2011 15th International Conference on Information Visualisation

Interactive Drug Design in Virtual Reality

Ching-Man Tse, Hongjian Li, Kwong-Sak Leung, Kin-Hong Lee, Man-Hon Wong
Department of Computer Science and Engineering, Chinese University of Hong Kong
{*cmtse,hjli,ksleung,khlee,mhwong*}*@cse.cuhk.edu.hk*

***Abstract*—Discovering new drugs for emerging diseases has been a challenging task. There are numerous drug design techniques including fragment-based and diversity-oriented methods but their accuracies and efficiencies are low. By incorporating visualisation, biomedical experts can interact with the process to produce drug-like ligands more efficiently. The paper presents an interactive drug design algorithm which generates lead candidates against a protein. A set of drug candidates, created by an inhouse fragment-based method and docked on the target protein, are visualised in the virtual reality settings. Biomedical experts can investigate and select some of the ligands for further processing, aided with distance and bonding information. It also assists the user to drag and rotate the ligand to the binding site they find suitable. The algorithm runs iteratively and improves the quality of lead candidates every step. The paper compares the quality of resulting ligands between interactive and automatic approaches.**

***Keywords*-Interactive Drug Design; Virtual Reality;**

I. Introduction

The momentum in searching for compounds of medical uses continues to grow as diseases are more difficult to cure owing to drug resistance. Out of the possible configurations of the molecules which may have medical purposes [1], only a minority of the molecules were synthesised and exploited. It becomes crucial to design new molecules which can be of medical values but not explored. Computational methods have been employed because of low cost and high efficiency.

There are numerous computational techniques, mainly fragment-based and diversity-oriented, to generate drug candidates for wet-lab experiments [2], [3]. Fragment-based method constructs a library of structurally diverse small molecules that could become fragments of active drugs [4], [5]. The drug candidate starts with low affinity for the target in which is systematically altered and enlarged, generating high affinity, drug-like lead compound. Diversity-oriented method produces a library of structurally diverse drug-like compounds, usually from common intermediates [6], [7]. The compounds are then screened and high affinity candidates are optimised for further analysis. The accuracy of both strategies rely on the screening procedure, which could be a docking program in a recent approach [8].

In addressing the accuracy issues, there are interactive approaches which enable optimisation to the compound using the user's knowledge about structure-activity relationship [9]. The algorithm visualises a set of drug candidates for the user to choose from and generate new candidates based on their choice using evolutionary algorithm. It overcomes the difficulties in creating the fitness function to assess drug design. However, it considers solely on the structure of the drug candidates while the target protein is often known especially for pharmaceutical companies.

We propose an interactive algorithm for drug design with a known target protein. Visualisation plays an important role in displaying ligand-protein structure in 3D. The user can then investigate the structure and determine whether a drug candidate is viable. While there are good visualisers such as JMol and ligand editors by MolSoft, manipulating a drug candidate around requires adequate depth information. Virtual reality enables immersive experience to the user where conventional visualisation techniques cannot. The program generates a set of drug candidates by evolutionary algorithm and refined by chemical rules and docking. The user gives feedback by translating or rotating the candidates and remove them if found unsuitable. Our study shows the resulting lead compound is lighter in molecular weight which is more readily absorbed and has comparable affinity to the automatic method.

The detailed design and algorithm is presented in section III. Experiments are described in section IV A discussion on the advantages and improvements of the interactive approach is given in section V.

II. Related Work

There are many attempts in combining visualisation with the biomedical field. Visualisation and sometimes virtual reality are brought into place to help solve complex biomedical problems such as protein docking and drug design.

1) Interactive Drug Design: A drug design tool Molecule Evoluator [9] uses atom-based evolutionary approach to explore multiple configurations of drug candidates. The tool displays numerous configurations in 2D for the user to choose from and modify. The evaluation however depends completely on the user. In addition, when the target protein is available, the tool cannot take advantage of the information by estimating the affinity such as docking.

