Computational and Structural Biotechnology Journal 21 (2023) 3002–3009

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

Computational structural-based GPCR optimization for user-defined ligand: Implications for the development of biosensors

Lorenzo Di Rienzo [a,*], Mattia Miotto [a], Edoardo Milanetti [b,a], Giancarlo Ruocco [a,b]

[a] Center for Life Nano- & Neuro-Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161 Rome, Italy
[b] Department of Physics, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy

ARTICLE INFO

Article history:
Received 3 January 2023
Received in revised form 17 April 2023
Accepted 4 May 2023
Available online 9 May 2023

Keywords:
Bio-sensors
Protein engineering
GPCR

ABSTRACT

Organisms have developed effective mechanisms to sense the external environment. Human-designed biosensors exploit this natural optimization, where different biological machinery have been adapted to detect the presence of user-defined molecules. Specifically, the pheromone pathway in the model organism Saccharomyces cerevisiae represents a suitable candidate as a synthetic signaling system. Indeed, it expresses just one G-Protein Coupled Receptor (GPCR), Ste2, able to recognize pheromone and initiate the expression of pheromone-dependent genes. To date, the standard procedure to engineer this system relies on the substitution of the yeast GPCR with another one and on the modification of the yeast G-protein to bind the inserted receptor. Here, we propose an innovative computational procedure, based on geometrical and chemical optimization of protein binding pockets, to select the amino acid substitutions required to make the native yeast GPCR able to recognize a user-defined ligand. This procedure would allow the yeast to recognize a wide range of ligands, without a-priori knowledge about a GPCR recognizing them or the corresponding G protein. We used Monte Carlo simulations to design on Ste2 a binding pocket able to recognize epinephrine, selected as a test ligand. We validated Ste2 mutants via molecular docking and molecular dynamics. We verified that the amino acid substitutions we identified make Ste2 able to accommodate and remain firmly bound to epinephrine. Our results indicate that we sampled efficiently the huge space of possible mutants, proposing such a strategy as a promising starting point for the development of a new kind of S.cerevisiae-based biosensors.

1. Introduction

The development of sensors able to sensitively and selectively detect environmental small molecules is of paramount importance for synthetic biology applications and it is still a difficult task [1,2]. Monitoring the presence of some molecules is important in many fields, such as environmental, medical, or food sample control [3,4].

However, sensing the surrounding environment is one of the most important tasks a cell has to fulfill in order to perform its physiological activities. Therefore, evolution devised very effective mechanisms to recognize the presence of a huge variety of chemicals and to respond to such stimuli. From this point of view, biosensors are built to exploit naturally evolved biological elements able to signal target chemicals [5], obviously modified to detect user-defined molecules. In the past years, many protein systems have been chosen to work as sensors: allosteric transcription factors [6], binding proteins that undergo computational design [7,8] or others [9–11].

Among these, G-Protein Coupled Receptors (GPCRs) are good candidates as synthetic signaling systems [12]. These proteins represent the largest family of trans-membrane proteins, expressed in virtually any tissue [13]. Indeed, the binding with a ligand in the extracellular environment promotes a protein conformational change. Thus, the associated G-protein can detect this switch in the cytosol. These proteins are therefore associated with a very wide variety of physiological and pathological responses [14–16]. Moreover, GPCRs share a very conserved architecture, a seven-helices trans-membrane domain. They are usually encoded by small genes, making the transfer of engineered GPCRs in other tissues or species easy [17].

In this framework, *Saccharomyces cerevisiae* constitutes one of the most used models in biotechnological research. Indeed, such yeast is characterized by a high growth rate and stability, and low

* Corresponding author.
E-mail address: lorenzo.dirienzo@iit.it (L. Di Rienzo).

https://doi.org/10.1016/j.csbj.2023.05.004

cost. These characteristics make its engineering very useful in performing both theoretical and technological settings [18]. In addition, the easiness of yeast genetic modifications makes it a cell system very suitable for the deletion, installation, and transformation of genes and genetic circuits, even more with the advent of CRISPR genome editing [18,19].

Interestingly, the haploid yeast of mating type MATa expresses just one GPCR (Ste2), whose function is to control the *pheromone* pathway [20,21]. As a consequence, the engineering of such a yeast represents an established and promising strategy [22,23]. Indeed, the pheromone pathway can be modified with two main steps. First, the native yeast GPCR has to be substituted with another one, able to recognize the user-defined small molecule. Later, the last 5 residues of the Gα associated with the inserted GPCR have to be introduced in the native yeast Gα. This strategy can pair the inserted GPCR with the yeast signaling pathway, while secondary modifications in the yeast can furnish experimental evidence of the GPCR-small molecules activation [22,24].

However, this strategy presents two major limitations. Indeed, it is necessary the prior knowledge of the ligand-associated GPCR, effectively limiting the set of possible chemicals that can be considered for engineering. In addition, also the association between GPCR and the corresponding G-protein has to be known *a-priori*, a condition not always satisfied.

To overcome these limitations, we developed a computational procedure aiming to select the minimum set of amino acid substitutions that make the native yeast GPCR Ste2 able to recognize a user-defined ligand. This result is achieved by modifying locally the wild-type (WT) GPCR to mimic the geometry and the physical-chemicals features of a protein binding pocket experimentally observed to be in interaction with such a small molecule.

However, the activation mechanism of Ste2 has been recently experimentally studied in great detail [25], revealing the conformational changes that Ste2 undergoes upon agonist ligand binding. The modification of the functional agonist-Ste2 pair has to be performed finely, to preserve the possibility for the selected small molecule to activate the GPCR.

Once identified the compound, we preliminary identify in the literature a protein-ligand interaction whose structural details have been previously deposited on Protein Data Bank [26]. Here, we select epinephrine as the test ligand, whose interaction with the beta2 adrenoceptor has been determined in X-ray crystallography [27].

This protocol aims to mutate Ste2 GPCR to reproduce the geometrical and physicochemical properties of the protein pocket experimentally observed in interaction with the compound. In particular, on one side we describe the geometry of protein regions employing a method based on the 2D Zernike polynomials formalism [28]. In this protocol, we describe a portion of molecular surface as a 2D function and we expand it on the basis of such polynomials: the norms of the expansion coefficients are an ordered set of numerical descriptors, compactly summarizing the geometrical properties of a protein region. In the past years, such a description, independent of the orientation of the proteins in the space, has proven their efficacy in several molecular systems and applications, including the GPCR-ligand interaction [29–35]. Moreover, it is worth noting that the compactness of Zernike descriptors permits an easy evaluation of patch similarity, adopting a metric to calculate the dissimilarity between two sets of descriptors. This feature allows us to use Zernike descriptors in computational engineering protocols we recently developed [36–38]. In addition, we quantify the chemical properties of surface regions using an *in-silico* derived amino acid hydrophobicity scale [39]. Each surface point is labeled with the hydrophobicity of the residue generating it, where the hydrophobicity of the pocket is obtained by averaging the values regarding all the region surface points.

