Statistical Potential to Improve Antibody-Antigen Docking

Alexander Sadovnikov*†, Timofei Ermak*, Pavel Yakovlev*
*BIOCAD, Saint-Petersburg, Russian Federation
†ITMO University, Saint-Petersburg, Russian Federation
{sadovnikovav, ermak, yakovlev}@biocad.ru

***Abstract*—In this paper, we describe the development of a novel statistical potential for the prediction of antibody-antigen complexes (docking), which play key role in *in silico* immunotherapy discovery. The developed statistical potential is then used to improve the accuracy of an existing docking algorithm. We also present a new dataset for the development and comparison of different statistical potentials and docking algorithms. One of the key features of the developed dataset is that it can be obtained almost automatically, with few optional manual steps, using the pipeline introduced in this paper.**

I. Introduction

A. Docking

Protein-protein interactions play fundamental role in living systems functioning. Immune system reactions, cell signaling, and many other intermolecular processes are based on protein-protein complexes interactions. Describing such complexes from the biophysical and structural points of view plays a crucial role in understanding cell's functions. This makes the problem of protein-protein complex prediction, also called *docking*, one of the most important tasks being solved during rational drug design — the process of drug development, which heavily relies on studying the structure and functions of molecules.

Up to this day the problem of docking remains one of the most difficult tasks in structural bioinformatics, which in the majority of cases cannot be solved efficiently neither by human nor by the existing algorithms [1]. One of the main reasons for this is the fact that obtaining accurate results for such task demands an exhaustive search in large conformational space.

During the process of drug development, the docking problem can occur up to several hundred times. Therefore, solving this task in short time ranges is very important. On the other hand, algorithmic optimizations of the search often rely on narrowing down the space of search, which may result in the eventual loss of accuracy. Hence, the problem of docking should be solved both efficiently and accurately.

B. Grid-based docking

First efficient algorithm for docking of two molecules was introduced in 1992 by Katchalski-Katzir [2].

Optimal rotation and translation of one molecule relative to the other molecule are searched. Step of searching optimal translation is calculated using grids' correlation, what improves step's efficiency compared to the naive approach.

First of the two molecules that form the complex, which in terms of the docking problem is called *receptor*, and the second molecule, which is called *ligand*, are being placed into 3-dimensional grids. The value of each cell of the grids is being calculated according to the following formula:

$$\text{cell}_{ijk} = \begin{cases} 1, & \text{if on surface} \\ p, & \text{if inside} \\ 0, & \text{otherwise.} \end{cases}$$

Here, the value of 1 is assigned to the grid cells that contain surface-atoms of the molecule. Value of p, which lays in the range $(0, 1)$, is assigned to cells that only contain internal atoms of the molecule. p can be different for receptor and ligand. The value of 0 is assigned to all other cells of the grid.

Once we have obtained grids for the receptor and the ligand, we calculate the correlations of the two grids for all possible translations in 3-dimensional euclidean space of ligand's grid respective to receptor's grid using the following formula:

$$\sum_{i=1}^{N}\sum_{j=1}^{M} A_{ij} \cdot B_{ij},$$

where A and B are the receptor's and the ligand's grids, and N and M are sizes of grids' dimensions. Conformation with maximal correlation value corresponds to the mutual position of the receptor and the ligand that has maximal geometric complementarity.

Therefore, the problem of finding optimal translation of one molecule relative to the other molecule comes down to finding the conformation with maximal grids' correlation.

Expressing molecules' shape complementarity in terms of the correlation of two grids allowed for the use of FFT-based approach to the exhaustive search of ligand's grid translation that maximizes the correlation of two grids. The whole algorithm significantly reduced theoretical and practical time it takes to dock two molecules [2]. The process of docking two molecules using correlation grids is now referred to as *grid-based docking*.

Overall, shape complementary method accounts only for geometry of the molecules and does not take into account their physical properties. Therefore, the method has very limited applications and low accuracy in general.

C. Statistical potentials and DARS

In 2006, Kozakov and coworkers introduced a novel docking algorithm *Piper*, which later became the core of the web-

server *ClusPro* [3]. Up to this day, ClusPro remains the best fully automated docking server [1].

