Journal of the Taiwan Institute of Chemical Engineers 41 (2010) 623–635

Contents lists available at ScienceDirect

Journal of the Taiwan Institute of Chemical Engineers

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

Review

Current developments of computer-aided drug design

Hung-Jin Huang [a,1], Hsin Wei Yu [a,1], Chien-Yu Chen [a], Chih-Ho Hsu [a], Hsin-Yi Chen [a], Kuei-Jen Lee [b], Fuu-Jen Tsai [b,c], Calvin Yu-Chian Chen [a,b,*]

[a] Laboratory of Computational and Systems Biology, School of Chinese Medicine, China Medical University, Taichung 40402, Taiwan
[b] Department of Bioinformatics, Asia University, Taichung 41354, Taiwan
[c] Department of Medical Genetics, Pediatrics and Medical Research, China Medical University Hospital and College of Chinese Medicine, China Medical University, Taichung 40402, Taiwan

ARTICLE INFO

Article history:
Received 26 November 2009
Received in revised form 12 March 2010
Accepted 20 March 2010

Keywords:
Molecular simulation
Molecular dynamics (MD)
Docking
Computer-aided drug design (CADD)
Quantitative structure–activity relationship (QSAR)

ABSTRACT

The continuous advancement in molecular biology and information technology aided the development of a rich molecular simulation repertoire that can be applied in system biology, proteomics, molecular biology, bioinformatics, and materials science. We attempt to introduce the latest developments in drug design based on computational techniques, including protein structure modeling, docking, binding site prediction, quantitative structure–activity relationship (QSAR), and molecular dynamics simulation. Furthermore, a brief discussion on current docking issues, including accuracy of protein structure and protein–ligand interaction, is also included. Weight equation and rules and a new concept on flexibility are also described here as possible solution for these issues.

Contents

* Corresponding author. Current address: Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Tel.: +1 617 353 7123.
E-mail addresses: ycc@mail.cmu.edu.tw, ycc929@MIT.EDU (C.-C. Chen).
[1] These authors contributed equally.

1876-1070/$ – see front matter © 2010 Taiwan Institute of Chemical Engineers. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.jtice.2010.03.017

1. Introduction

The research fields in chemical engineering have always been changing and evolving, from the field of applied industrial chemistry at the beginning of the last century, through the revolutionary reformulation of unit operations, transport phenomena and engineering science in the 1960s, to the extensive use of computing technology and the incorporation of molecular biology over the last two decades. This latter change is gradually being adopted by prestigious research institutes and universities, including the Department of Chemical and Biomolecular Engineering of Johns Hopkins University, which has shifted research focuses to biological related issues and revised the engineering undergraduate and graduate teaching curricula to integrate biomolecular modeling into process dynamics and control. Indeed, the integration of computational force and molecular biology, such as to simulate the behavior of molecules, is becoming a mainstream in the chemical engineering research and has received much attention from the pharmaceutical industries.

Molecular simulations are an interdisciplinary science with different applications in different research area. In polymer science, much use of molecular simulation has been put in studying fluxional behavior. As for monitoring heat transfer in semi-conductor, simulation is used for studying thermal profile. In informatics, the emphasis is on developing more powerful, fast simulation models that can accurately predict or describe scientific phenomenon. As for uses in drug design and bioinformatics, there are two different research focuses concurrent in the scientific field. The first is to design new mathematic algorithm for more realistic calculation. The second is to apply current algorithms on molecular biology research, such as simulating protein–ligand or protein–protein interactions, and use the result for further biomolecular experiments.

For a medicinal product to reach patients, commonly, more than 8 years of time and millions of dollars in investment are required to finish the long tedious drug development process. Furthermore, only a handful can finish the clinical trial and pass the strict inspection of drug regulatory agency, despite that thousands of new therapeutic candidates are being discovered in laboratories every year. However, the recent advances in technologies, namely automated platform, computational chemistry and computer-aided drug design (CADD), are now offering a fast track to some limiting factors of therapeutic discovery as well. Computer-aided drug design (CADD), that offers an *in silico* alternative to medicinal chemistry techniques for studying the structure and predicting the biological activity of drug candidates, has the advantages of both speed and low cost and is becoming an indispensable program of major pharmaceutical companies.

There are two major application areas of CADD, namely structure-based drug design and ligand-based drug design. Structure-based drug design relies on three-dimensional knowledge of the receptor structure and its active sites to investigate interaction, binding energy and steric relationship between ligand and receptor. Ligand-based designing approach, on the other hand, relies on knowledge of ligands that interact with target of interest. This technique employs statistical methods to link structural features to biological activities and attempts to identify specific structural features of a ligand required for interacting with its target. Both structure-based and ligand-based techniques can be applied in the initial drug discovery process and aid the discovery of a lead compound which serves as the starting basis for further modification to improve pharmacokinetics, solubility, selectivity, potency or stability. Two of the great advantages of CADD lie in the ability of fast screening a large molecule databank and the accelerated time steps of identifying notable medicinal chemistry features. These characteristics are extremely beneficial in designing multi-target medicinal products.

As a short review to introduce the basic of molecular simulation in drug design, we arbitrarily categorized CADD into three major sections: (a) structure-based drug design, (b) ligand-based drug design, and (c) molecular dynamics. A summary of CADD process flowchart is shown in Fig. 1. In addition, to the brief overview, we will address on issues encountered by biochemists on using docking programs, including the low hit rate of docking programs (Kontoyianni *et al.*, 2005; Warren *et al.*, 2006) and the low replicability of predicted protein–ligand interaction (Baxter *et al.*, 1998). For these issues, we will attempt to improve the docking results by introducing a parameter accounting for the flexibility of protein and ligand. Overall, we hope the readers can gain much insight into CADD by our concise introductions on major- and sub-topics, practical examples of relevant application cases, and sample curricula of CADD courses.

2. Structure-based drug design

To design a medicinal product for treating a disease or relieve a symptom, a clear understanding of the disease pathway and relevant processes is crucial for selecting a therapeutic target. Thus, in the past we have employed programs such as GeneGo and KEGG to build pathway maps of hypoxia-inducible factor in brain injury, shown in Fig. 2, for identifying critical signal or transcription pathways, specific protein–protein interactions and relationships between upstream and downstream proteins. This background knowledge has proven helpful to us for selecting key therapeutic target.

2.1. Protein structure determination

For structure-based drug design, a priority before investigating receptor–ligand relationship is to obtain the target structure. There are two major methods for protein structure determination by physical measures, X-ray diffraction and NMR (Marti-Renom *et al.*, 2000). The solved protein structures can be readily found at Protein Data Bank (www.rcsb.org/); however, for proteins that have not been solved or are difficult to isolate, modeling approach can be used.

2.1.1. Homology modeling

Homology modeling is a fast method to obtain protein structures that can not only be used in studying rational drug design but also for protein–protein interaction and site-directed mutagenesis (Josa *et al.*, 2008; Mohan *et al.*, 2009; Sujatha *et al.*, 2009). Proteins lacking structural information could be constructed if they have over 30% sequence identify with their related homologous proteins (templates) (Marti-Renom *et al.*, 2000). This modeling strategy has been widely applied in many researches and in our past studies as well (Chen, 2008a,b,c; Chen, 2009a,b,c,d;