Our computational protocol consists of several steps. Initially, we search on the Ste2 inner surface for the region most similar to the one experimentally seen in interaction with epinephrine, in terms of both shape and hydrophobicity descriptors. Selected that region as a template, we perform ten Monte Carlo (MC) simulations to optimize the similarity with the experimental pocket. In particular, with computational mutagenesis, we randomly mutate a residue calculating in each step the new shape and hydrophobicity descriptors. Each mutation is accepted according to a cost function taking into account both the similarity with the experimental pocket in terms of shape and hydrophobicity and a penalty for the introduction of mutations with respect to the Ste2 WT. As a result, we identify a set of Ste2 mutants characterized by a significantly increased similarity with the epinephrine experimental pocket and a reduced number of mutations with respect to WT Ste2.

To test the effectiveness of our protocol, we performed molecular docking and molecular dynamics simulations. Indeed, we docked epinephrine on the examined Ste2 region, both for WT and optimized versions of the GPCR. Molecular dynamics simulations, performed with the best docking poses obtained for both WT and optimized proteins, confirm that the residue substitutions introduced with our protocol make the Ste2-epinephrine complex stable.

Such results indicate that procedure is efficient in the sampling of the huge space of possible GPCR mutants, outputting sets of mutations that make the GPCR able to interact with a given ligand. Therefore, this strategy could represent a promising starting point for the development of S.cerevisiae-based biosensors.

2. Results and discussions

2.1. Computational protocol

In this section, we describe the computational protocol we developed to identify the amino acid substitutions to make Ste2 GPCR able to recognize epinephrine. The main steps involved are schematically shown in Fig. 1.

Preliminary, it is necessary to identify a protein-ligand interaction whose structural details are known. For this case, we select the beta2 adrenoceptor-epinephrine complex (PDB code: 4ldo) [27]. After the computation of both the protein and ligand molecular surfaces, we define the epinephrine binding site (EBS) as the set of protein molecular surface points whose distance from any ligand surface point is lower than 3 Å (Fig. 1.a). We aim to modify Ste2 to design a region as similar as possible to this EBS, in terms of geometry and chemistry. To do this, we computed the 2D Zernike Descriptors of this EBS, summarizing its morphological properties in 121 ordered numerical coefficient [28]. Moreover, using the hydrophobicity scale we developed in [39], we assigned to each EBS surface point the hydrophobicity of the residue generating it. We obtain the EBS mean hydrophobicity averaging over all the surface points constituting it.

Thus, we need to select where on Ste2 it is reasonable to perform our protocol. Recently it has been determined the x-ray structure of the Ste2-pheromone molecular complex (PDB code: 7ad3) [40]. From this structure, we extract the set of residues involved in pheromone recognition (closer to any pheromone atoms more than 4 Å). Since these residues are in interaction with the physiological ligand, this region constitutes the potential area of design.

However, the complex structure furnishes a bound configuration of the Ste2 GPCR. Since we aim to design starting from the unbound configuration, we adopted as template structure the Ste2 unbound structure, as furnished by Alpha Fold [41]. We hence calculated the molecular surface of Ste2 and we sampled the regions regarding the residues we previously saw in interaction with the physiological ligand (Fig. 1.b). Indeed, we selected in each surface all the points