2) Virtual Reality in Biomedical Field: The function and interaction of proteins depend heavily on their conformations which are best visualised in virtual reality. The advantages of comparative visualisation was investigated in [10]. There is a shortcoming in their visualiser that it does not support the standard Protein Data Bank [11] format. An attempt to

1550-6037/11 $26.00 © 2011 IEEE
DOI 10.1109/IV.2011.72

assist protein docking using virtual reality was made in [12]. The user guides a ligand to the binding site of a protein with real-time feedback from their system. The user needs to control the whole process for the algorithm to work.

III. Methods and Design

The program consists of an interface which displays the ligand-receptor pair in virtual reality and a computational module to provide feedback. We first explain the algorithms in the computational module and then describe the features of the interface.

A. Hybrid Drug Synthesis

The drug synthesis algorithm employed in our program combines the strength of fragment-based and diversity-oriented approaches. A diversity-oriented approach creates structurally diverse molecules with their molecular masses usually close to those of drug-like compounds. In addition, it explores molecules that escape the attention of human or even nature. A fragment-based approach produces numerous sets of compounds not represented in existing libraries, filling the gap of 'chemical space'. It is suggested that tremendous amount of compounds can be represented in a library of few fragments.

The program accepts a small molecule and a receptor of Protein Data Bank (PDB) format. The small molecule is then be optimised to bind into the receptor. The program uses a fragment library ranging from hydroxide to benzene to add onto the small molecule. Evolutionary algorithm (EA) is employed to generate multiple instances of the small molecule and maintain diversity. It is based on biological principles such as natural selection and usually involves mutation and crossover operations. There are four operators implemented, namely, mutation, crossover, merge and split. The detail of the operators are as follows.

(1) Mutation: This replaces a hydrogen atom in a small molecule with a random fragment from the library. A hydrogen atom on the fragment is removed such that a non-hydrogen atom is attached to the small molecule. The small molecule is optimised by rotating the fragment to achieve minimal torsion.

(2) Crossover: Two small molecules are selected and exchange some of their fragments. The implementation resembles genetic programming that a subtree of two different molecules are selected and swapped. The small molecule is in a graph-based representation and cannot contain incomplete cycles. Thus, when a random atom on the fragment is selected, excluding the common initial compound, all dependent atoms are checked and selected for the exchange. Two small molecules are optimised and generated.

(3) Merge: An operation similar to crossover that two small molecules are chosen. All fragments on one of the small molecule are transferred to the other small molecule, removing some hydrogen atoms. Only the larger small molecule is returned, decreasing the number of small molecules in the set by one.

(4) Split: This transfers some randomly chosen fragments on a small molecule to the initial common compound. The valences of the small molecule which loses fragments is corrected by adding hydrogen atoms. Two small molecules are optimised and returned, increasing the number of small molecules in the set by one.

Apart from the mentioned operators, there are two special methods to produce larger variance to the small molecules which is restrictive to invoke.

Join Ring: This is an advanced mode of mutation that the ring in the fragment is joined to a ring on the small molecule, forming consecutive rings. Both the fragment and the small molecule must contain ring structures to perform. Two adjacent atoms on the ring are checked to find a matching pair in another ring. If the rings cannot be joined, the usual mutation is carried instead.

Decrease Bond Order: If there is a double or triple bond in the small molecule, the order of that bond is decreased by one. Hydrogen atoms are added to correct the valences. The small molecule is optimised and contains more bonds to add fragments.

Using the operators and methods mentioned, a set of small molecules is generated. The user can investigate the structures of the generated small molecules. However, it usually takes tens of different small molecules for the evolutionary algorithm to work well. The user may not want to explore the whole population. In addition, the decision of the user can benefit from computing affinity between the small molecules and the receptor. A docking algorithm is integrated as the fitness function of the evolutionary algorithm. An external docking program AutoDock Vina [13] is integrated because it is fast and relatively accurate. Its bundled tool, MGLTools, computes the flexible side chains and partial charges of the small molecule and receptor.

A compound follows certain chemical rules. If the graph representation is evolved freely, invalid molecule may arise. The Lipinski's rule of five [14] is applied to check the validity of the resulting small molecules evolved. The rules describe the number of acceptors, donors, molecular weight and solubility of a compound to be drug-like.