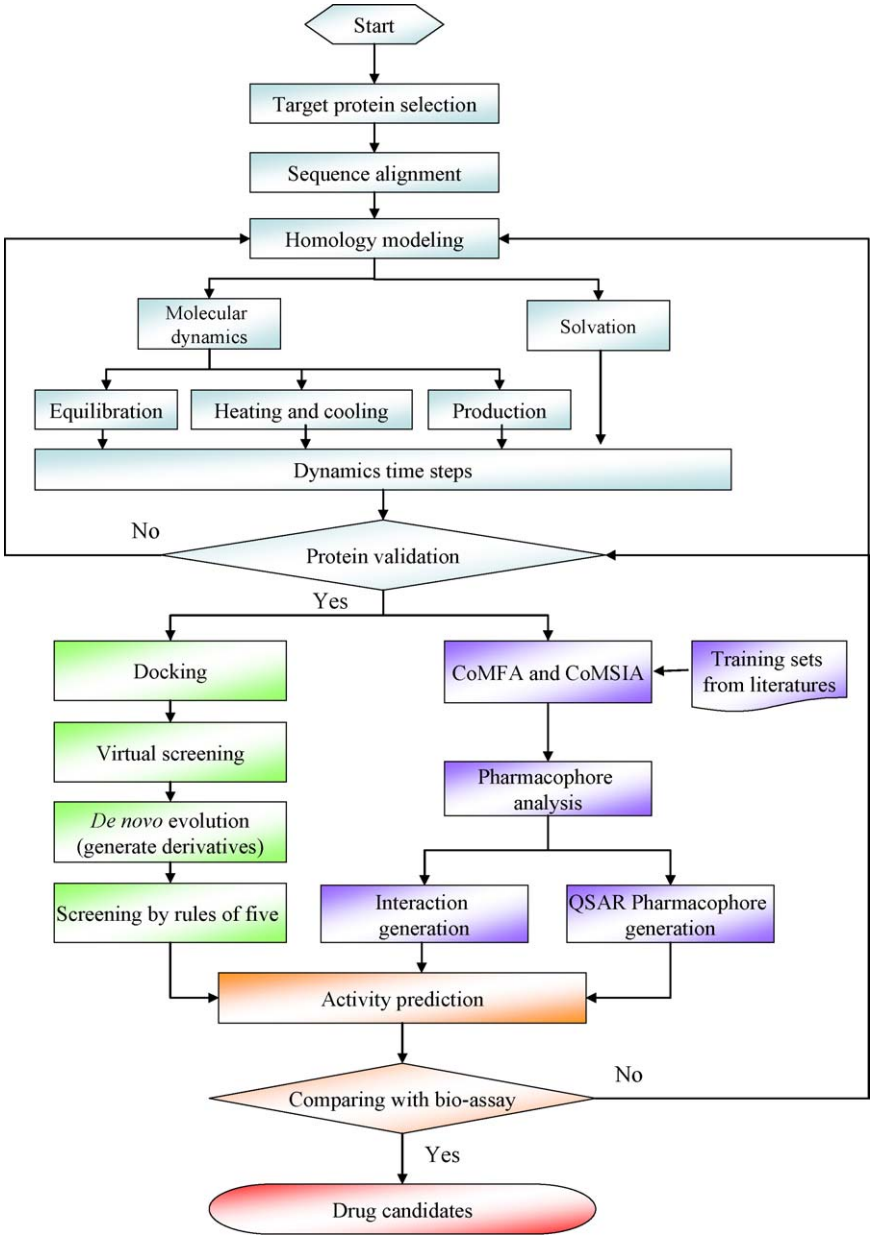

Fig. 1. An example of a computer-aided drug design flowchart.

Chen, 2010a,b; Chen and Chen, 2007; Chen *et al.*, 2008a,b; Chen *et al.*, 2009, 2010; Ding *et al.*, 2008; Lin *et al.*, 2009; Sheu *et al.*, 2009).

For our studies on H1N1 influenza virus, we have used homology modeling approach to construct hemagglutinin and neuraminidase structures from newly identified viral protein sequences and solved protein structures. A past sequence alignment result of the H1 and N1 sequences to templates, shown in Fig. 3, shows that both H1 and N1 have sequence identity and similar over 75%, which gives us high confidence in using the templates for modeling.

The modeled structures can be further modified in model refinement to be consistent with the experiment data in covalent bonds, geometry, and energy configuration. Force fields, such as CHARMM, AMBER, CVFF, CFF91, and GROMOS can also be applied to molecules for calculating energy minimization, which uses the function (Payne *et al.*, 1992) shown below:

$$E_{\text{total}} = E_{\text{stretching}} + E_{\text{bending}} + E_{\text{dihedral}} + E_{\text{out-of-plane}} + E_{\text{crossterms}} + E_{\text{VdW}} + E_{\text{coulombic}} \quad (1)$$

To ensure the rationality of the modeled structures, checks on stereochemistry, energy profile, residue environment, and structure similarity are often needed. Stereochemistry considers the bond angles and lengths, the dihedral angles of major chains, and the non-covalent bonds of amino acid residues within a protein. Two examples of our stereochemistry check are shown in Fig. 4. For our modeled hemagglutinin, the Ramachandran graph shows that 94.4% of H1 residues are in the allowed region while only 2.5% are in the disallowed region. Similarly, for modeled N1 structure, 91.4% of residues are found in the allowed region and only 2.5% of residues are in the disallowed region.

Energy profile is based on Profile-3D that analyzes the compatibility of amino acid sequences with three-dimensional environment (Al-Lazikani *et al.*, 2001). The Profile-3D graphs of H1 and N1 model are shown in Fig. 5. There are several factors that can influence the verify score. In conditions where the hydrophobic residues are folded on protein surface or the polar residues are folded into protein core, decreases in verify scores are likely to be seen. For regions that have verify scores above zero are considered

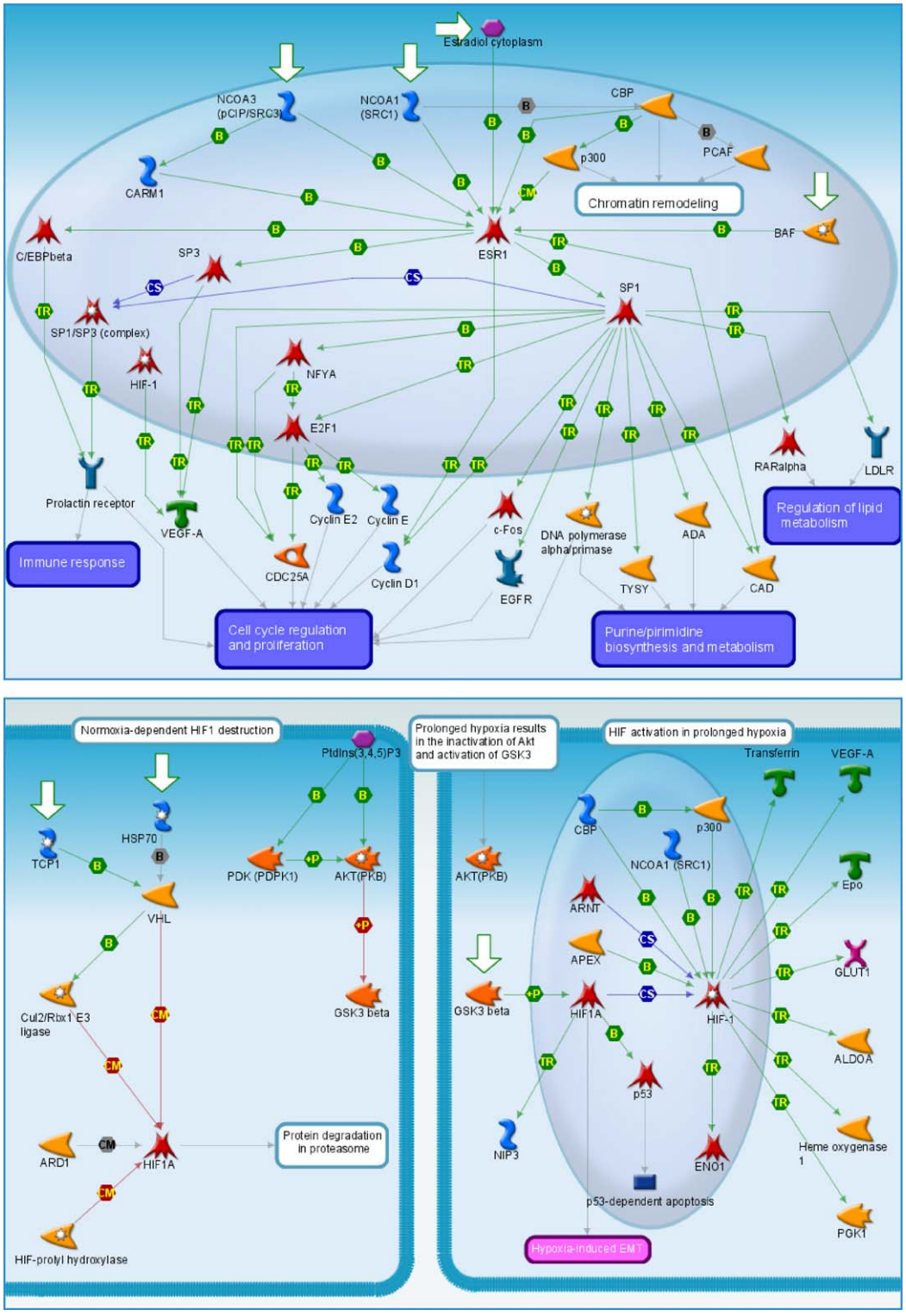

Fig. 2. The signal pathway of HIF protein.

to have stable folding and more energetic favor three-dimensional structure.