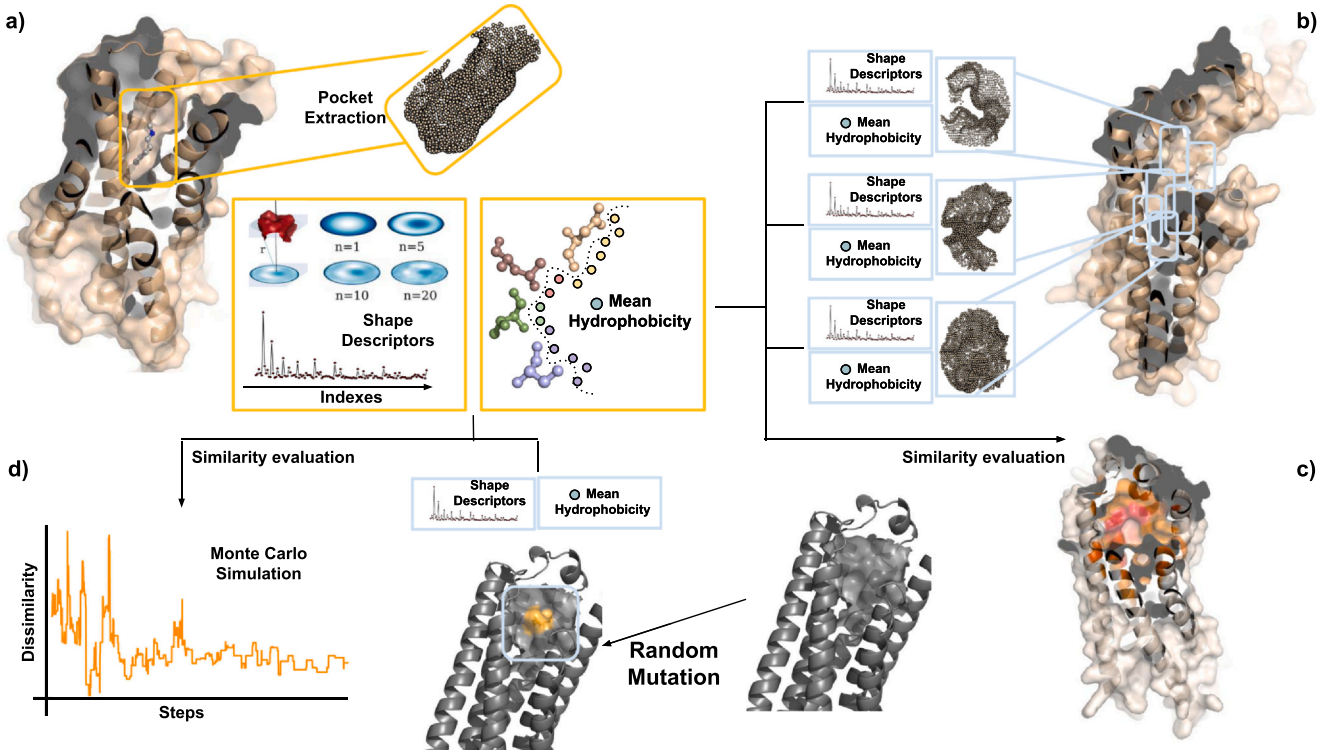

Fig. 1. Overview of the computational pipeline. a) Cartoon representation of the beta2 adrenoceptor-epinephrine complex (pdb code: 4ldo). After the computation of both the protein and ligand molecular surfaces, epinephrine binding site (EBS) is defined as the set of protein molecular surface points whose distance from any ligand surface point is lower than 3 Å. Shape and hydrophobicity of the extracted pocket are evaluated. **b)** Cartoon representation of Ste2 protein. The protein patches located in the protein region where pheromone is bound were extracted and characterized with shape and hydrophobicity descriptors. They were compared with the descriptors of EBS. **c)** Ste2 surface colored according to the local similarity with EBS. **d)** In the Monte Carlo optimization procedure the dissimilarity between the designed pocket and EBS is minimized, in terms of the shape and hydrophobicity descriptors.

generated by the ligand recognition involved residues, and we build around all of them a patch considering all the surface points with a distance to the patch center lower than 7.5 Å. This radius has been chosen to construct patches of the same size as EBS. In this way we defined over 15,000 patches, thus characterizing their properties with both Zernike shape descriptors and mean hydrophobicity.

Therefore, we compared all these patches with EBS, in terms of shape and hydrophobicity. In the molecular representation of Fig. 1.C we schematically report the results of such analysis, where the redder the molecular surface of Ste2 is the more the corresponding patch is similar to EBS. We, therefore, used the most similar patch among these studied as the starting point for our design protocol (see next section for the detail on how we made such a choice). In this way, we make sure to start with the best possible patch, so that the optimization protocol can start from an advantageous template. The selected patch is composed of the following residues: ASN 132, ASP 124, GLN 135, GLN 51, HIS 94, LEU 102, SER 104, SER 107, THR 131, THR 192, TYR 101, TYR 106, TYR 128, TYR 98, VAL 127.

Lastly, we run 10 independent MC simulations to make the selected regions as similar as possible to EBS (See Fig. 1.d). In each step one of the previously listed residues is mutated in one of the other possible 19. The physicochemical descriptors of the mutated surface were calculated and compared with EBS. The mutations have been accepted according to a cost function studying the similarity with EBS, as well as a penalty term for the introduction of a new mutation. This ensures to not change too much the original Ste2 sequence.

In the next section, we will first discuss the details of the choice of the starting template region and later the technicalities of our MC simulations. Later, we will see how the optimization protocol outputs regions with improved similarity with EBS in terms of several characteristics. Lastly, we will examine the outcome of independent docking and molecular dynamics simulations.

2.2. Selection of the Ste2 design region and analysis of Monte Carlo exploration

The selection of a region on the Ste2 surface to undergo the MC optimization processes designed is an important issue. Indeed, the probability to design a binding site that actually can bind epinephrine in such processes increases if the starting region is *a-priori* characterized by good compatibility.

As explained before, we built over 15,000 patches on Ste2. For all of them, we computed both Zernike shape descriptors and the hydrophobicity value and we thus compare them with EBS. We report in Fig. 2.a–b the distribution of the distances between EBS and all the sampled patches on the Ste2 surface, for shape and hydrophobicity respectively. In addition, in the top central panels, we show a molecular representation of these distances where a high color intensity indicates a low distance.

To consider both aspects with the same weight, we normalized separately these distances with the Z-score. Therefore, each patch is labeled with 2 normalized values reflecting its similarity with EBS. Now it is possible to define a *dissimilarity score*, simply summing the shape and hydrophobicity Z-score. Where the dissimilarity score is low the patch has good compatibility for both descriptors. Lastly, we performed a smoothing procedure averaging the dissimilarity score of all the neighboring patches to avoid possible singularities. Hence, we selected the patch characterized by the dissimilarity score minimum as the region to work with. The green dotted lines in Fig. 2.a–b represent the shape and hydrophobicity distances between the selected patch and EBS: as evident, this procedure selects a patch characterized by a low distance in both the factors.

Identified the selected patch, we performed 10 independent MC simulations. In each step, a residue is mutated, where shape and

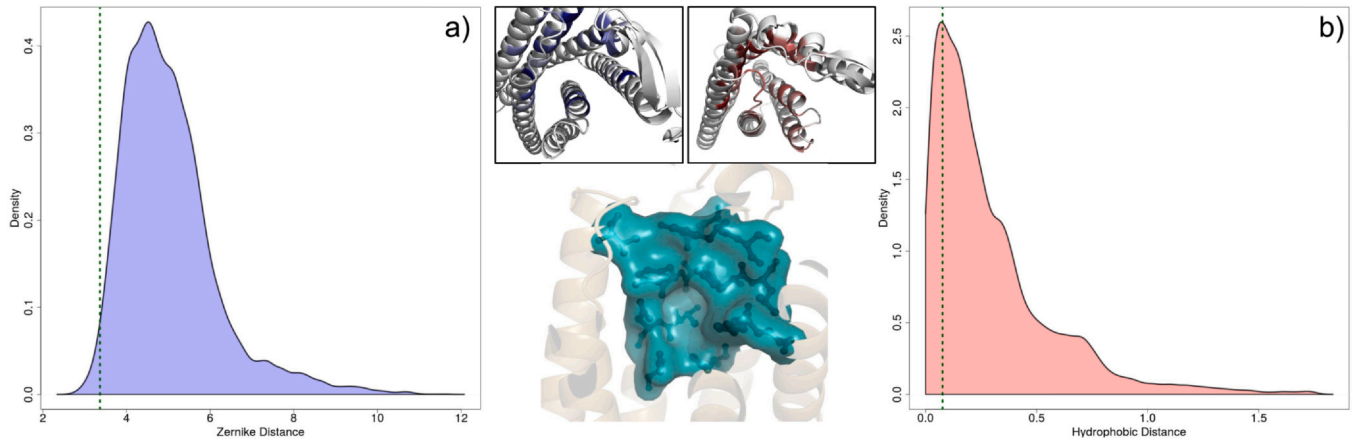

Fig. 2. Selection of the Ste2 pocket to optimize. a) Density distribution of the Zernike distances between EBS and all the sampled patches on Ste2 surface. **b)** Same as in panel a) but for the hydrophobic distance. In the central panels, we report the molecular representation of Ste2 colored according to the shape or hydrophobicity dissimilarity, in the left and right boxes respectively. The bottom central figure represents the selected pocket on Ste2 surface.

