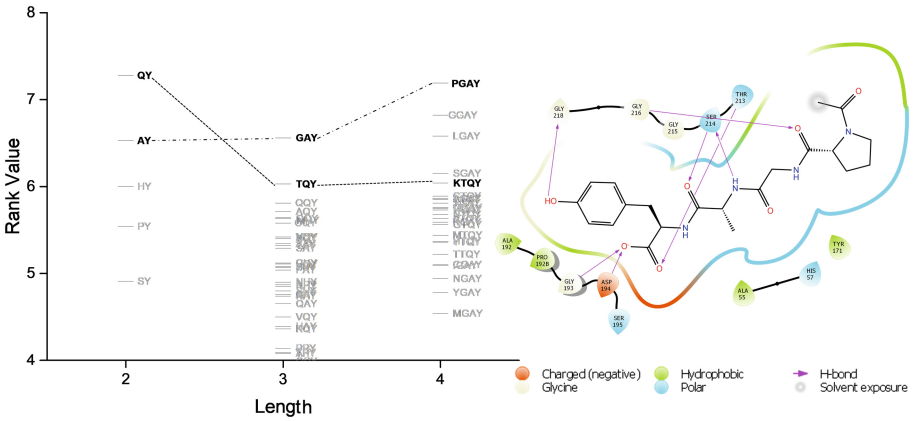

Fig. 5. DynDock protocol trend. Length is the number of amino acids in the ligand and the rank function is the normalized binding energy. On the left are shown the ligand PGAY interactions with the binding site.

To build a ligand candidate formed by ten residues, traditional approaches of structure-based drug design provide 10 million of millions of sequences to analyze ($20^{10} = 1{\cdot}10^{13}$). DynDock method limits the calculation to a number of sequences to analyze of few thousands.

4 Conclusion

The method here described for Drug Design is an easy way to perform fully molecular docking reducing drastically the number of possible sequences. It permits also the easy distribution on computer grids to further reduce the analysis time.

This hierarchical approach has several advantages over traditional docking. First, the flexibility of the receptor, essential for its function, is fully considered and modeled with the modern AMBER14 force field. Second, DynDock takes into account the receptor distortion to avoid unrealistic and improbable interactions. Third, the sequential procedure guarantees to find a series of peptides with high binding energies without a sensible decrease of computational performances. Fourth, the sequentiality of the investigation makes DynDock ideal for parallelization or for use on grids. Finally, the molecular dynamics can be used also to perform a steered molecular dynamics of the ligand out from the receptor to estimate the residence time. This improvement will be the object of an upcoming paper.

References

1. Sliwoski, G., Kothiwale, S., Meiler, J., Lowe, E.W.: Computational methods in drug discovery. Pharmacol. Rev. **66**, 334–395 (2014)
2. Salum, L.B., Polikarpov, I., Andricopulo, A.D.: Structure-based approach for the study of estrogen receptor binding affinity and subtype selectivity. J. Chem. Inf. Model. **48**, 2243–2253 (2008)
3. Ferreira, L.G., Dos Santos, R.N., Oliva, G., Andricopulo, A.D.: Molecular docking and structure-based drug design strategies. Molecules **20**, 13384–13421 (2015)
4. Sessa, L., Biasi, L.D., Concilio, S., Cattaneo, G., De Santis, A., Iannelli, P., Piotto, S.: A new flexible protocol for docking studies. Commun. Comput. Inf. Sci. **587**, 117–126 (2016)
5. Lin, J.-H.: Accommodating protein flexibility for structure-based drug design. Curr. Top. Med. Chem. **11**, 171–178 (2011)
6. Durrant, J.D., McCammon, J.A.: Molecular dynamics simulations and drug discovery. BMC Biol. **9**, 71 (2011)
7. de Ruyck, J., Brysbaert, G., Blossey, R., Lensink, M.F.: Molecular docking as a popular tool in drug design, an in silico travel. Adv. Appl. Bioinf. Chem. AABC **9**, 1 (2016)
8. Geng, C., Narasimhan, S., Rodrigues, J.P., Bonvin, A.M.: Information-driven, ensemble flexible peptide docking using HADDOCK. In: Schueler-Furman, O., London, N. (eds.) Modeling Peptide-Protein Interactions. MMB, vol. 1561, pp. 109–138. Springer, New York (2017). https://doi.org/10.1007/978-1-4939-6798-8_8
9. Sessa, L., Concilio, S., Piotto, S.: Molecular dynamics and morphing protocols for high accuracy molecular docking. In: Piotto, S., Rossi, F., Concilio, S., Reverchon, E., Cattaneo, G. (eds.) Advances in Bionanomaterials. LNB, pp. 85–96. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-62027-5_8
10. James, M.N.G., Sielecki, A.R., Brayer, G.D., Delbaere, L.T.J., Bauer, C.A.: Structures of product and inhibitor complexes of Streptomyces griseus protease A at 1.8 Å resolution: a model for serine protease catalysis. J. Mol. Biol. **144**, 43–88 (1980)
11. Bartholomae, M., Buivydas, A., Viel, J.H., Montalban-Lopez, M., Kuipers, O.P.: Major gene-regulatory mechanisms operating in ribosomally synthesized and post-translationally modified peptide (RiPP) biosynthesis. Mol. Microbiol. **106**(2), 186–206 (2017)