This result was achieved with the help of adding new terms to the correlation function maximized during the stage of grid-based docking. Firstly, geometry term became an approximation of van der Waals energy, unlike the geometry term in Katchalski-Katzir's work that approximates basic shape complementarity. Secondly, term accounting for electrostatic interaction and a *statistical potential* term were introduced.

The statistical potential term captures the information about how probable it is that the given conformation of the receptor and the ligand is a *native* conformation, i.e. a conformation that can be observed in nature. Estimation of probability is based on statistics retrieved from a particular set of biological complexes that the statistical potential is trained on.

The idea of the application of the statistical potential to the problem of docking was not new for 2006, as several prior works explored the possibilities of construction and application of such potentials [4]. What made the result achieved by Kozakov stand out from its predecessors is the novel way of obtaining the training set for the statistical potential and the developed technique of representation of statistical potential in terms of grids' correlation, which made it possible to incorporate statistical potentials into grid-based docking stage of ClusPro's pipeline.

One of the most commonly used techniques used for the development of the statistical potentials is the inverse Boltzmann approach [5]. Its idea is to express free energy of the molecular system's state via the probability of this state's existence, using Boltzmann's distribution. This approach is very general and the results of its application may vary depending on the definition of the system's state. In the case of the development of the statistical potential for the problem of docking, state is usually defined as a pair of atoms.

Statistical potentials are just 2-dimensional tables. Rows and columns of these tables correspond to types of atoms taken into consideration in a given statistical potential. For example, if there are two atoms of types I and J, then energy of their interaction according to a statistical potential with table SP is equal to the value of the item SP_{IJ} of the table. If there are two molecules forming a complex, then the energy of that complex in a statistical potential with table SP is calculated in the following manner:

$$E_{SP} = \sum_{i=1}^{N} \sum_{j=1}^{M} I_{d(i,j)<D} \cdot SP_{ty(1,i)ty(2,j)},$$

where N and M are the numbers of atoms in the first and the second molecules respectively, $I_{d(i,j)<D}$ is equal to 1, if atoms of molecules with indices i and j are at the distance less then D, and is equal to 0 otherwise, $ty(i,x)$ is the function that maps the atom with index x in the molecule i to its type in the statistical potential SP.

Each item of the statistical potential's table is calculated using the following formula derived from the Boltzmann's distribution:

$$SP_{IJ} = -RT \ln \frac{p_{IJ}^{obs}}{p_{IJ}^{ref}},$$

where R is the gas constant, T is the temperature, p_{IJ}^{obs} is the probability of atoms I and J interacting in a training set and p_{IJ}^{ref} is the probability of atoms I and J interacting in a so called *reference state*, which is the approximation of space of all possible molecular complexes. Most of the differences between statistical potentials come from the training sets that statistical potentials are built on and from the way p_{IJ}^{ref} is being calculated.

ClusPro up to this day uses statistical potential called *DARS* [3], [6], introduced in 2006. The main idea of DARS is to use an artificially generated set of physically wrong conformations, *decoys*, as a reference state. To generate this artificial set, one needs to take a number of valid molecular complexes and dock their components using grid-based docking algorithm that takes into account only the shape complementarity correlation term. Conformations obtained as the result of this action are independent of any atom-specific interactions by construct: there is no bias towards any pair of interacting atom types. So in terms of atom interactions, complexes can be called "random". And that fact makes it reasonable to define reference state for statistical potential as a large set of such decoys.

Despite DARS showing good results at discriminating improbable conformations from probable for some types of complexes, it has a number of drawbacks.

For complexes where *antibody*, a protein produced by the immune system for the elimination of pathogenic molecules, acts as a receptor, DARS has been demonstrated to have poor discriminatory power [6].

The less obvious drawback of the DARS potential is the fact that it does not take into account the distances between interacting atoms. Binary metric for detecting the interaction between two atoms is used: only atoms that are less distant than some cutoff distance D are considered as interacting. Such binary classification may yield imprecise results: two equal pairs of atoms placed at *different* distances both of which are less than D are said to *equally* contribute to the energy of the complex.