While the evolutionary approach can work on its own, there are several disadvantages. Firstly, the docking method may not be accurate enough in optimisation and sometimes incorrectly estimates the potency of the lead compound. Secondly, the compounds are difficult to compare. It is necessary to compare the physical and chemical properties among compounds which cannot simply represented by similarity index or subgraph differences. Lastly, evolutionary algorithm requires parameters tuning to work well. For this complex problem, the number of parameters is high which does not appeal to the users. In addition, it is more effective to adaptively adjust the parameters based on situations.

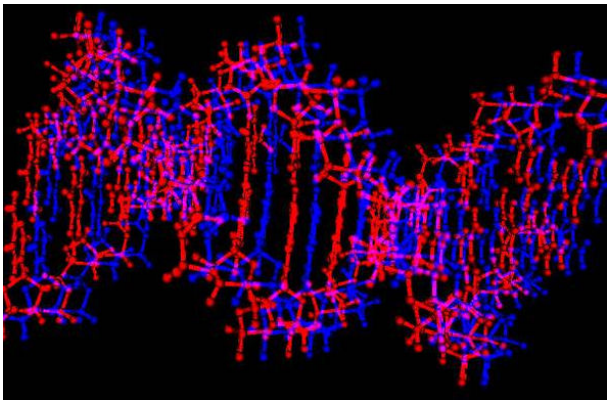

Figure 1. Stereo effect in anaglyph mode (red-and-blue). Using a spectacles with red and blue filter on each side, the disparity between the two superimposed DNA creates a perception of depth. The DNA appears in white in VR. This figure shows the effect without the spectacles.

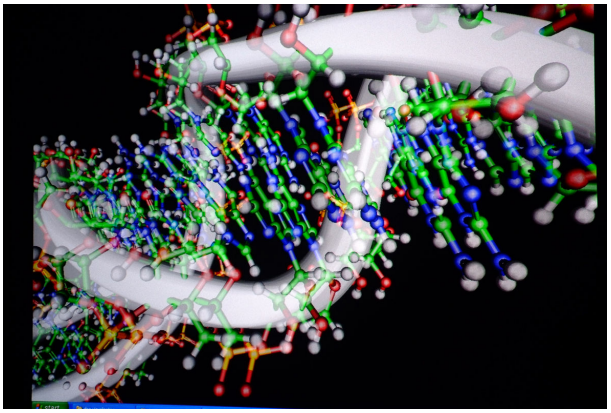

Figure 2. Stereo effect in shutter mode. Superposed images are displayed alternatively in temporal frames. The spectacles synchronise with the monitor such that each eye perceive the correct image. This figure shows the effect withour the spectacles.

We, therefore, introduce an interactive solution incorporating human intelligence. The fitness and similarity are decided by the users using their expert knowledge. The parameters of the evolutionary algorithm are adjusted based on the feedback of the user. This interactive approach relies on a rigorous visualisation implementation.

B. Interactive Interface in Virtual Reality

Realistic visualisation is important in biomedical research, especially in the field of drug design. Virtual reality (VR) environments allow a detailed inspection on molecular structure and offer a different quality than standard 3D representation.

The interface is developed in C++ and OpenGL programming language which is portable across platforms. The interface can visualise PDB format in standard 3D, red-and-blue stereo or shutter stereo mode. When paired with the drug design algorithm, the interface provides feedback to the user and computes the parameters for the algorithm. The standard 3D visualisation enables users without equipment to produce virtual reality environment to use the drug design algorithm.

The red-and-blue stereo mode (Figure 1) is a compromise in visualising structures in single colour which works on common monitors. It could display depth information better than the standard 3D which is crucial when judging the binding position. The shutter stereo mode (Figure 2) requires a high frequency monitor and special spectacles, but provides the best experience. Molecules visualised in this mode are rich in colour and depth information is presented precisely.