2.1.2. Folding recognition

Also known as "threading," folding recognition was brought up in 1991 by Bowie and colleagues whom employed this method to describe the environment of residues interactions. Folding recognition calculates the probabilities of the 3D structures could form by given protein sequences (Mishra, 2009). Both the environment of residues interactions and the protein surface area are considered in the threading protocol. Structure with the highest probability is recommended to construct the protein model.

2.1.3. Ab initio protein modeling

The *ab initio* method is based on physical principles, residue interaction center and lattice representation of a protein to build the target (Adrian-Scotto and Vasilescu, 2008; Deepa and

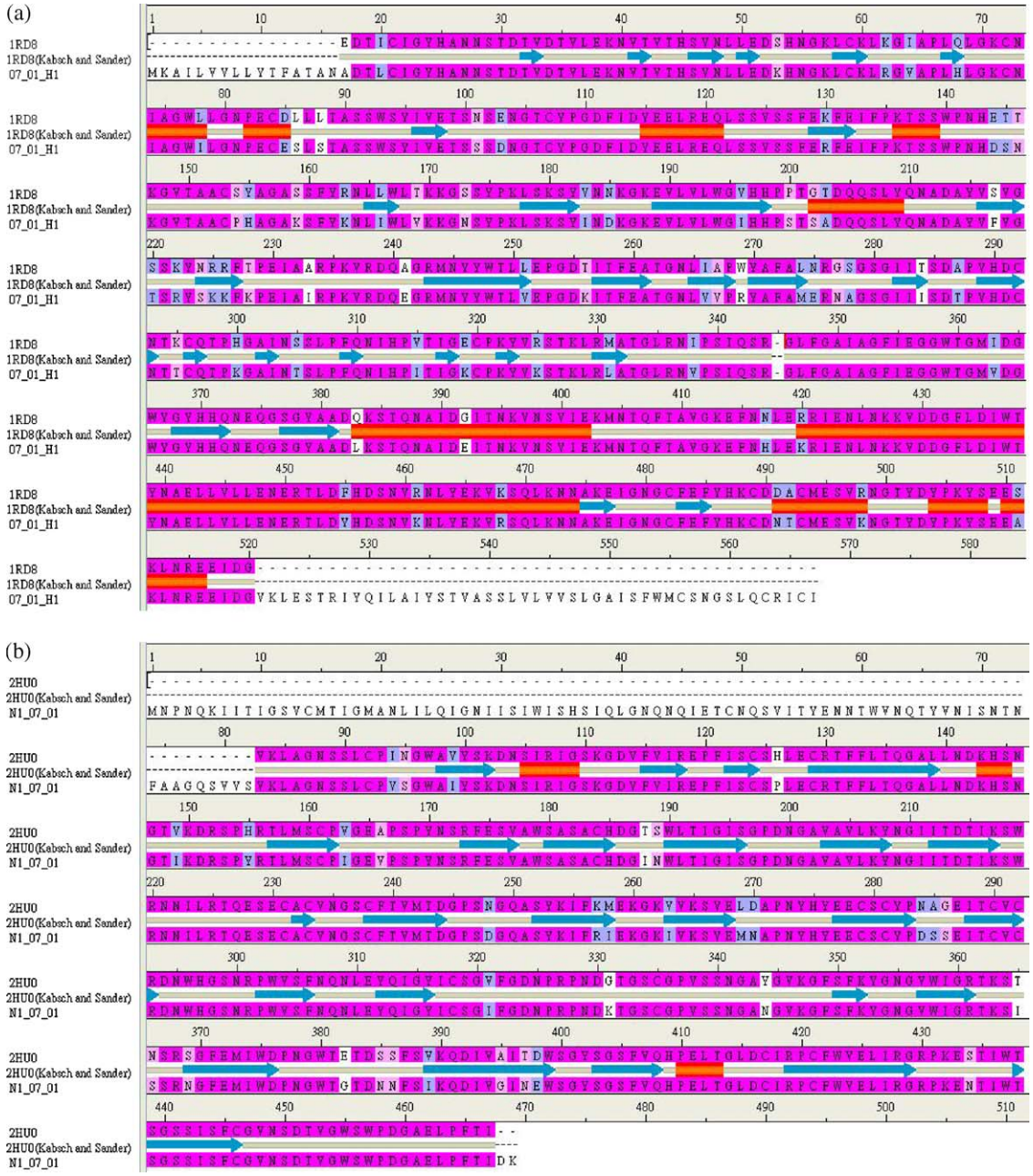

Fig. 3. The sequence alignment of (a) H1 (75.4% sequence identity and 79.8% sequence similarity) and (b) N1 (83.6% sequence identity and 92.1% sequence similarity).

Kolandaivel, 2008). This method is extremely useful when the other protocols fail to predict an unknown protein structure (Huang *et al.*, 1998). However, the identity and accuracy given by *ab initio* modeling could be lower than other approaches. Protein folding is not only a physical action, but also involves many biochemical actions originated from inherent residues interaction (Sippl, 1993). Based on this concept, *ab initio* method hypothesizes that: when a protein folds, it would tend to achieve the most energetically favorable state (Luthy *et al.*, 1992).

2.1.4. Hot spot prediction

Another important issue in structure-based drug design is to determine the ligand active site. While the active site may be determined via ligand location in the crystal lattice after X-ray crystallography, this method is not possible for proteins that cannot be crystallized. Several binding site determination methods have been invented to address this issue and FTMAP (Brenke *et al.*, 2009) is one of the recently developed methods being investigated in our lab.

The primary strategy of FTMAP utilizes small molecular fragments as a probe for exploring protein surface. Spots where molecular fragments clustered are predicted to be the favorable druggable sites. Significant hydrogen bonds and non-bounded interactions can also be explored between the probes and protein. In addition, the structure of the molecular probes can be the starting basis for designing new medicinal products.

The reliability of FTMAP has been confirmed in the past by comparing the predicted with the experiment results of Allen *et al.* (1996) and Mattos *et al.* (2006). Consistencies in binding site location and protein–probe interactions have been observed in the comparisons (Brenke *et al.*, 2009). Landon *et al.* (2009) have also conducted FTMAP researches with molecular experiments verifications as well.