As for the solution of this problem, in 2001 a distant-dependent potential was developed for the needs of the *folding* problem [7]. This potential can be combined with DARS's method of obtaining a reference state in order to overcome said problem of accounting for the distances between atoms when calculating the total energy of the complex.

D. Antibody-antigen complexes and aADARS

Antibody-antigen complexes are complexes, in which the antibody acts as a receptor and a ligand is a molecule that binds to antibody. Such ligands are also referred to as *antigens*. Antibodies play a vital role in modern-day drug development, because they proved themselves to be efficient in the treatment of cancer and autoimmune diseases. About one fourth of all drugs approved by FDA during the last five years are monoclonal antibodies [8]. Therefore, the problem of accurate docking for antibody-antigen complexes is particularly relevant today.

In 2012, aADARS statistical potential was introduced [9]. It was designed to improve DARS's performance on antibody-

antigen complexes. Such an improvement was achieved by training statistical potential on a specific dataset containing only antibody-antigen complexes. Another feature of the statistical potential was the fact that its table was asymmetric. This was done in order to discriminate atoms of antibody and antigen, which are presumed to have different properties.

This statistical potential has the same problem as DARS in terms of not taking into consideration the distances between interacting atoms.

But what is more important is that aADARS was developed in 2012 and since then there have been no reports on potential's updates. The same can be said about non-antibody DARS itself.

Protein Data Bank, or just *PDB*, is the world's largest storage of molecules' 3-dimensional structures. PDB acted as a source of data for the development of both DARS and aADARS. As it has been shown in the latest report on the state of PDB, the number of structures available in the storage grows exponentially each year [10]. Therefore, there is clear evidence that both statistical potentials can be made more accurate with the use of new data.

Unfortunately, no tools exist for the automatic creation of said potentials using relevant data. And that poses another challenge for the creation of updated versions of DARS and aADARS.

E. Benchmark for Docking and Statistical Potential

To conduct a benchmark of a newly developed docking algorithm or statistical potential, a particular set of test structures should be collected.

Every entry in the benchmark's dataset should contain:

1) Receptor-ligand complex that captures molecules' native interaction. Such complexes are called *bound*.
2) Receptor molecule as it is found in nature, independent of any interactions with other molecules. Such conformations of molecules are called *unbound*.
3) Ligand's unbound conformation.

During the benchmark of docking, unbound conformations of receptor and ligand are being docked and the result is being compared with the bound complex. This way of testing an algorithm is chosen because it simulates the real-life situation: one has structures of receptor and ligand as they exist in nature and wants to predict their complex as close as possible to the native one.

The same dataset is applicable to the task of benchmarking and comparing statistical potentials. For every complex in the dataset, unbound receptor and unbound ligand are being docked with the help of grid-based docking algorithm that does not use statistical potential correlation term. Algorithm produces a number of conformations of the molecules which are then sorted with respect to the statistical potential energy. After that, a number of conformations, that are close to the bound complex and that made it into top-N poses, is calculated. The more near-native complexes are in the top-N solutions – the better.

Most well-known open-source docking benchmark is ZLab dataset [11]. Latest version of the dataset was gathered from the PDB in 2015. Both DARS and aADARS statistical potentials have been benchmarked using ZLab [3], [9].

If one wants to benchmark an antigen-antibody statistical potential, ZLab has a major drawback: it contains 230 entries in total, only 28 of which are bound antigen-antibody complexes with both of their unbound components present in the dataset. To test the discriminatory power of statistical potential properly, clearly, more data is required.

There is also evidence that such data indeed can be obtained: SAbDab database, which accounts for the structures from the PDB that contain antibodies and which is updated regularly, lists at least 2730 bound antibody-antigen complexes [12]. So there is real hope that unbound parts can be found for a more significant number of complexes than those already contained in ZLab.

F. HEDGE

In 2018, BIOCAD introduced its own docking algorithm, *HEDGE*, the novelty of which lies in its performance capabilities: all stages of the algorithm are computed solely on GPU with ability to use several devices at once [13]. This allows for more exhaustive search of conformational space in less computational time.

HEDGE has not incorporated a statistical potential yet, so its accuracy could be significantly improved by implementing such potential.