There are several display styles the interface supports which visualises different levels of information. The supported styles include DNA backbone model, ball-and-stick model, cartoon representation and van der Waals filling model. The van der Waals filling model requires surface mesh computation which can take up to half a minute

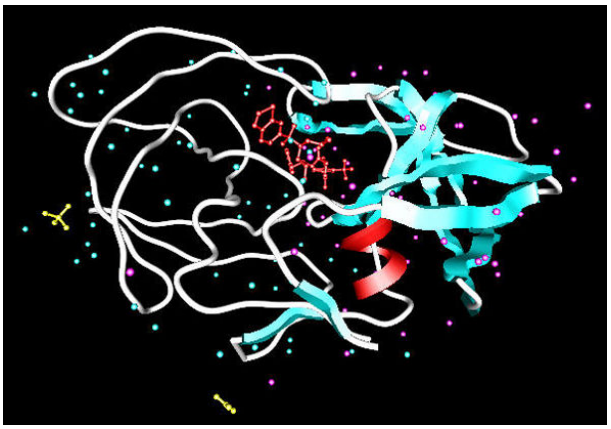

Figure 3. Cartoon representation of 3XME. Alpha helices and beta strands are represented by ribbons and arrows respectively. The ligand 1F1 is presented by ball-and-stick model inside the protein

upon the initial rendering. After the initial rendering, all models can be manipulated in real-time in stereo on common machines.

The ball-and-stick model (primary structure) is used to display the chemical aspects of the molecules. This is the only model available to the small molecules. This mode is used to visualise the bonding between a small molecules and its receptor.

The cartoon representation (secondary structure) reveals information on the shape and size of a protein. This mode uses simplified symbols such as arrows and ribbons to represent beta strands and alpha helices of the protein.

The van der Waals filling model (surface representation) visualises each atom in the protein using their van der Waals radius. This mode presents surface information of the protein which allows investigation on the binding sites.

The display capabilities of our interface may not be comprehesive as the mature display tools such as JMol,

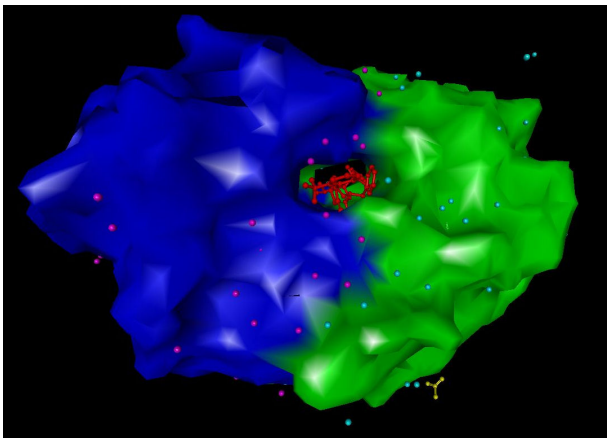

Figure 4. Surface model of 3XME. The protein is coloured according to its chains and a binding site is clearly shown.

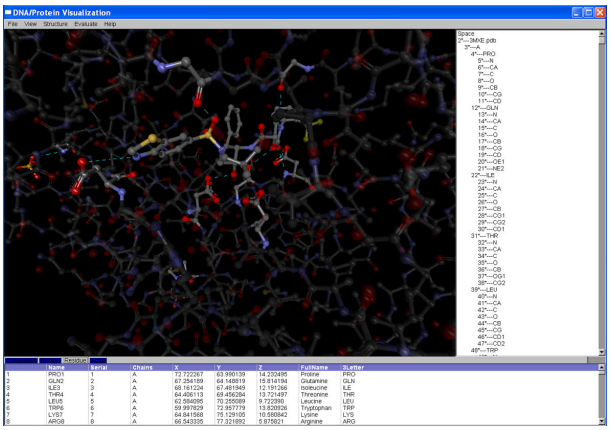

Figure 5. Interface of the interactive drug design program. Table on the right displays atom information in a hierarchy. Table on the bottom shows information in group such as chains and residues. The hydrogen bonds between the ligand and protein 3XME are displayed in the main window, with non-interacting residues faded out

which supports a variety of visualisation choices. When compared with the editing tool from Molsoft, their tool has more modification options and a well-established evaluation function. Nevertheless, the interface we developed has more features than the existing tools. It includes multiple stereo modes for different hardware configurations. In addition, our tool evaluates multiple drug candidates and assigns fragments automatically which can reduce the effort of the user.

The interface has a collision detection algorithm integrated to support dragging by the user. The user can select the protein by atom, residue or chain levels. It is done by directly clicking on the visualised proteins or the names in the list which contains the sequences in text form. In addition, bonding between the small molecule and the protein can be highlighted. The residues which bond directly with the small molecule is shown normally while the other fades out. The bonds are calculated using hydrogen bond or distance criteria.