hydrophobicity descriptors of the resulting patch are calculated. We defined the *shape similarity* S_s and the *hydrophobicity similarity* S_H as the difference between the patch and the EBS corresponding descriptors. In each step of MC the cost function is:

$$\Delta H = a \cdot \Delta S_s + b \cdot \Delta S_H + c \cdot (M_{new} - M_{old}) \cdot M_{new} \quad (1)$$

where a, b and c are constant parameters. M represents the number of mutations with respect to Ste2 WT. The purpose of such a term is to keep M low to obtain an engineered mutant not too different from the WT. The values of the constants a–c were selected to ensure that the magnitudes of these three terms are similar. This way, all three terms contribute roughly equally to the decision of whether to accept or reject the mutation under consideration.

The beta factor of the simulations is initially kept low and then progressively increased, to let the system exit from its local minimum before freezing it in another configuration. It is worth noting that the two physicochemical terms in the cost function can represent two main aspects: morphology and chemistry. The addition in the cost function of other terms (such as polarity, solvation free energy, residue size, …) could result in the addition of redundant information.

In Fig. 3.a we show the trends that each of these three terms exhibits in a MC simulation. In the left panel, we show the distance between EBS and the optimized patch, as a function of the MC steps. As evident, the shape similarity reached after the optimization is significantly increased with respect to the original patch. In the central panel, we report the difference in average hydrophobicity between EBS and the optimized patch. Lastly, in the right panel, we report the number of mutations as a function of the number of steps. We defined threshold values for all these terms: where the proposed structure is characterized by a value better than the threshold we marked it with an orange point. Where the three threshold conditions are jointly satisfied, we accepted the corresponding mutant as an optimized candidate structure. In Fig. 3.b a joint representation is offered with a 3D plot: the two points in red represent the structure explored in one MC simulation with a low shape and hydrophobicity dissimilarity with EBS, obtained with a low number of mutations with respect to Ste2 original sequence.

Working with 10 independent MC simulations, in this way we obtained 39 optimized candidate structures that satisfy the conditions. The color maps reported in Fig. 3.c express the phase space exploration that occurred during all the 10 MC simulations. Indeed, lighter colors in the pixel are more frequently is to find during MC a structure with these properties.

2.3. Chemico-physical evaluation of the proposed mutants

As discussed before, during the MC simulations we identified 39 sets of mutations that would modify the Ste2 structure to reproduce EBS, with a little number of residues involved to not shatter the overall fold. In particular, our hypothesis is that these mutants should be able to properly bind epinephrine.

Analyzing all these mutants we produced the results shown in Fig. 4. In panel a, we show the frequency of mutations for each of the 15 residues involved in the pocket formation. In addition, each bar of the barplot is subdivided according to the inserted residues. It is interesting to note the presence of mutation hotspots, where 3 positions (HIS 94, GLN 135, THR 192) undergo a residue substitution in over 80% of the proposed mutants. The three WT residues are polar or charged, usually substituted with aromatic or non-polar lateral chains. Conversely, to increase shape and hydrophobicity compatibility just rarely is necessary to mutate 6 out of 15 positions (TYR 106, SER 104, GLN 51, TYR 128, TYR 98, SER 107 exhibit a frequency of mutations lower than 10 %), originally populated by aromatic or polar side chains.

To verify if actually the mutations we proposed make the patch on Ste2 more similar to EBS, we performed a chemical-physical analysis involving some of the most used indices representing amino-acids properties. Hence, we calculated such properties for EBS-forming residues and for both original and mutated Ste2 patch residues. In this way, we compare the EBS residues' characteristics with both mutated and original Ste2 patches (Fig. 4.b). We calculated hydrophobicity [39], net charge, size [42], polarity [43], solvation free energy [44]. For each of these properties, the red dot represents the difference between the original Ste2 pocket and EBS, while the green boxplot reports the distributions for differences between EBS and the 39 mutants identified. Importantly, green boxplots are closer to 0 than red dots, meaning that the mutated patches are characterized by residues more similar to EBS in almost all these properties. Only the size descriptor does not show a clear signal: this is probably because the protocol prefers smaller residues since their reduced size makes them suitable for a variety of substitution combinations.

2.4. Molecular dynamics validation of the optimization protocol

To verify the efficacy of the optimization protocol we developed, we performed independent computational validation of our results through molecular docking and molecular dynamics. In other