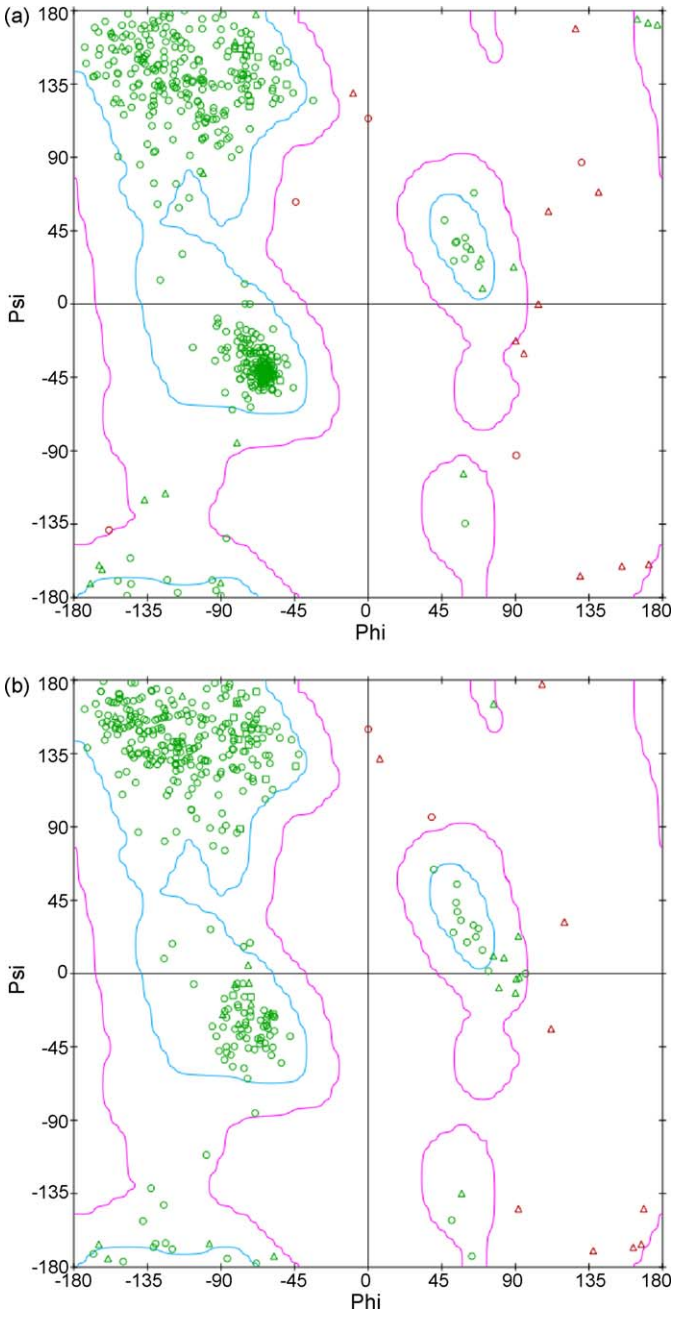

Fig. 4. The Ramachandran plot of (a) H1 and (b) N1.

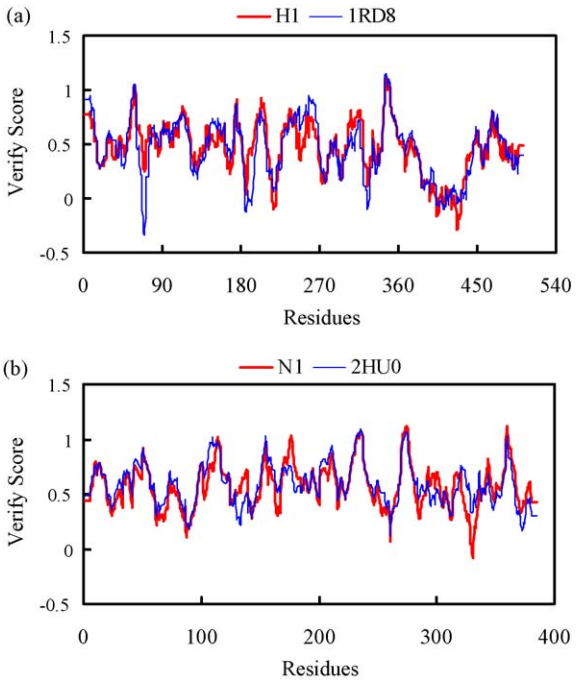

Fig. 5. The verify score of (a) H1 and (b) N1.

We have used FTMAP in our researches on protein phosphodiesterase-5 (PDE-5) to determine potential hot spots. Potential druggable sites predicted by FTMAP are shown in Fig. 6, and histograms of calculated protein–probe interactions are shown in Fig. 7.

2.2. Docking

2.2.1. Autodock

The software AutoDock, developed by Olsen's laboratory in the Scripps Research Institute, is a program for docking small flexible ligands into a rigid 3D structure (Goodsell and Olson, 1990). A set of grid is used to describe the 3D structure, based on the AMBER force field, and generated with AutoGrid to calculate van der Waals and coulombic interactions. In version 1.0 and 2.0, the genetic algorithm and simulated annealing were utilized for searching the best binding model, but version 3.0 incorporated Lamarckian Genetic Algorithm (LGA) in the search, and the efficiency was greatly enhanced than the previous versions. In the current version, the linear regression analysis is used to obtain a free-energy scoring function, based on the AMBER force field, and a larger set of diverse receptor–ligand complexes is kept constant while the side-chains in the 3D structure are flexible. The applications of AutoDock are immense, including but not limited to computer-aided structure-based drug design, X-ray crystallography analysis, high throughput virtual screening, combinatorial library design, and protein–protein interaction study (Rajakrishnan *et al.*, 2008).

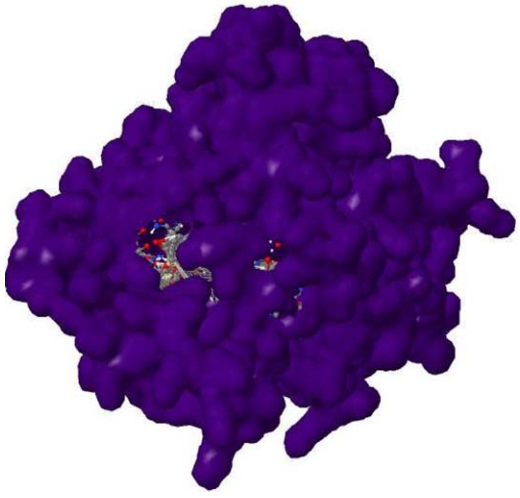

Fig. 6. PDE-5 hot spots predicted by FTMP. Small molecular fragments used include acetaldehyde, ethanal, ethane, acetamide, acetonitrile, acetone, methylamine, benzene, dimethyl ether, urea, N,N-dimethylformamide, ethanol, benzaldehyde, phenol, isopropanol.

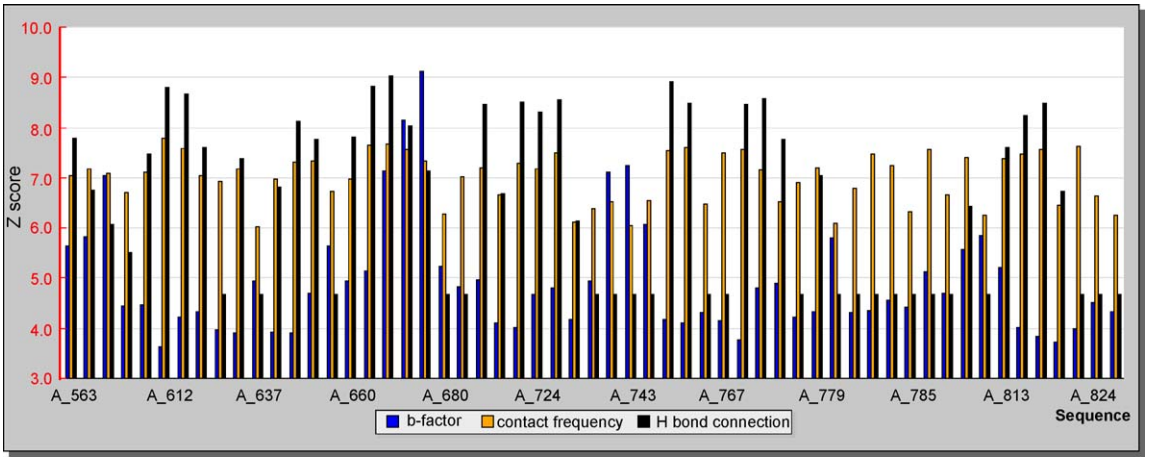

Fig. 7. Summary of H-bond interaction and non-bonded interaction found between PDE-5 protein residues and small molecular probes.

2.2.2. CDOCKER

The CDOCKER protocol is a CHARMM-based docking algorithm (Wu *et al.*, 2003) and retains all the advantages of full ligand flexibility. Ligand conformations are generated from the initial ligand structure by high temperature molecular dynamics (MD) simulation. The random conformations are refined based on grid-based simulated annealing and full molecular mechanics minimization. In the docking procedure, CDOCKER uses a sphere to define an active site, so the knowledge of the binding site is not required.