When the interface is used with the drug design algorithm, the initial small molecule and the receptor must be specified. Two parameters, the number of small molecules generated and investigated, are defined by the user. A sufficient number of small molecules to be generated can encourage diversity among the population and reduce the chance of premature convergence. The user only needs to investigate a portion of the generated small molecules. A subset is selected for investigation according to the affinity, calculated by the docking program, and the molecular weights of the small molecules. The initial population is then created by mutating the specified small molecule.

During each generation, the user investigates a set of small molecules and decides whether to accept or remove them. Only one ligand-protein pair can be investigated at a time due to the complexity of the structure to be drawn. The user is assisted by a variety of information to make the decision. The statistics of the small molecule, including number of donors, acceptors, estimated logP and molecular weight, are displayed. The docked structure is visualised in virtual reality with possible hydrogen bonds drawn. The structure can be rotated and zoomed in for more details. The user can cycle through different model styles and highlight the interacting residues. The small molecule can be dragged and rotated, hydrogen bonds are updated in real-time. The user can specify to use the modified structure in the next generation of the drug design algorithm.

Based on the decision of the user, parameters of the drug design algorithm is dynamically calculated. The adaptive parameters includes mutation rate and crossover chance. These parameters are changed each generation depending the investigation result. When the user removes a majority of the small molecules for investigation, the mutation rate becomes high in order to create structurally diverse small molecules, which covers more ‘chemical space’. In contrast, when the user accepts most small molecules, crossover is encouraged to create similar compound of the accepted ones. The molecular weight also plays a role in changing the parameters. There is an implicit molecular weight in the algorithm, which is the average of molecular weights of the small molecules for investigation. After the user accepts a certain number of small molecules, their average molecular weight is calculated. If the new weight is smaller than the implicit weight, it means that the user prefers smaller compound. The split operator will be invoked more often. The opposite applies that merge operator is used more often when the averaged weight is larger than the implicit weight. The two advanced methods, join ring and decrease bond order, need to be instructed by the user and are not used normally. The program can be stopped any time when the user is satisfied with the generated small molecules.

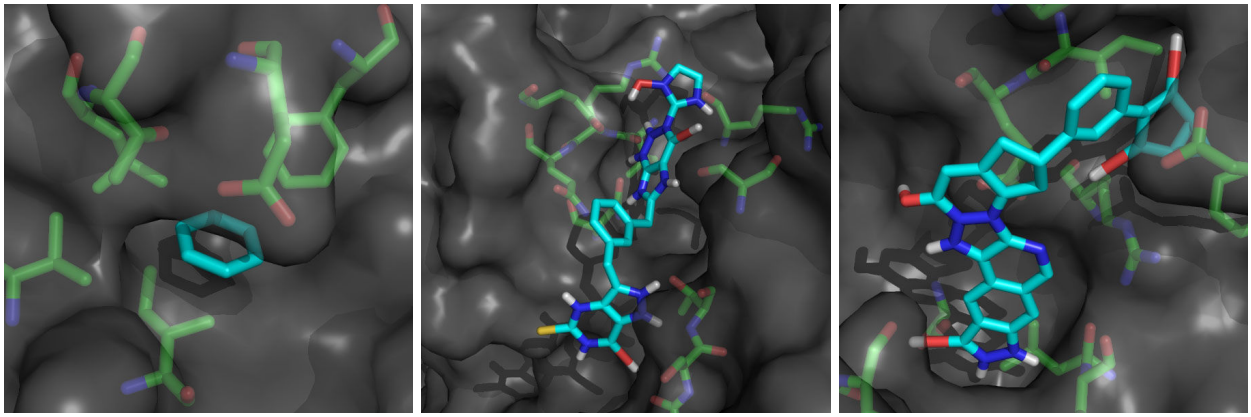

Figure 6. Putative binding site on EV71 protease. (a) Initial scaffold in binding site. (b) Resulting lead candidate by the automatic algorithm. (c) Resulting lead candidate by the interactive algorithm.