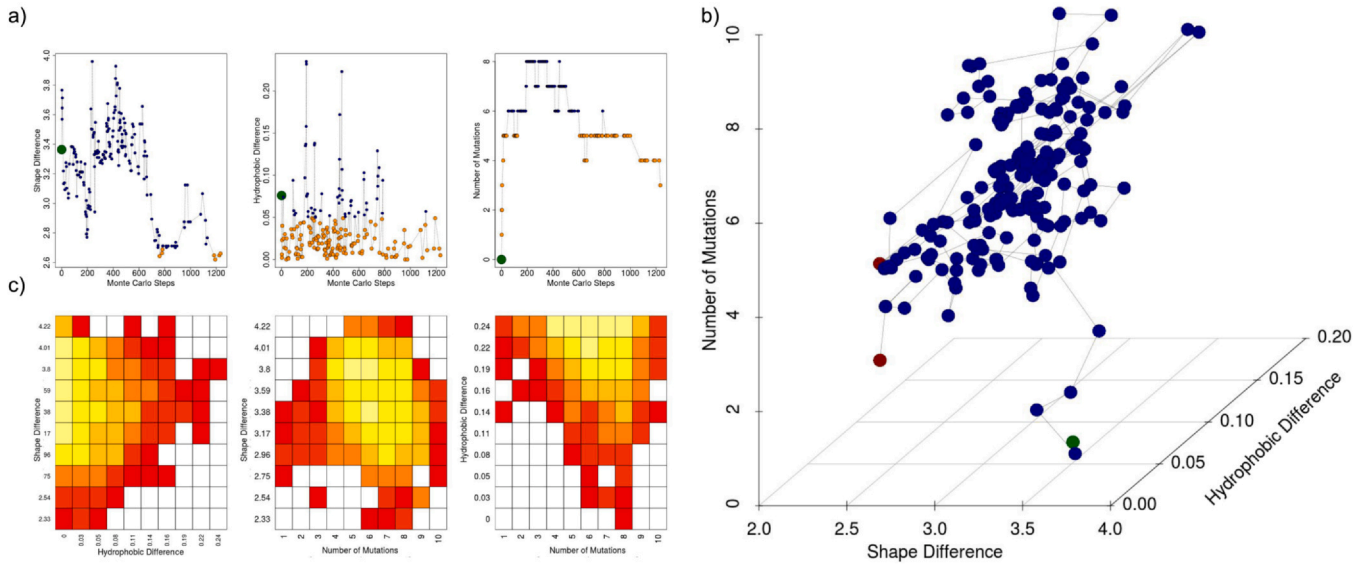

Fig. 3. Results of the Monte Carlo procedure. a) Differences in shape, hydrophobicity, and the number of mutations as a function of the Monte Carlo steps for a representative run of the optimization procedure. The orange points represent the mutants satisfying the acceptance condition for the examined descriptor. **b)** Same as in a) but here points are represented in the 3D space. The red points represent the mutants satisfying at the same time the three conditions. **c)** Color maps representing the probability of visiting a certain region of the phase space that occurred during the 10 performed MC simulations. Colors range from red to yellow as the occupancy probability increases. White regions correspond to unvisited regions.

words, we would like to demonstrate if the designed Ste2 has acquired the capability to recognize and form a stable complex with epinephrine.

Therefore, we first docked epinephrine on both the WT and the 39 mutants of Ste2 we identified using AutoDock [45,46]. We, therefore, obtained 125 independent docking poses of WT Ste2-epinephrine complex. Considering all the optimized Ste2 structures, we collected 4832 independent docking poses of the complex with epinephrine (See Materials and Methods section for details).

Autodock associates every output pose with a predicted binding affinity: for both wt and engineered Ste2 we selected the GPCR-epinephrine pose labeled with the best binding affinity. In Fig. 5.a we reported a molecular representation of such poses: in the top panel the one regarding the WT Ste2 (predicted binding affinity − 5.7 kcal/mol) while in the bottom panel the one involving the engineered version of Ste2 (predicted binding affinity − 6.4 kcal/mol).

To further verify the acquisition of GPCR-epinephrine stability due to residue substitutions, we used the 2 best docking poses as a starting point for molecular dynamics simulations. Indeed, using Gromacs [47] we performed a 200-ns long molecular dynamics of both systems in three replicas. Remarkably, the two systems show very different behaviors.

Indeed, in Fig. 5.b we show the RMSD regarding the epinephrine atoms, after superimposing using as a template the backbone atoms of the protein. In the y-axis is reported the logarithm of RMSD to allow easy comprehension of the general trend, while in the inset the boxplots highlight the distributions of RMSD values. As evident, the red lines, representing the epinephrine in complex with WT-Ste2, underline a predictable binding instability. Very differently, when we analyzed the complex with the optimized Ste2 (blue lines) all three replicas exhibit low values of RMSD, proving the stability of the epinephrine-engineered pocket interaction.

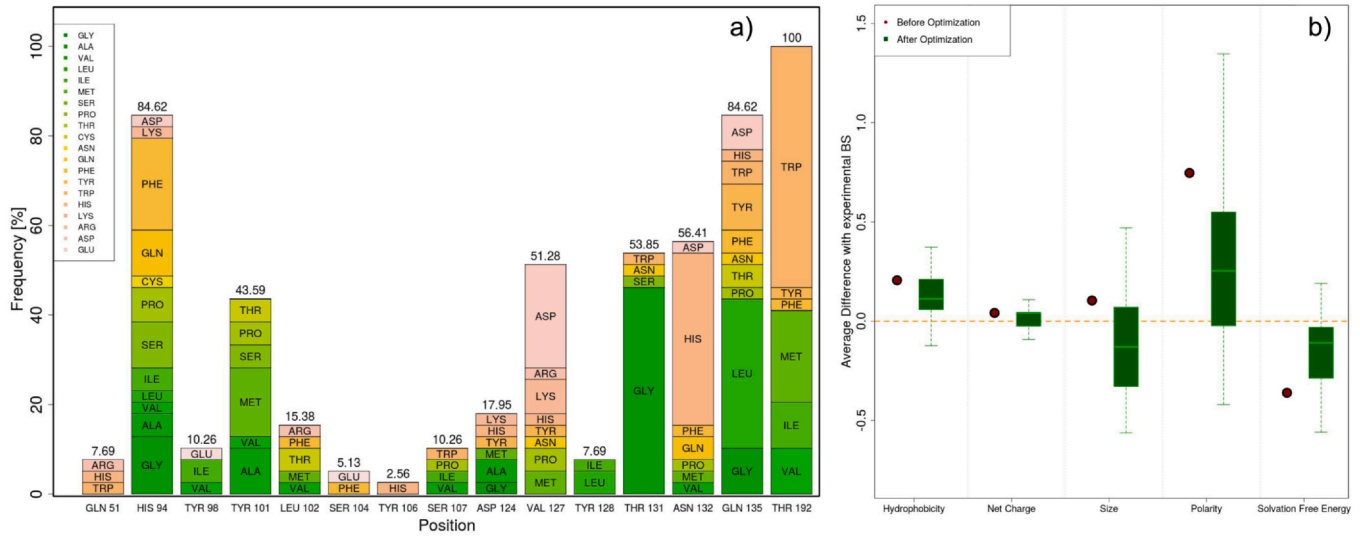

Fig. 4. Analysis of the observed mutations. a) Frequency of mutations for each of the 15 residues involved in the pocket formation of the Ste2 protein during the MC simulation. Each bar of the barplot is subdivided according to the inserted residues. **b)** Boxplot representation of the distributions of the differences between the hydrophobicity, net charge, size, polarity, and solvation free energy of the WT with respect to the mutated versions of the Ste2 protein.