2.2.3. Flexible docking

The flexible docking protocol retains receptor flexibility during docking of flexible ligands (Koska *et al.*, 2008). The target receptor side-chain conformations are calculated in the first step and are generated by the ChiFlex algorithm. The ChiFlex algorithm creates various protein conformations with different side-chain orientations. The second step is providing low energy conformations of ligands for the docking process. The LibDock program is used for this docking process, which indicates the binding site where ligand polar and non-polar groups may be bound to the favorable positions of protein. The next step is to remove similar ligand poses. The refinement is performed in the final steps; the side-chains are refined using the ChiRotor algorithm and the CDOCKER for structure simulated annealing and energy minimization of each ligand pose. Overall, flexible docking can optimize the flexibility of the side-chains (Braun *et al.*, 2008). However, it generally requires extensive computing resources and generates more data than general rigid docking protocol.

2.2.4. LigandFit

LigandFit is a grid-based method for calculating receptor–ligand interaction energies, which is crucial in initial ligand shape match to the receptor binding site (Akten *et al.*, 2009; Ramalho *et al.*, 2009; Venkatachalam *et al.*, 2003). The LigandFit protocol contains three essential steps for docking ligands to the specified site: definition of the active site, analysis of ligand conformations, docking of ligands to a selected site, and scoring of the predicted poses. The first step is to determine the active site of a protein with known 3D structure. As afore mentioned this can be achieved by locating the ligand within the protein structure. If ligands are not available in the active site, flood-filling algorithm of LigandFit can be used to determine possible cavity region on protein surface. In the second step, LigandFit utilizes Monte Carlo (MC) method to generate the ligand conformations. As soon as one conformation is generated, it will be employed to dock with the receptor. The third step is the estimation of binding affinity (score) by grid-based energy calculation of the energy between the ligand and the receptor. The ligands will have to be docked into the receptor before calculating the docking scores. The docking scores can be calculated according to the following scoring functions: Dock Score, LigScore1, LigScore2, PLP1, PLP2, Jain, PMF, PMF04, Ludi energy estimate 1, Ludi energy estimate 2, and Ludi energy estimate 3.

2.2.5. Transmembrane protein modeling

Despite there are several prescription medicines that target transmembrane protein, such as HER2 and GABA receptor, at present a bottleneck lies in accurately analyzing transmembrane protein structure due to difficulties in crystallization. However, in addition to modeling transmembrane protein, considerations need to be paid to the influence of phospholipid cell membrane. Thus, a simplified force field for phospholipid bilayer can be included into the simulation process. In Accelrys Discovery Studio 2.5, membrane force field option is built based on CHARM (Im *et al.*, 2003; Spassov *et al.*, 2002). A snap shot of transmembrane protein simulation is presented in Fig. 8.

Although, simulating cell membrane may offer a more realistic insight into protein behavior, this method is not matured yet due to the exclusion of the mass of phospholipids in calculation. Thus, we believe that much development is still needed in this area.

2.3. Binding free energy

All the docking protocols discussed above do not include functions for calculating binding free energy in their protocols. To

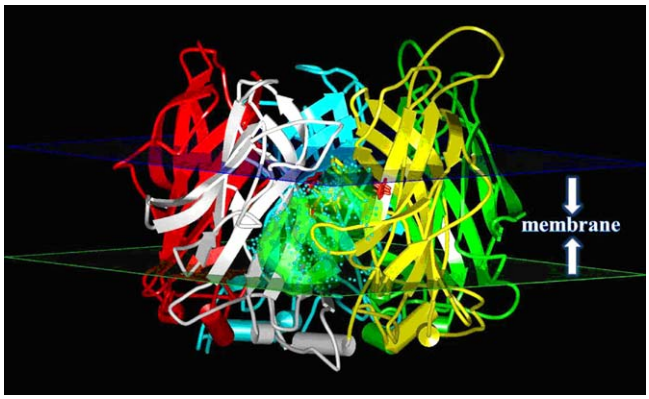

Fig. 8. A modeled GABA receptor with membrane force field.

calculate the binding free energy, information on the energy status of the protein–ligand complex, free ligands and unbound protein must be pre-determined. The energy is calculated using the formula (Kollman *et al.*, 2000):

$$\text{energy of binding} = \text{energy of complex} - \text{energy of ligand} - \text{energy of receptor}. \quad (2)$$

2.4. Flexibility of protein–ligand complex

Our research team has proposed a weight equation and rules (Chen, 2009d), attempting to improve the accuracy of the consensus scoring. Although our results supported weight score over consensus score, there are still areas needed for further development. In here, we will propose a rough concept, based on the flexible nature of protein and drug molecule. Currently, the rotation and fluctuation of protein and drug molecules can be simulated by using molecular dynamics. However, molecular dynamics simulations require extensive computing unit and time. Hence, it is impractical to perform large scale screening of a molecule database with molecular dynamics. Therefore, our current experiments are limited to virtual screening of database and then to simulations of a few possible candidates.

Hereby, we propose a concept:

(1) The difference in result of flexible docking and LigandFit is due to difference in flexibility of molecules, such that:

$$\text{flexibility} = \text{score of LigandFit} - \text{score of flexible docking}$$

(2) The result of molecular simulation is related to flexibility, and a positive relationship can be obtained in flexibility vs. molecular dynamics.

Furthermore, we hope to introduce flexibility parameter into docking algorithm to closely monitor real life situation. The equation is shown below:

$$\text{real docking score} = \text{docking score} \times \text{flexibility}$$

2.5. De novo evolution

After docking program, we can modify ligands by two methods (shown in Fig. 9). The first method is based on active site features to identify functional groups that can establish strong interactions with the receptor. Then, the functional groups can be linked or attached to the original ligand scaffolds. The second method uses

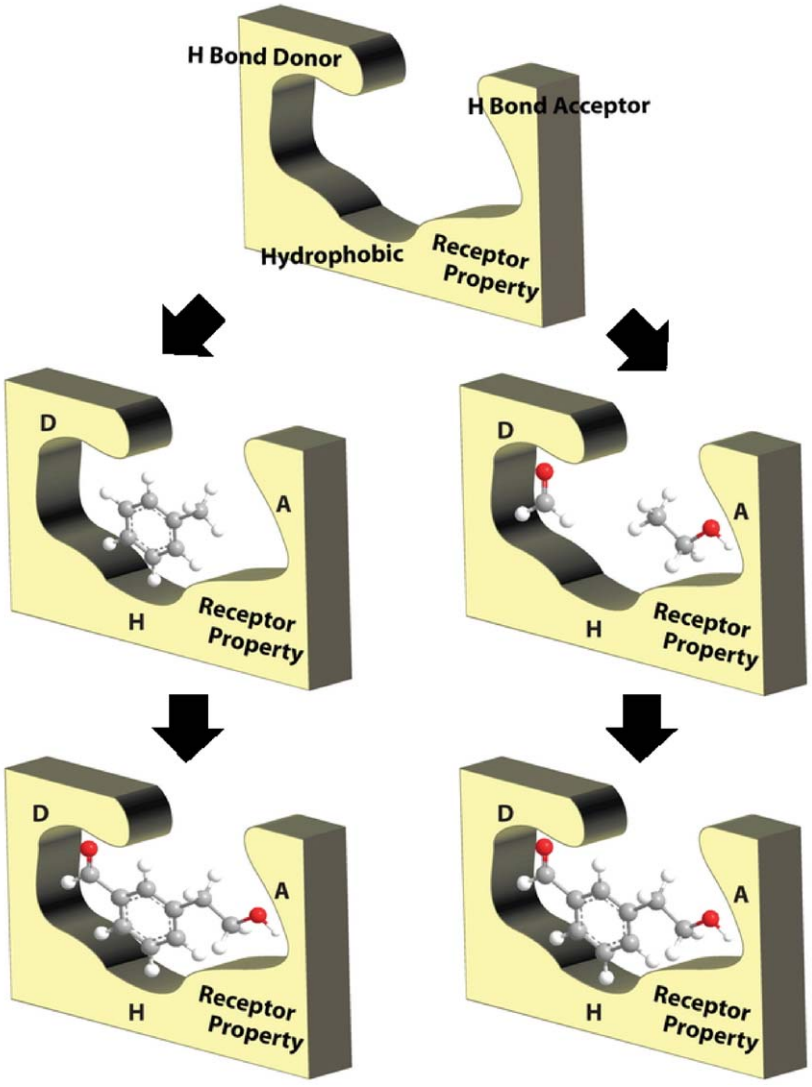

Fig. 9. The concept of *De novo* evolution.