IV. Experiments

We conducted experiments on the program using several receptors and compounds from the ZINC database [15]. We will explain one of the experiments which illustrates the idea of the interactive drug design algorithm. An automatic drug design algorithm, AutoGrow, was chosen for comparison. This algorithm is fragment-based and utilises evolutionary algorithm.

The receptors chosen are glycogen synthase kinase 3 beta (GSK3β), HIV reverse transcriptase, HIV protease and Human Entrovirus 71 3C protease. Their corresponding PDB ID are 1J1B, 2ZD1, 3KFN and 3OSY where the first three have known binding sites and inhibitors. We screened all compounds in the ZINC databases against the receptors using the docking program. Five initial compounds were selected which have high affinity and work on the automatic drug design program. The fragment library used for both programs was the same which contained 46 small fragments. For the interactive program, a population of 50 was set. The number of small molecules for investigation was 10. For the automatic program, it used the same population size. The number of elitists was set to 10. Their program required to set the number of molecules produced by mutation and crossover, where they were set to 20 and 20 respectively.

Both programs were executed for eight generations. We only interacted with the first generation of small molecules to save effort. The program was run the end which took about two hours in total. The interactive apporach had an average of 10% better affinity. Moreover, the automatic program is not aware of physico-chemical property such that the molecular weight of the resulting compounds were 60% higher on the extreme, which were not drug-like.

In figure 6, the comparison indicates that the automatic program misplaced the binding region of the initial scaffold. The interactive approach created a compound of 8.9% better affinity and comparable molecular weight. In addition, the resulting compound of the interactive approach contained large consecutive rings not in the initial compound which is difficult to achieve in automatic methods.

V. Discussions

In this paper, we have described an interactive approach over the evolutionary algorithm for drug design. The graph representation preserves most of the information available in the structures. It is also suitable for fragment-based methods which have the potential benefit of easier synthesis. We used standard PDB format for displaying and processing which effectively utilises online resources. In addition, the interface can be used as a standard visualiser to display structure in virtual reality when not paired with the drug design algorithm. We have achieved real-time interation with the molecule in stereo mode, while the evaluation process may take longer time.

Adopting an interactive approach, we overcome the most difficult aspect in evolutionary algorithm for drug design, creating the appropiate fitness function. The evolutionary algorithm is specially designed to dynamically adjust itself each generation by the decision of the user. We have decided not to display all the small molecules in the population because the user may not want to screen tens or hundreds of compounds before passing them back to the algorithm. Instead, a portion of small molecules, which the algorithm thinks they are good, are visualised for investigation. The parameters of the evolutionary algorithm are calculated based on the decision about the selected set of small molecules. Through this design, we have abstracted the complicated parameter tuning process into simple reasoning.

There are many advantages to user interaction. The most attractive advantage to drug design is the feedback from the users can produce compounds which can be synthesised more easily in laboratory. With more expert knowledge, the resulting compounds will be more likely to be synthesisable. Another advantage is the user can use whatever domain

knowledge they have, whereas the developers may not possess. It is particularly difficult to assume all required knowledge is integrated beforehand. A third advantage is user-friendliness. Tuning the evolutionary algorithm can be tedious when there are a lot of parameters. The interactive approach reduces the tuning process to simple questions: whether to accept a small molecules or use certain methods. Nonetheless, using an interactive program can be more time-consuming than an automatic approach. The outcome may not be objective which depends on the perception of the user.

Considering the feedback of the users, we have planned to allow greater flexibility in the interface such as real-time modification to the small molecules. At the moment of writing, only translation and rotation are available to the user. With more explicit chemical rules implemented, the user can modify atoms or fragments on the small molecules. In addition, with recent capability to display in virtual reality on webpages, it is beneficial to have a web-front which encourages a larger user base.

CONCLUSIONS

We have designed and implemented an interactive approach to the drug design. Through visualisation on the docked structures in virtual reality, domain experts can investigate and manipulate with higher precision. Incorporating human intelligence in the fitness function, we overcome difficulties associated with physico-chemical properties of compounds. It also combines the domain knowledge of the user and processing power of computers.