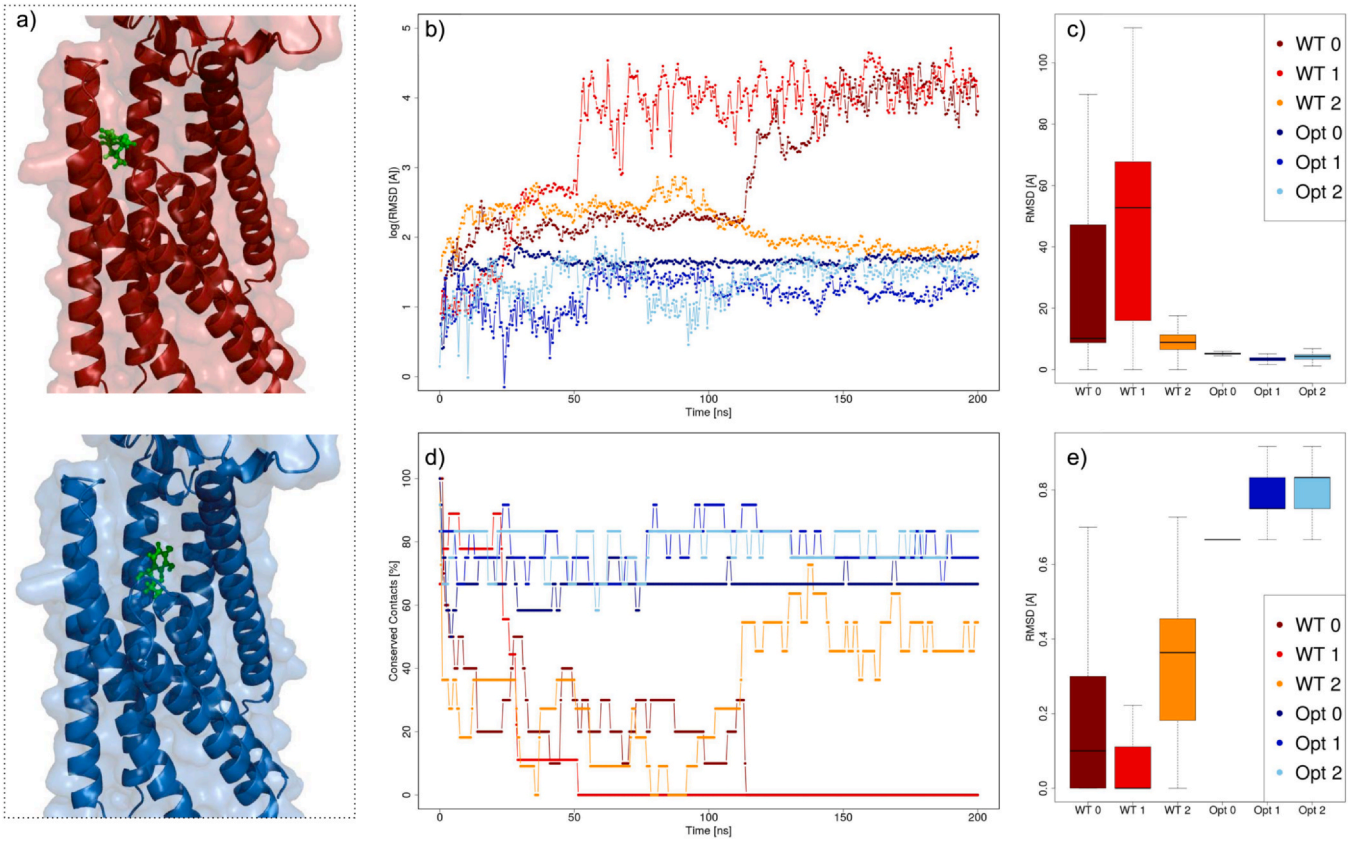

Fig. 5. Molecular dynamics-based validation. a) Molecular representation of the docking poses of Ste2-epinephrine complexes obtained with AutoDock of the WT Ste2 (top) and the engineered version of Ste2 (bottom). **b)** Root Mean Square Deviation (RMSD) as a function of the molecular dynamics simulation time of the epinephrine atoms for three replicas of the WT and optimized systems. **c)** The boxplots highlight the distributions of RMSD values. **d)** Percentage of conserved contacts of the epinephrine atoms with the Ste2 protein as a function of the molecular dynamics simulation time for three replicas of the WT and optimized systems. **e)**The boxplots highlight the distributions of percentages of conserved contacts.

A similar conclusion can be drawn by looking at the percentage of molecular contacts conserved along the simulations (Fig. 5.c). Indeed, a residue is defined as in contact if any of its atoms are closer than 4 Å to any epinephrine atom. Defined the list of residues in contact in the starting structure, we studied the percentage of them that remain in contact along the dynamics. In the WT Ste2 simulations (red curves), the vast majority of initial contacts are lost during the simulations, since the ligand moves away from its pocket. Conversely, the amino acid substitution we inserted made the pocket more suitable for epinephrine recognition, since the blue curves highlight a high percentage of conserved contacts.

3. Conclusions

The ideation of biosensors is very often based on biological systems, whose evolution has developed rapid and effective signaling mechanisms. Because of its stability and the easiness of genetic handling, one of the most used organisms is the haploid version of S.Cerevisiae. This system employs just one GPCR pathway controlled by one GPCR, Ste2, whose pheromone recognition causes the expression of signaling-dependent genes.

Here, we described a computational protocol for the engineering of WT Ste2 to recognize a user-defined compound. The effectiveness of this method paired with secondary genetics modifications furnishing experimental evidence of Ste2 activation could, at least in principle, open the way to the development of a sensor for a very wide range of molecules. Indeed, when a molecule has been previously seen in interaction with a protein, we perform a computational mutagenesis Monte Carlo simulation to identify the set of amino acid substitutions that modify a local region of Ste2 to maximize the similarity with the experimental region that already proved to recognize the examined molecule. Our formalism works only with local protein shape and hydrophobicity descriptors.

As a test case, we modified Ste2 to make it able to recognize epinephrine. It is worth noting that the choice of the test small-molecule is very challenging for the computational protocol we discussed here. Epinephrine is a very different molecule with respect to pheromone, testing the range of applicability of our method. Interestingly, the physicochemical properties of the pockets we designed are more similar to the experimental target than the original WT Ste2 region. More remarkably, molecular docking and molecular dynamics independent testing highlight that the indicated residue substitutions confer high stability to the Ste2-epinephrine complex.