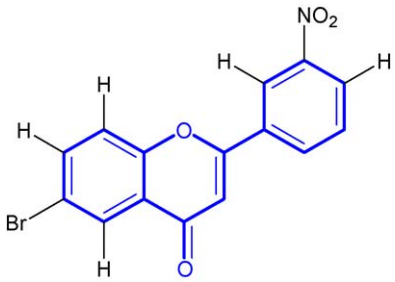

Fig. 10. The core atom (blue) of the training set used for designing GABA receptor inhibitor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

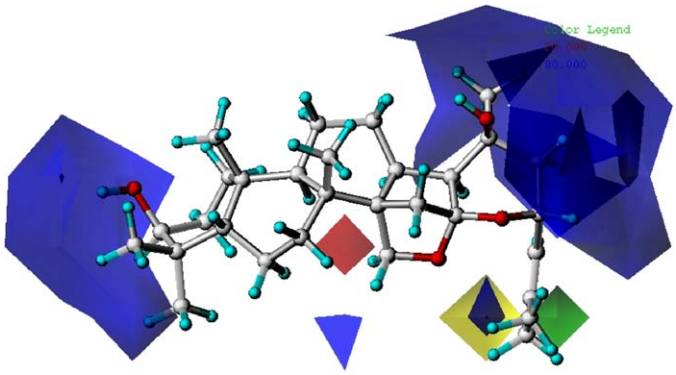

Fig. 11. CoMFA contour map, with steric favor region in green and disfavor region in yellow. The electropositive contribution is in blue and electronegative region in red (Chen, 2009a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the original ligand scaffolds to develop derivatives that can complement the receptor.

3. Ligand-based drug design

When the target protein structure is unknown or cannot be predicted by modeling techniques, the ligand-based drug design is the alternative protocol. This method uses statistical approaches to correlate ligand activity to structural information (Singer and William, 1967).

3.1. Quantitative structure–activity relationship (QSAR)

Quantitative structure–activity relationship is a widely used technique in drug designing process. It employs statistics and analytical tools to investigate the relationship between the structures of ligands and their corresponding effects. Hence, mathematical models are built based on structural parameters to describe this structure–activity relationship. Before, 2D-QSAR was widely used to link structural property descriptors (such as hydrophobicity, steric, electrostatic and geometric effects) to molecular biological activity; the results were often analyzed with multiple regression analysis. One of the most commonly used 2D-QSAR methods was proposed by Hansch (Clayton and Purcell, 1969; Hansch, 1969). However, because 2D-QSAR cannot accurately describe the correlation between the 3D spatial arrangement of the physiochemical properties, and the biological activities, recently 3D-QSAR approaches have been adapted.

In the past, we had used QSAR for drug design research, including GABA-A (Chen, 2009a) and mPGES-1 (Chen, 2009c). Using the case of GABA-A receptor as an example, the alignment core for the training set is shown in Fig. 10 and the contour maps from CoMFA and CoMSIA are shown in Fig. 11 and Fig. 12 separately.

In here, we will describe two frequently applied 3D-QSAR methodologies: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA).

3.1.1. CoMFA

Comparative molecular field analysis (CoMFA) is established on the concept that the biological activity of a molecule is dependent

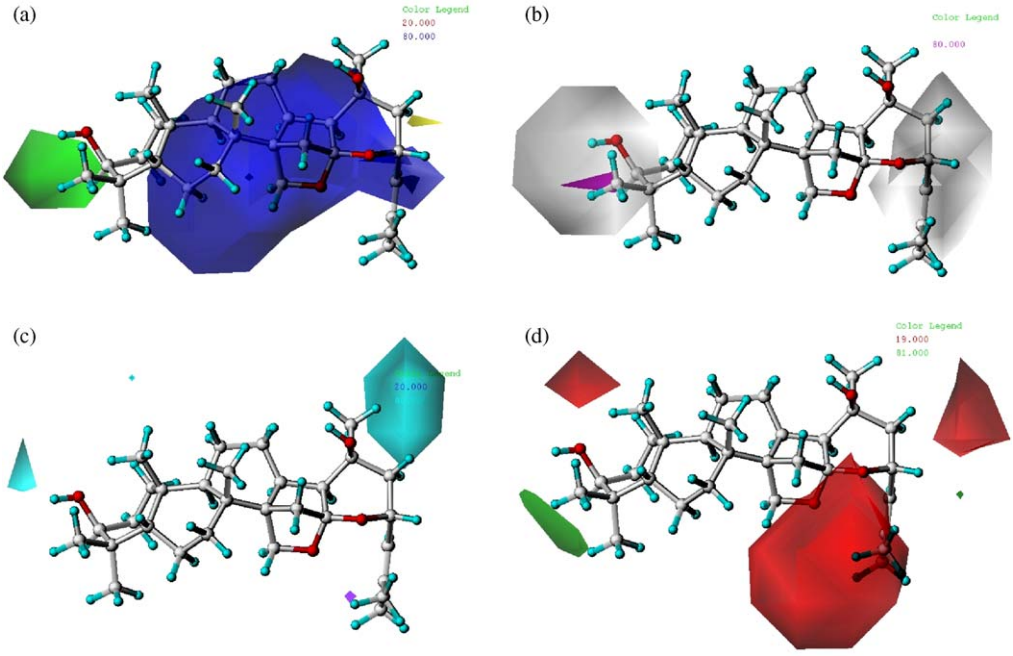

Fig. 12. CoMSIA contour maps. (a) Steric region: favor (green) and disfavor (yellow). Electropositive favored region (blue). (b) Hydrophobic region: favor (purple) and disfavor (white). (c) Hydrogen bond donor region: favor (cyan) and disfavor (purple). (d) Hydrogen bond acceptor: favor (green) and disfavor (red) (Chen, 2009a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