12. Harish, B., Uppuluri, K.B.: Microbial serine protease inhibitors and their therapeutic applications. International J. Biol. Macromol. **107**, 1373–1387 (2017)
13. Piotto, S., Di Biasi, L., Concilio, S., Castiglione, A., Cattaneo, G.: GRIMD: distributed computing for chemists and biologists. Bioinformation **10**, 43–47 (2014)
14. Wagner, J.R., Sorgentini, D.A., Añón, M.C.: Relation between solubility and surface hydrophobicity as an indicator of modifications during preparation processes of commercial and laboratory-prepared soy protein isolates. J. Agric. Food Chem. **48**, 3159–3165 (2000)
15. UniProt Consortium: UniProt: the universal protein knowledgebase. Nucleic Acids Res. **45**, D158–D169 (2016)
16. Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y.: CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. J. Chem. Theory Comput. **12**, 405–413 (2015)
17. Krieger, E., Vriend, G.: YASARA view-molecular graphics for all devices-from smartphones to workstations. Bioinformatics **30**, 2981–2982 (2014)
18. Di Biasi, L., Fino, R., Parisi, R., Sessa, L., Cattaneo, G., De Santis, A., Iannelli, P., Piotto, S.: Novel algorithm for efficient distribution of molecular docking calculations. Commun. Comput. Inf. Sci. **587**, 65–74 (2016)
19. Piotto, S., Di Biasi, L., Fino, R., Parisi, R., Sessa, L., Concilio, S.: Yada: a novel tool for molecular docking calculations. J. Comput. Aided Mol. Des. **30**, 753–759 (2016)
20. Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R., Ben-Tal, N.: ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics **20**, 1322–1324 (2004)
21. Copeland, R.A., Pompliano, D.L., Meek, T.D.: Drug–target residence time and its implications for lead optimization. Nat. Rev. Drug Discovery **5**, 730–739 (2006)
22. Piotto, S., Sessa, L., Iannelli, P., Concilio, S.: Computational study on human sphingomyelin synthase 1 (hSMS1). Biochim. Biophys. Acta (BBA) Biomembr. **1859**, 1517–1525 (2017)
23. Casas, J., Ibarguren, M., Álvarez, R., Terés, S., Lladó, V., Piotto, S.P., Concilio, S., Busquets, X., López, D.J., Escribá, P.V.: G protein-membrane interactions II: effect of G protein-linked lipids on membrane structure and G protein-membrane interactions. Biochim. Biophys. Acta (BBA) Biomembr. **1859**, 1526–1535 (2017)
24. 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**, 1509–1517 (2014)