However, it has to be noted that here we focused on the GPCR-ligand binding stability, while additional efforts have to be eventually devoted to ensuring that the examined ligand can activate Ste2, promoting its conformational switch [48,49]. Indeed, our results seem promising since a stable binding could be the first and necessary step to build a reliable biosensor. But binding alone doesn't ensure Ste2 activation. For this reason, further research is needed to properly establish new computational protocols to understand if the binding with the selected ligand can activate the GPCR. Eventually, it is necessary to understand if it is possible to identify other residue substitutions that, while maintaining a stable ligand binding, can promote the GPCR activation upon ligand binding. On the other hand, it is worth noting that the procedure in its current version can design binding sites for antagonist or inverse agonist ligands, for whose activation it is not necessary. This could

be of interest if one wants to modulate the activity of physiological ligands.

In conclusion, we think that our findings can pave the way to a very general procedure to reprogram S. cerevisiae to bind with a very wide range of molecules.

4. Materials and methods

4.1. Molecular surface

All the molecular surfaces in this work have been calculated using the DMS software with standard parameters [50].

When the Ste2 surface was sampled, all the patches were constructed centering it on a surface point and selecting the patch points closer than 7.5 Å to the patch center. We defined 5 patches per Å^2, obtaining 15,114 patches. It has to be noted that we sampled the surface generated by the residues experimentally seen in interaction with the pheromone.

4.2. Zernike descriptors

The surface points of a patch were projected onto a plane, using a conical symmetry to maintain the morphology information [28]. Hence, a patch is described as a 2D function $f(r, \phi)$. It can be expanded on the basis of Zernike polynomials as follows:

$$f(r, \phi) = \sum_{n=0}^{\infty} \sum_{m=0}^{m=n} c_{nm} Z_{nm}, \tag{2}$$

where the expansion coefficients, called the Zernike moments, are:

$$\begin{aligned} c_{nm} &= \frac{(n+1)}{\pi} \langle Z_{nm} | f \rangle \\ &= \frac{(n+1)}{\pi} \int_0^1 dr r \int_0^{2\pi} d\phi Z^*_{nm}(r, \phi) f(r, \phi) \end{aligned} \tag{3}$$

The Zernike polynomials $Z_{nm}(r, \phi)$ are composed of a radial and an angular part:

$$Z_{nm} = R_{nm}(r) e^{im\phi}. \tag{4}$$

The radial function, given n and m, can be written as:

$$R_{nm}(r) = \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k! \left(\frac{n+m}{2} - k\right)! \left(\frac{n-m}{2} - k\right)!} r^{n-2k} \tag{5}$$

Since the set of polynomials form an orthonormal basis, it results that, for each couple of polynomials, the following relation holds:

$$\langle Z_{nm} | Z_{n'm'} \rangle = \frac{\pi}{(n+1)} \delta_{nn'} \delta_{mm'} \tag{6}$$

The knowledge of all the coefficients $\{c_{nm}\}$ permits the complete reconstructions of the original function. In this formalism, the detail level of the description can be set by modulating the order of expansion, $N = \max(n)$.

The modulus of the coefficient ($z_{nm} = |c_{nm}|$) is independent of the original phase since its value does not change if a rotation around the origin is performed. The z_{nm} constitutes the Zernike invariant descriptors.

The shape similarity between two patches is evaluated by comparing their corresponding Zernike invariants. Here, we quantified the similarity between two patches as the Euclidean distance between their invariant vectors. We used the order of expansion N = 20, therefore dealing with 121 invariant descriptors for each patch.

4.3. Hydropathy of patches

Each molecular surface point has been generated from one exposed residue. Each amino acid is characterized by a hydrophobic value [39]. We associate each patch point with the hydrophobic value of the residue generating it. The hydrophobicity of a patch is obtained by averaging all the patch points' hydrophobicity.

4.4. Monte Carlo simulations

Each computational mutagenesis has been performed using the SCWRL4 software [51].

As described in Section 2.2, in each step of the MC simulation we perform a random mutation generating a patch with different shape and hydrophobic properties. We define the *shape similarity* S_s and the *hydrophobicity similarity* S_H as the difference between the generated patch and the EBS corresponding descriptors.

The MC cost function is defined as:

$$\Delta H = a^* \Delta S_s + b^* \Delta S_H + c^* (M_{new} - M_{old}) M_{new} \tag{7}$$

where $a = 1$, $b = 5$ and $c = 0.1$. M_{new} and M_{old} represent the number of mutations with respect to the WT regarding the new and the last accepted mutation, respectively. Each mutation is accepted with the probability:

$$P = \begin{cases} 1 & if\ \Delta H < 0 \\ e^{-\beta \Delta H} & if\ \Delta H \geq 0 \end{cases} \tag{8}$$

where β is the temperature factor. β values start from 4 and go to 20 with a jump of 4: each of 8 β values is maintained for 500 steps for a total of 2500 steps.

4.5. Molecular docking

We docked epinephrine on both the WT and the 39 mutants of Ste2 we identified using AutoDock-vina [45,46].

For each run, we used as a receptor the AlphaFold structure of Ste2 (WT or engineered), and ligand the epinephrine structure extracted from the pdb 4ldo. We restrained the possible docking poses on a 20 × 20 × 20 Å box centered on the centroid of the residues forming the pocket.

For each GPCR-epinephrine docking, we perform 20 experiments each obtained, using 20 randomly generated seeds. With each seed we obtained 20 different poses. We consider two docking poses as independent if the RMSD between their coordinates, after structural protein superposition, is higher than 1.5 Å. After all this procedure, we deal with 125 independent docking poses regarding WT Ste2 and 4832 independent docking poses regarding 39 engineered Ste2.

4.6. Molecular dynamics simulations

The topology files for epinephrine were generated using SwissParam web-server [52]. All the molecular dynamics simulations were performed using Gromacs 2020.6 [47]. Each system was minimized using the steepest descent algorithm. Next, a 0.1 ns-long thermalization of the system was run sequentially in NVT and NPT environments 0.1 ns at 2fs time-step. The temperature was kept constant at 300 K with v-rescale thermostat [53]. Lastly, the final pressure (1 bar) was set using Parrinello-Rahman barostat [54]. For both the WT and engineered Ste2-epinephrine complexes we perform three replicas of 200 ns-long molecular dynamics simulations.

CRediT authorship contribution statement

Lorenzo Di Rienzo: Investigation, Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Mattia Miotto:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Edoardo Milanetti:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Giancarlo Ruocco:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Competing Interests

The authors declare no conflict of interest.