ACKNOWLEDGEMENTS

We thank Bobby Tang and Bean Lam for developing the interface of the program. The research is supported by the grant CUHK414708 from the Research Grants Council of the Hong Kong SAR, China.

REFERENCES

[1] R. S. Bohacek, C. McMartin, and W. C. Guida, "The art and practice of structure-based drug design: A molecular modeling perspective," *Medicinal Research Reviews*, vol. 16, no. 1, pp. 3–50, 1996. [Online]. Available: http://dx.doi.org/10.1002/(SICI)1098-1128(199601)16:1%3C3::AID-MED1%3E3.0.CO;2-6

[2] P. J. Hajduk, W. R. J. D. Galloway, and D. R. Spring, "Drug discovery: A question of library design," *Nature*, vol. 470, no. 7332, pp. 42–43, Feb 2011. [Online]. Available: http://dx.doi.org/10.1038/470042a

[3] G. Schneider and U. Fechner, "Computer-based de novo design of drug-like molecules," *Nature Reviews Drug Discovery*, vol. 4, no. 8, pp. 649–663, Aug 2005. [Online]. Available: http://dx.doi.org/10.1038/nrd1799

[4] I. D. Kuntz, "Structure-based strategies for drug design and discovery," *Science*, vol. 257, no. 5073, pp. 1078–1082, Aug 1992.

[5] P. J. Hajduk and J. Greer, "A decade of fragment-based drug design: strategic advances and lessons learned," *Nat Rev Drug Discov*, vol. 6, no. 3, pp. 211–219, Mar 2007. [Online]. Available: http://dx.doi.org/10.1038/nrd2220

[6] R. S. Bohacek and C. McMartin, "Multiple highly diverse structures complementary to enzyme binding sites: Results of extensive application of a de novo design method incorporating combinatorial growth," *Journal of the American Chemical Society*, vol. 116, no. 13, pp. 5560–5571, Jun 1994. [Online]. Available: http://dx.doi.org/10.1021/ja00092a006

[7] D. R. Westhead, D. E. Clark, D. Frenkel, J. Li, C. W. Murray, B. Robson, and B. Waszkowycz, "Pro-ligand: an approach to de novo molecular design. 3. a genetic algorithm for structure refinement," *J.Comput.Aided Mol.Des*, vol. 9, no. 2, pp. 139–148, Apr 1995.

[8] J. D. Durrant, R. E. Amaro, and J. A. McCammon, "Autogrow: a novel algorithm for protein inhibitor design," *Chem.Biol.Drug Des*, vol. 73, no. 2, pp. 168–178, Feb 2009.

[9] E.-W. Lameijer, J. N. Kok, T. Back, and A. P. IJzerman, "The molecule evoluator. an interactive evolutionary algorithm for the design of drug-like molecules," *Journal of Chemical Information and Modeling*, vol. 46, no. 2, pp. 545–552, 2006. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/ci050369d

[10] E. Moritz and J. Meyer, "Interactive 3d protein structure visualization using virtual reality," in *Proceedings of the 4th IEEE Symposium on Bioinformatics and Bioengineering*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 503–. [Online]. Available: http://portal.acm.org/citation.cfm?id=998667.998837

[11] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic acids research*, vol. 28, no. 1, pp. 235–242, January 1 2000.

[12] A. Anderson and Z. Weng, "Vrdd: applying irtual eality visualization to protein ocking and esign," *Journal of Molecular Graphics and Modelling*, vol. 17, no. 3-4, pp. 180 – 186, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/B6TGP-3YN94J2-3/2/fe8e17c8cb8f8f0c3309af690b83b04a

[13] O. Trott and A. J. Olson, "Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J.Comput.Chem.*, vol. 31, no. 2, pp. 455–461, Jan 2010.

[14] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv.Drug Deliv.Rev.*, vol. 46, no. 1-3, pp. 3–26, Mar 2001.

[15] J. J. Irwin and B. K. Shoichet, "Zinc: A free database of commercially available compounds for virtual screening," *Journal of Chemical Information and Modeling*, vol. 45, no. 1, pp. 177–182, 01/01 2005, doi: 10.1021/ci049714+; M3: doi: 10.1021/ci049714+. [Online]. Available: http://dx.doi.org/10.1021/ci049714+