G. Aim of this paper

As it has been shown in the subsections I-C and I-D, both DARS and aADARS statistical potentials can be improved. In this work, we aim to develop a novel statistical potential, designed to work solely with antibody-antigen complexes, in order to improve HEDGE's performance for this type of structures. Statistical potential is going to be built with the use of all the data available in the PDB about said type of complexes at this moment.

Also, we aim to create a tool that could, given a training dataset, produce a developed statistical potential built on this set. In the future this tool will help keeping developed statistical potential up to date with all the new data that becomes available in the PDB.

Discriminative power of the newly developed statistical potential is going to be tested and compared to the DARS and aADARS in the manner described in I-E.

In order to achieve this, a new dataset will be gathered since, as it has been described in the subsection I-E, ZLab dataset lacks antibody-antigen complexes. There is strong evidence that more data for the task of benchmarking antibody statistical potential can be obtained from the PDB.

The benchmarking dataset will be complemented with a tool for its automatic updates.

H. Paper's structure

The paper is organized as follows. The next section presents a detailed research plan of what is needed to be done in order to achieve the goals set in the subsection I-G. The third section presents results that have been obtained up to now. The final section draws up conclusions that can be made from this study so far, as well as presents short summary of the work that will have been done before the conference.

II. Research plan

A. Plan

The plan for achieving goals set in section I-G looks as follows:

1) Confirm the hypothesis that the usage of statistical potentials will improve HEDGE's performance.
2) Collect the dataset of antibody-antigen complexes to train the new statistical potential. Collect the benchmarking dataset to compare the performances of different statistical potentials.
3) Develop DARS-like statistical potential built on a new training set and compare its performance to original DARS and aADARS.
4) Add the accounting for distances to the new potential, similar to how it was done in in 2001 distant-dependent potential.
5) Compare both versions of developed potential with DARS and aADARS.
6) Create a tool that, given a training set, builds two versions of the developed statistical potential using provided training set.

Following subsections present a detailed description of every item of the plan.

B. Confirming statistical potentials' applicability

As the main motivation for creating a novel statistical potential is the improvement of HEDGE's accuracy, it was important to confirm the hypothesis that adding statistical potential correlation term into the HEDGE's grid-based docking correlation function would significantly improve algorithm's performance.

In order to do that, a statistical potential had to be implemented in HEDGE. We chose the original DARS for this task, as it can be used for different types of complexes. Next step was the creation of a grid-based docking stage benchmark for HEDGE. Final step was the comparison of algorithm's performance with and without DARS using chosen benchmark dataset.

C. Collecting training and validation sets

As the DARS and aADARS were built in 2006 and 2012 respectively, there exists a handful of new antibody-antigen complexes that can be included in the training set to improve the performance of developed statistical potential.

On the other hand, the existing dataset for benchmarking statistical potentials, ZLab, contains too few entries of the antibody-antigen type. So for that reason, again, new data should be gathered from the PDB.

The big difference between the validation and the training sets is that for the validation set we need not only antibody-antigen complexes themselves, but also the unbound versions of the receptor and the ligand. Therefore, the pipeline for gathering data for the validation set has to be more complex: the process of finding unbound conformations of molecules in PDB is a difficult task due to the diversity of data stored in the database.

D. Developing DARS-like potential

Next step is the creation of a statistical potential that, similarly to DARS, uses a decoy set of complexes as a reference state. The only difference of this version of the developed statistical potential from the original DARS would be the training set: our potential will be built on a set of antigen-antibody complexes, unlike DARS that was built on a heterogeneous training set.

For the comparison of the developed statistical potential with DARS and aADARS, the technique described in I-E will be employed. All statistical potentials will be benchmarked on the data collected during the previous step.

This comparison will show the viability of the idea that DARS's method of obtaining a reference state can be applied as is to the problem of developing an antibody-antigen statistical potential.

E. Accounting for distances in the statistical potential

The following step of developing a statistical potential is to combine the DARS's way of creating reference state and distant-dependent potential's method of accounting for distances between interacting atoms. In theory, this should significantly increase statistical potential's accuracy, but such combination of two ideas has never been properly researched.