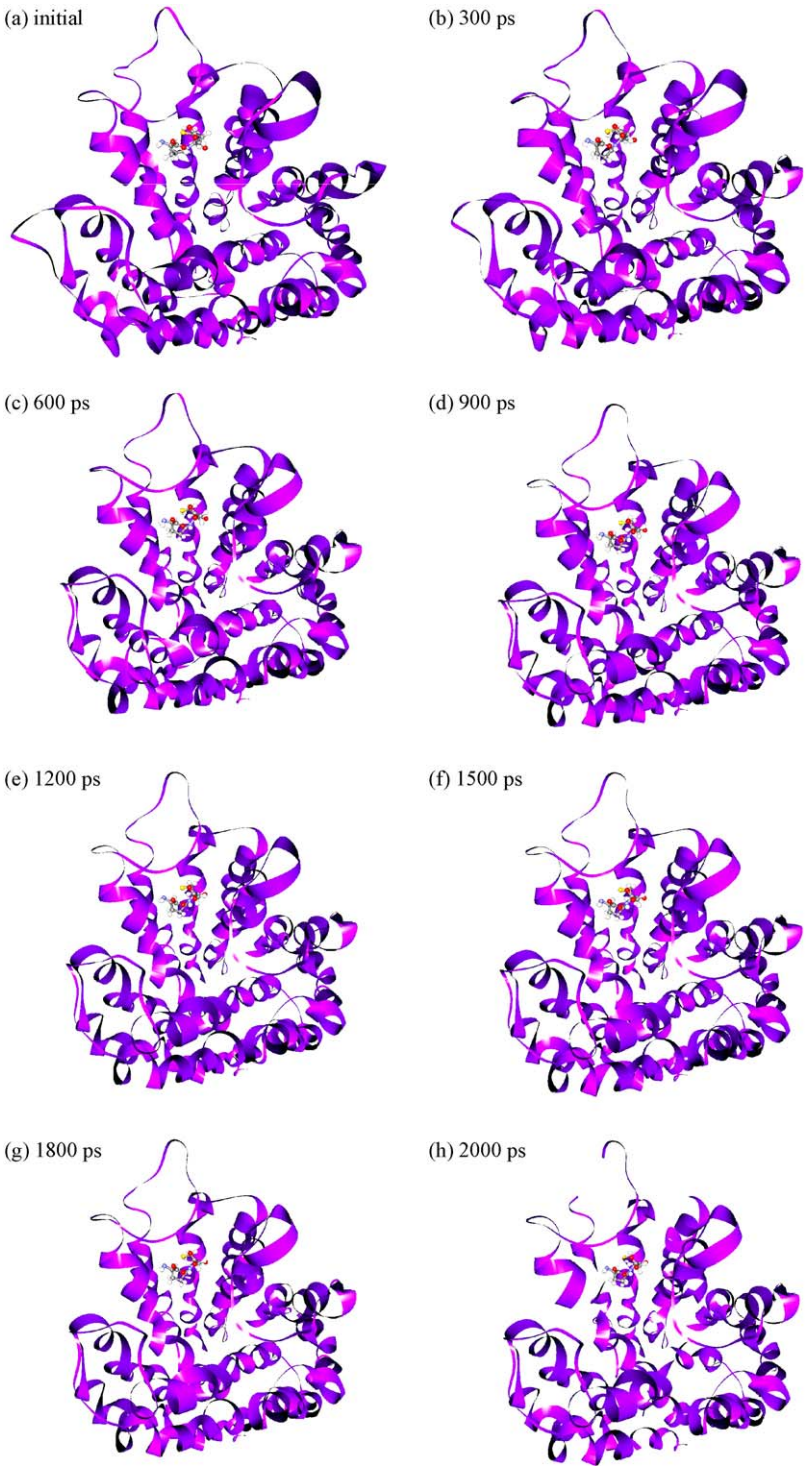

Fig. 13. Standard dynamics simulation of mPGES-1. Snapshots taken at (a) initial conformation, (b) 300 ps, (c) 600 ps, (d) 900 ps, (e) 1200 ps, (f) 1500 ps, (g) 1800 ps, and (h) 2000 ps. The protein is quite dynamic and has loop movement and fluctuation in structures. However no large movement in protein backbone is observed in the simulation time.

of the surrounding molecular fields, such as steric and electrostatic fields. The steric and electrostatic fields were calculated by CoMFA using Lennard–Jones potential, and coulombic potential, respectively. Although this method has been widely adopted, it has several problems. Both potential functions changes dramatically near the van der Waals surface of the molecule and thus, cut-off values are often required. In addition, alignment of ligands must be conducted before energy calculation, but the orientation of the superimposed molecules is correlative to the calculation grid. It could cause large changes in CoMFA results. Moreover, in order to examine both fields in the same PLS analysis, a scaling factor needs to be added to the steric field (Cramer *et al.*, 1989).

3.1.2. CoMSIA

Comparative molecular similarity index analysis (CoMSIA) is a method developed recently as an extension of CoMFA. The CoMSIA method includes more additional field properties; these are: steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen

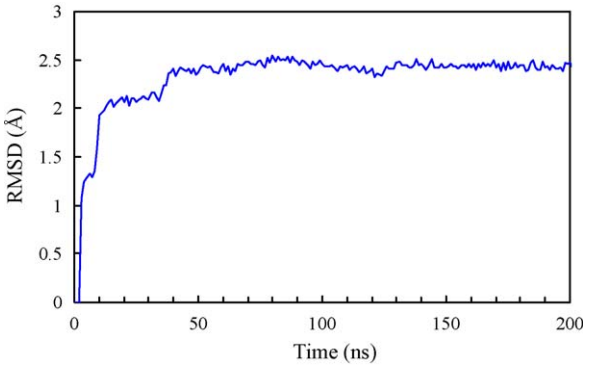

Fig. 14. Root mean square deviation (RMSD) of M2 influenza proton channel as a function of the simulation time. The root mean square deviation (RMSD) of M2 shows that the protein gradually adapts and stabilizes at a configuration after 100 ns of simulation.

bond acceptor. CoMSIA is insensitive to the orientation of the aligned molecules and correlates to the grid by using Gaussian function. Furthermore, the improved function algorithm is least influenced by the relative distance to the van der Waals surface. Overall, this model can offer a more accurate structural–activity relationship than CoMFA (Klebe *et al.*, 1994).

4. Molecular dynamics simulations

Molecular dynamics (MD) simulation is one of the important tools in the theoretical study of biological molecules. Because molecular systems generally contain a large number of particles, it is impossible to analyze such complex systems. By using numerical methods, molecular dynamics simulation can avoid such analytic intractability.

During simulation, atoms and molecules are allowed to interact for a period of time. The motion for every atom is calculated and can be played to examine the overall behavior (Mccammon *et al.*, 1977). Overall, the background algorithm for a MD simulation includes: (1) the determination of the initial positions and velocities of every atom; (2) the calculation of forces applied on the investigated atom using inter-atomic potentials; (3) the progression of atomic positions and velocities through a short-time period. These new positions and velocities are then turned into new inputs to step 2, and when steps 2 and 3 are repeated, each repetition forms an additional time step.

Table 1
The lecture of computer-aided drug design.

Lesson	Lecture description
1	Introduction of modeling software
2	Protein structure prediction
3	Structure-based drug design
4	Docking
5	Virtual screening
6	Pharmacology and molecular simulations
7	Molecular dynamics
8	Protein folding prediction
9	Mid-term examination
10	Ligand-based drug design
11	CoMFA
12	CoMSIA
13	HypoGen
14	Scoring function
15	Chinese herb database and NCI database
16	Weight rules (Chen's weight rules and equation) (I) [Chen, 2009d]
17	Weight rules (Chen's weight rules and equation) (II) [Chen, 2009d]
18	Final examination

Table 2
The lecture of structure bioinformatics.

Lesson	Lecture description
1	Protein structure
2	Protein structure prediction
3	Homology modeling
4	Force fields
5	Folding recognition
6	Hot spot and binding site
7	Programming for binding affinity
8	*Ab initio* protein modeling
9	Perl programming
10	Protein–protein interaction
11	Modeling for lipid bilayer
12	Docking
13	Flexible docking
14	Binding free energy
15	3D-QSAR
16	NCI database
17	Traditional Chinese medicine database
18	Final examination

Molecular dynamics is now routinely employed to study the structure, dynamics and thermodynamics of biological molecules and their complexes. It provides detailed information on the fluctuations and conformational changes of proteins. Nucleic acids structural information can be investigated using this method as well (Borkar *et al.*, 2010; Roy and Thakur, 2010). In addition, solvent molecules can also be investigated on the impacts of overall protein structural changes.

In our studies, we are especially interested at studying the entering of drug molecule into target protein and the associated protein–ligand interaction. In the case of mPGES-1, we have applied molecular dynamics to study structural changes after binding of target to ligand (Fig. 13). An RMSD graph of M2 influenza proton channel is also shown (Fig. 14) to illustrate the difference between the protein at a specific time and the initial reference.

5. Sample course syllabuses

An introductory syllabus on CADD, including structure-based drug design, ligand-based drug design and molecular dynamics, is shown in Table 1. The syllabus for structural bioinformatics is shown in Table 2, and this course emphasizes on physics and molecular simulation algorithms. Moreover, syllabus of the principle and application of molecular simulation is shown in Table 3.

Table 3
The lecture of molecular simulation.

Lesson	Lecture description
1	An overview of molecular simulation
2	Monte Carlo methods
3	Free energy
4	Free-energy calculations
5	An overview of molecular dynamics
6	Force fields: AMBER, CHARMM, OPLS, GROMOS (I)
7	Application of force fields
8	Effects of solvents
9	Algorithms and computations
10	Choosing the time step
11	The Lennard–Jones potential
12	FENE potential
13	EAM potential
14	Potential for covalent carbon
15	Simulated annealing
16	Softwares (AMBER, CHARMM, VASP (DFT), XMD, CPMD)
17	Group presentation
18	Final examination

6. Conclusion

From the aforementioned introduction, it is easy to see that molecular simulation has a vital role in drug design and CADD, whether it is in protein modeling, in docking or in molecular dynamics. In addition to these, we hope our flexibility concept can greatly increase the hit rate and the accuracy of protein–ligand interaction. This concept is different from our previous weight equation of which requires IC_{50} to obtain the parameter in the algorithm. By introducing flexibility in docking protocol, we hope that the simulation can be more close to real life events. With the advancement in computing facilities and software algorithms, many simulation works that require supercomputer in the past can be done in a workstation. By implementing molecular simulation into biomolecular researches, not only the research steps can be accelerated, but also the vast investment in money can be saved. In the future, molecular simulation and computer-aided drug design can greatly influence the development of pharmaceutical industry and become a necessity before molecular experiments.