Acknowledgements

The research leading to these results was also supported by European Research Council through its Synergy grant programme, project ASTRA (grant agreement No 855923) and by European Innovation Council through its Pathfinder Open Programme, project ivBM-4PAP (grant agreement No 101098989).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.05.004.

References

[1] Beltrán J, Steiner PJ, Bedewitz M, Wei S, Peterson FC, Li Z, et al. Nat Biotechnol 2022:1–7.
[2] Zimran G, Feuer E, Pri-Tal O, Shpilman M, Mosquna A. ACS Synth Biol 2022.
[3] Yáñez-Sedeño P, Agüí L, Villalonga R, Pingarrón J. Anal Chim Acta 2014;823:1.
[4] Mello LD, Kubota LT. Food Chem 2002;77:237.
[5] Eggins BR. Biosensors: an introduction. Springer-Verlag; 2013.
[6] Taylor ND, Garruss AS, Moretti R, Chan S, Arbing MA, Cascio D, et al. Nat Methods 2016;13:177.
[7] Polizzi NF, DeGrado WF. Science 2020;369:1227.
[8] Bick MJ, Greisen PJ, Morey KJ, Antunes MS, La D, Sankaran B, et al. Elife 2017;6:e28909.
[9] Shui S, Gainza P, Scheller L, Yang C, Kurumida Y, Rosset S, et al. Nat Commun 2021;12:1.
[10] d'Oelsnitz S, Nguyen V, Alper HS, Ellington AD. ACS Synth Biol 2022;11:265.
[11] Rottinghaus AG, Xi C, Amrofell MB, Yi H, Moon TS. Cell Syst 2022;13:204.
[12] Urban DJ, Roth BL. Annu Rev Pharm Toxicol 2015;55:399.
[13] Bjarnadóttir TK, Gloriam DE, Hellstrand SH, Kristiansson H, Fredriksson R, Schiöth HB. Genomics 2006;88:263.
[14] Armbruster BN, Roth BL. J Biol Chem 2005;280:5129.
[15] Wettschureck N, Offermanns S. Physiol Rev 2005;85:1159.
[16] Spiegel AM, Weinstein LS. Annu Rev Med 2004;55:27.
[17] Conklin BR, Hsiao EC, Claeysen S, Dumuis A, Srinivasan S, Forsayeth JR, et al. Nat Methods 2008;5:673.
[18] Rowe JB, Taghon GJ, Kapolka NJ, Morgan WM, Isom DG. J Biol Chem 2020;295:8262.
[19] Besada-Lombana PB, McTaggart TL, Da Silva NA. Curr Opin Biotechnol 2018;53:39.
[20] Bardwell L. Peptides 2005;26:339.
[21] Wang Y, Dohlman HG. Science 2004;306:1508.
[22] Dowell SJ, Brown AJ. Recept Channels 2002;8:343.
[23] Kapolka N, Taghon G, Rowe J, Morgan W, Enten J, Lambert N, et al. Proc Natl Acad Sci 2020;117:13117.
[24] Dowell SJ, Brown AJ. G protein-coupled receptors in drug discovery. Springer; 2009. p. 213–29.
[25] Velazhahan V, Ma N, Vaidehi N, Tate CG. Nature 2022;603:743.
[26] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. Nucleic Acids Res 2000;28:235.
[27] Ring AM, Manglik A, Kruse AC, Enos MD, Weis WI, Garcia KC, et al. Nature 2013;502:575.
[28] Milanetti E, Miotto M, Di Rienzo L, Monti M, Gosti G, Ruocco G. Comput Struct Biotechnol J 2021;19:29.
[29] Venkatraman V, Yang YD, Sael L, Kihara D. BMC Bioinform 2009;10:407.
[30] Daberdaku S, Ferrari C. BMC Bioinform 2018;19:35.
[31] Sandomenico A, Di Rienzo L, Calvanese L, Iaccarino E, D'Auria G, Falcigno L, et al. Biomedicines 2020;9:20.
[32] Di Rienzo L, De Flaviis L, Ruocco G, Folli V, Milanetti E. J Comput-Aided Mol Des 2022;36:11.
[33] Daberdaku S, Ferrari C. Bioinformatics 2019;35:1870.
[34] Sael L, Li B, La D, Fang Y, Ramani K, Rustamov R, et al. Protein: Struct Funct Bioinform 2008;72:1259.
[35] Di Rienzo L, Milanetti E, Ruocco G, Lepore R. Front Mol Biosci 2021;8:749784.
[36] Di Rienzo L, Milanetti E, Testi C, Montemiglio LC, Baiocco P, Boffi A, et al. Comput Struct Biotechnol J 2020;18:2678.
[37] Di Rienzo L, Monti M, Milanetti E, Miotto M, Boffi A, Tartaglia GG, et al. Comput Struct Biotechnol J 2021;19:3006.
[38] De Lauro A, Di Rienzo L, Miotto M, Olimpieri PP, Milanetti E, Ruocco G. Front Mol Biosci 2022;9.
[39] Di Rienzo L, Miotto M, Bò L, Ruocco G, Raimondo D, Milanetti E. Front Mol Biosci 2021;8:626837.
[40] Velazhahan V, Ma N, Pándy-Szekeres G, Kooistra AJ, Lee Y, Gloriam DE, et al. Nature 2021;589:148.
[41] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Nature 2021;596:583.
[42] Brock DJ, Mayo O. Biochemical genetics of man. New York: Academic Press; 1972.
[43] Grantham R. Science 1974;185:862.
[44] Eisenberg D, McLachlan AD. Nature 1986;319:199.
[45] Trott O, Olson AJ. J Comput Chem 2010;31:455.
[46] Eberhardt J, Santos-Martins D, Tillack AF, Forli S. J Chem Inf Model 2021;61:3891.
[47] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. SoftwareX 2015;1:19.
[48] Fleetwood O, Carlsson J, Delemotte L. Elife 2021;10:e60715.
[49] Alberstein RG, Guo AB, Kortemme T. Curr Opin Struct Biol 2022;72:71.
[50] Richards FM. Annu Rev Biophys Bioeng 1977;6:151.
[51] Krivov GG, Shapovalov MV, Dunbrack Jr RL. Protein: Struct Funct Bioinform 2009;77:778.
[52] Zoete V, Cuendet MA, Grosdidier A, Michielin O. J Comput Chem 2011;32:2359.
[53] Bussi G, Donadio D, Parrinello M. J Chem Phys 2007;126:014101.
[54] Parrinello M, Rahman A. Phys Rev Lett 1980;45:1196.