F. Detailed comparison with DARS and aADARS

Both versions of the developed potential will be compared to DARS and aADARS. The resume will be given for the advantages and disadvantages of all four statistical potentials as their performance on the benchmark dataset will be thoroughly examined. For the task of statistical potentials' comparison the same method will be used as in II-D.

G. Tool for creating statistical potentials

The final step will be developing a fully automated tool for creating DARS-like statistical potentials with option of accounting for the distances between atoms. The idea is that this tool will be able to create potentials not only for antibody-antigen type of complexes, but for any given training dataset.

III. Results

A. Confirming statistical potentials' applicability

We've confirmed the hypothesis that the use of statistical potentials in the correlation function of HEDGE's grid-based

docking stage will improve its accuracy. In order to confirm this, DARS was employed.

The performances of HEDGE's grid-based docking with and without the use of statistical potential have been compared in the following manner.

For every complex in the benchmark, 1,600,000 conformations are obtained using HEDGE's grid-based docking algorithm. These conformations are then sorted by their correlation value. After that, every pose in Top-N conformations obtained for each complex, where N varies from 1,000 to 10,000 with step 1,000 and from 10,000 to 50,000 with step 10,000, is being compared to the native-complex using RMSD, a standard metric for the task of comparing molecules' structures. RMSD is basically a root-mean-squared distance between the corresponding atoms of two structures. RMSD for our case is calculated between the *interaction interface*, i.e. the atoms of the receptor and the ligand that are closer than 10 Å to the other molecule, of the native-complex and the interaction interface of the predicted conformation. For every Top-N, we detect whether there exist a conformation distant from the native complex by no more than K Å, where K is chosen to be 1, 2, 5, and 10 Å. In the end, for each pair of (N, K), we calculated the number of complexes in the benchmark, for which there was found at least one conformation in the Top-N results, the interaction interface of which has RMSD of no more than K Å to the interaction interface of the native complex.

As a benchmark dataset, we chose ZLab's rigid category that contains 151 complexes of different types. Since we wanted to confirm the hypothesis of statistical potential's applicability to the improvement of the accuracy of the algorithm in general, there was no need in focusing only on complexes of the antibody-antigen type. After filtering complexes for problems with their 3-dimensional structures, 128 complexes comprised the final benchmark.

Fig. 1 illustrates the improvement of accuracy that is achieved by using DARS statistical potential during the grid-based docking stage of HEDGE's pipeline. The value of every cell in the table is equal to the difference in number of complexes, for which conformations satisfying (Top-N, RMSD) constraints have been found for versions of algorithm with and without DARS.

We are mostly interested in the row corresponding to 2 Å, since structures so close to the native complex can be considered near-native. In this row there are cases where 30-42 more near-native complexes have been found with the help of DARS compared to the version of algorithm without the statistical potential. Other cells in the table are also positive, so the improvement of accuracy from the use of statistical potential is clear.

B. Collecting training and validation sets

So far, only the validation set has been collected. Collecting validation data is a much more difficult task compared to collecting a training set due to the process of finding unbound conformations of structures in PDB.

It is unclear how to automatically recognize that a structure is unbound. This process has to be automatic, because the manual search for such structures in PDB is very time-consuming.

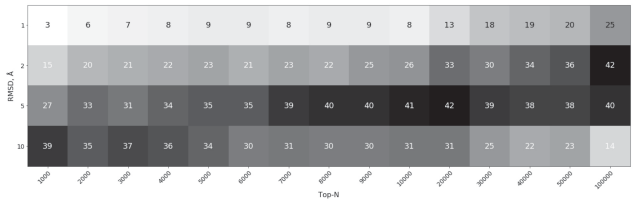

Fig. 1. HEDGE's grid-based docking stage's accuracy improvement from the use of statistical potentia

We developed a semi-automated pipeline for gathering such validation set. It's not fully automated, because sometimes there are cases that cannot be solved definitively by a computer. And for the resolution of such cases, human assistance is required. Despite this fact, the pipeline can be run fully automatically and it will still gather a proper validation set. Human assistance is needed in order to make this set larger by approving some complexes that the algorithm declined to include in the set automatically.

As a starting point for our pipeline, we took SAbDab database that contains 2730 antigen-antibody complexes [12]. Every complex in the database has a link to its structure in PDB.

The pipeline for finding unbound parts for complexes from SAbDab database is presented in Fig. 2.

For every complex from SAbDab, our pipeline can give one of three results:

1) Unbound conformations for both the ligand and the receptor. This result corresponds to the terminal state "Return all pairs of non-"doubtful" antibody-antigen structures" in Fig. 2.
2) Conformations of the ligand and the receptor that are presumed to be unbound, but require human's assessment. This result corresponds to the terminal state "Need human assistance" in Fig. 2.
3) Definitive answer that unbound conformations cannot be found for both the ligand and the receptor. This result corresponds to the terminal state "No unbound receptor and ligand for complex" in Fig. 2.

The pipeline consists of three main stages. Stage A is responsible for finding potential candidates for unbound structures for sequences of antibody and antigen that form the complex. Candidates are searched in PDB using sequence-based searching algorithm BLAST [14].

Sometimes the structures found during stage A may contain something other than desired antigen/antibody. Stage B is designed specifically to look into such cases and to confirm whether the found structures are really the unbound parts of the initial SAbDab's complex.

Stage C acts as a final filter for the candidates for unbound conformations. It is designed to check the found structures for the presence of small molecules, which may prevent the corresponding conformations from being unbound.

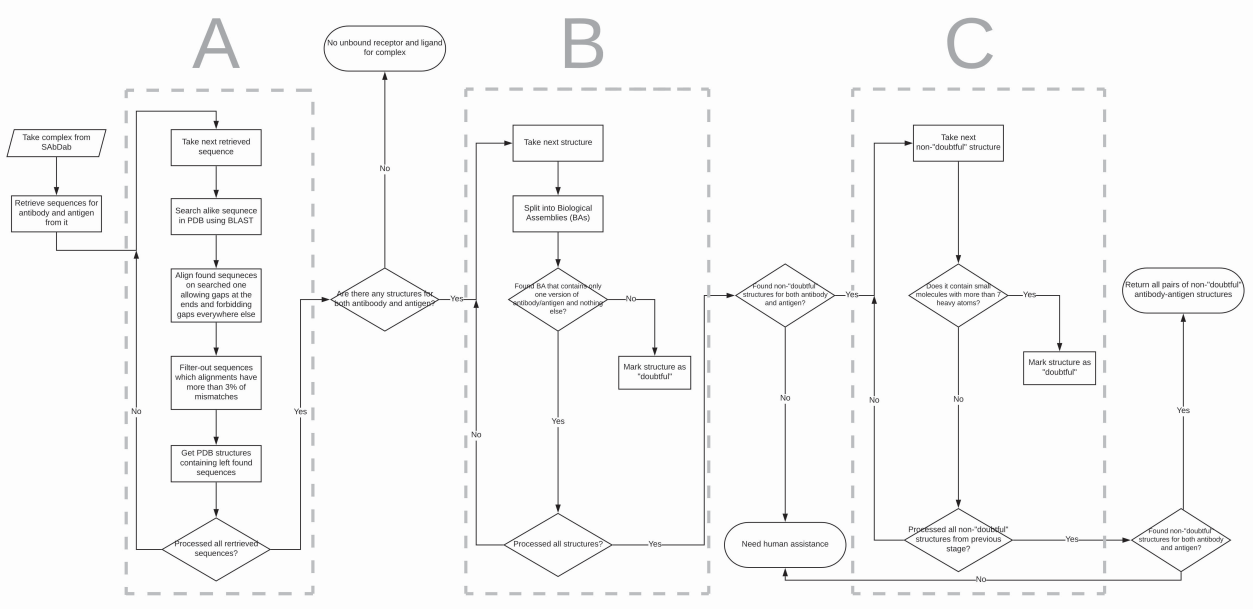

Fig. 2. The pipeline for finding unbound parts for complexes from SAbDab database

The current results of the pipeline's work on SAbDab database are as follows. The gathered validation set of antibody-antigen complexes has 100 complexes, for which the pipeline has given a definitive answer that it found required unbound parts, and 74 complexes more, for which the pipeline marked all the unbound parts as in need of human's assessment.

IV. Conclusion and future work

We have shown that the use of statistical potential significantly improves the performance of HEDGE, a novel GPU-driven docking algorithm developed by BIOCAD. This makes the development of a new statistical potential that favors complexes of type antibody-antigen reasonable.

We have also presented a plan for developing said statistical potential along with the plan for its comparison to existing solutions, such as DARS and aADARS.

A validation set for comparison of statistical potentials has been gathered using a developed pipeline for finding unbound parts of antibody-antigen complexes.

As for the future work, by the time of the conference, we plan to develop a DARS-like statistical potential for antibody-antigen complexes and conduct the comparison of developed statistical potential with DARS and aADARS.

References

[1] M.F. Lensink, S. Velankar, S.J. Wodak, "Modeling protein–protein and protein–peptide complexes: CAPRI 6th edition.", *Proteins: Structure, Function, and Bioinformatics*, 85.3, 2017, pp. 359-377.

[2] Katchalski-Katzir, Shariv, Eisenstein, Friesem, Aflalo and Vakser, "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques.", *Proceedings of the National Academy of Sciences*, 89(6), 1992, pp.2195-2199.

[3] D. Kozakov, R. Brenke, S.R. Comeau and S. Vajda, "PIPER: an FFT-based protein docking program with pairwise potentials.", *Proteins: Structure, Function, and Bioinformatics*, 65(2), 2006, pp.392-406.

[4] C. Zhang, G. Vasmatzis, J.L. Cornette and C. DeLisi, "Determination of atomic desolvation energies from the structures of crystallized proteins.", *Journal of molecular biology*, 267(3), 1997, pp.707-726.

[5] K. Olechnovič and Č. Venclovas, "VoroMQA: Assessment of protein structure quality using interatomic contact areas.", *Proteins: Structure, Function, and Bioinformatics*, 85(6), 2017, pp.1131-1145.

[6] G.Y. Chuang, D. Kozakov, R. Brenke, S.R. Comeau and S. Vajda, "DARS (Decoys As the Reference State) potentials for protein-protein docking.", *Biophysical journal*, 95(9), 2008, pp.4217-4227.

[7] H. Lu, and J. Skolnick, "A distance-dependent atomic knowledge-based potential for improved protein structure selection.", *Proteins: Structure, Function, and Bioinformatics*, 44(3), 2001, pp.223-232.

[8] B. G de la Torre and F. Albericio, "The pharmaceutical industry in 2018. An analysis of FDA drug approvals from the perspective of molecules.", *Molecules*, 24(4), 2019, p.809.

[9] Brenke, Hall, Chuang, Comeau, Bohnuud, Beglov, Schueler-Furman, Vajda and Kozakov, "Application of asymmetric statistical potentials to antibody–protein docking.", *Bioinformatics*, 28(20), 2012, pp.2608-2614.

[10] Goodsell, Zardecki, Di Costanzo, Duarte, Hudson, Persikova, Segura, Shao, Voigt, Westbrook and Young, "RCSB Protein Data Bank: Enabling biomedical research and drug discovery.", *Protein Science*, 29(1), 2020, pp.52-65.

[11] Vreven, Moal, Vangone, Pierce, Kastritis, Torchala, Chaleil, Jiménez-García, Bates, Fernandez-Recio and Bonvin, "Updates to the integrated protein–protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2.", *Journal of molecular biology*, 427(19), 2015, pp.3031-3041.

[12] Raybould, Marks, Lewis, Shi, Bujotzek, Taddese and Deane, "Thera-SAbDab: the Therapeutic Structural Antibody Database.", *Nucleic acids research*, 48(D1), 2020, pp.D383-D388.

[13] T. Ermak, A. Shehovtsov and P. Yakovlev, "HEDGE: Highly accurate GPU-powered protein-protein docking pipeline.", *SI Kabanikhin, Professor, Corresponding Member of the RAS, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS*, 2018, p.22.

[14] Altschul, Madden, Schäffer, Zhang, Zhang, Miller and Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.", *Nucleic acids research*, 25(17), 1997, pp.3389-3402.