Acknowledgements

The research was supported by grants from the National Science Council of China (NSC 98-2221-E-039-007-) and China Medical University (CMU98-CT-15, CMU97-276) and Asia University (CMU98-ASIA-09). This study is also supported in part by Taiwan Department of Health Clinical Trial and Research Center of Excellence (DOH99-TD-B-111-004) and Taiwan Department of Health Cancer Research Center of Excellence (DOH99-TD-C-111-005). We are grateful to the National Center of High-performance Computing for computer time and facilities.

References

Adrian-Scotto, M. and D. Vasilescu, "Quantum Molecular Modeling of Glycyl-Adenylate," *J. Biomol. Struct. Dyn.*, **25**, 697 (2008).

Akten, E. D., S. Cansu, and P. Doruker, "A Docking Study Using Atomistic Conformers Generated Via Elastic Network Model for Cyclosporin A/Cyclophilin A Complex," *J. Biomol. Struct. Dyn.*, **27**, 13 (2009).

Al-Lazikani, B., J. Jung, Z. Xiang, and B. Honig, "Protein Structure Prediction," *Curr. Opin. Chem. Biol.*, **5**, 51 (2001).

Allen, K. N., C. R. Bellamacina, X. Ding, C. J. Jeffery, C. Mattos, G. A. Petsko, and D. Ringe, "An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins," *J. Phys. Chem.*, **100**, 2605 (1996).

Baxter, C. A., C. W. Murray, D. E. Clark, D. R. Westhead, and M. D. Eldridge, "Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity," *Proteins*, **33**, 367 (1998).

Borkar, A., I. Ghosh, and D. Bhattacharyya, "Structure and Dynamics of Double Helical DNA in Torsion Angle Hyperspace: A Molecular Mechanics Approach," *J. Biomol. Struct. Dyn.*, **27**, 695 (2010).

Braun, G. H., D. M. Jorge, H. P. Ramos, R. M. Alves, V. B. da Silva, S. Giuliatti, S. V. Sampaio, C. A. Taft, and C. H. Silva, "Molecular Dynamics, Flexible Docking, Virtual Screening, Admet Predictions, and Molecular Interaction Field Studies to Design Novel Potential Mao-B Inhibitors,," *J. Biomol. Struct. Dyn.*, **25**, 347 (2008).

Brenke, R., D. Kozakov, G. Y. Chuang, D. Beglov, D. Hall, M. R. Landon, C. Mattos, and S. Vajda, "Fragment-based Identification of Druggable 'Hot Spots' of Proteins Using Fourier Domain Correlation Techniques," *Bioinformatics*, **25**, 621 (2009).

Chen, C. Y. C., "Discovery of Novel Inhibitors for c-Met by Virtual Screening and Pharmacophore Analysis," *J. Chin. Inst. Chem. Engrs.*, **39**, 617 (2008a).

Chen, C. Y. C., "A Novel Perspective on Designing the Inhibitor of HER2 Receptor," *J. Chin. Inst. Chem. Engrs.*, **39**, 291 (2008b).

Chen, C. Y. C., "Insights into the Suanzaoren Mechanism—From Constructing the 3D Structure of GABA-a Receptor to Its Binding Interaction Analysis," *J. Chin. Inst. Chem. Engrs.*, **39**, 663 (2008c).

Chen, C. Y. C., "Chemoinformatics and Pharmacoinformatics Approach for Exploring the GABA-A Agonist from Chinese Herb Suanzaoren," *J. Taiwan Inst. Chem. Engrs.*, **40**, 36 (2009a).

Chen, C. Y. C., "*De Novo* Design of Novel Selective COX-2 Inhibitors: From Virtual Screening to Pharmacophore Analysis," *J. Taiwan Inst. Chem. Engrs.*, **40**, 55 (2009b).

Chen, C. Y. C., "Pharmacoinformatics Approach for mPGES-1 in Anti-inflammation by 3D-QSAR Pharmacophore Mapping," *J. Taiwan Inst. Chem. Engrs.*, **40**, 155 (2009c).

Chen, C. Y. C., "Weighted Equation and Rules—A Novel Concept for Evaluating Protein–Ligand Interaction," *J. Biomol. Struct. Dyn.*, **27**, 271 (2009d).

Chen, C. Y. C., "Virtual Screening and Drug Design for PDE-5 Receptor from Traditional Chinese Medicine Database," *J. Biomol. Struct. Dyn.*, **27**, 627 (2010a).

Chen, C. Y. C., "Bioinformatics, Chemoinformatics, and Pharmainformatics Analysis of HER2/HSP90 Dual-targeted Inhibitors," *J. Taiwan Inst. Chem. Engrs.*, **41**, 143 (2010b).

Chen, C. Y. C. and K. T. Chen, "Novel selective inhibitors of hydroxyxanthone derivatives for human cyclooxygenase-2," *Acta Pharmacol. Sin.*, **28**, 2027 (2007).

Chen, C. Y. C., G. W. Chen, and Y. C. Chen, "Molecular Simulation: The Mechanism of HER2/neu Degradation by Inhibiting HSP90," *J. Chin. Chem. Soc.*, **55**, 297 (2008a).

Chen, C. Y. C., Y. F. Chen, C. H. Wu, and H. Y. Tsai, "What is the Effective Component in Suanzaoren Decoction for Curing Insomnia? Discovery by Virtual Screening and Molecular Dynamic Simulation. *J. Biomol. Struct. Dyn.*, **26**, 57 (2008b).

Chen, C. Y., Y. H. Cheng, D. T. Bau, H. J. Huang, F. J. Tsai, C. H. Tsai, and C. Y. C. Chen, "Discovery of potent inhibitors for phosphodiesterase 5 by virtual screening and pharmacophore analysis," *Acta Pharmacol. Sin.*, **30**, 1186 (2009).

Chen, C. Y., H. J. Huang, F. J. Tsai, and C. Y. C. Chem, "Drug Design for Influenza A Virus Subtype H1N1," *J. Taiwan Inst. Chem. Engrs.*, **41**, 8 (2010).

Clayton, J. M. and W. P. Purcell, "Hansch and Free-Wilson Analyses of Inhibitory Potencies of Some 1-Decyl-3-carbamoylpiperidines against Butyrylcholinesterase and Comparison of the Two Methods," *J. Med. Chem.*, **12**, 1087 (1969).

Deepa, P. and P. Kolandaivel, "Studies on Tautomeric Forms of Guanine–Cytosine Base Pairs of Nucleic Acids and Their Interactions with Water Molecules," *J. Biomol. Struct. Dyn.*, **25**, 733 (2008).

Ding, Q., L. Huo, J. Y. Yang, W. Xia, Y. Wei, Y. Liao, C. J. Chang, Y. Yang, C. C. Lai, D. F. Lee, C. J. Yen, Y. J. Rita Chen, J. M. Hsu, H. P. Kuo, C. Y. Lin, F. J. Tsai, L. Y. Li, C. H. Tsai, and M. C. Hung, "Down-regulation of Myeloid Cell Leukemia-1 through Inhibiting Erk/Pin 1 Pathway by Sorafenib Facilitates Chemosensitization in Breast Cancer," *Cancer Res.*, **68**, 6109 (2008).

Goodsell, D. S. and A. J. Olson, "Automated Docking of Substrates to Proteins by Simulated Annealing," *Proteins*, **8**, 195 (1990).

Hansch, C. A., "Quantitative Approach to Biochemical Structure–Activity Relationships," *Acc. Chem. Res.*, **2**, 232 (1969).

Huang, E. S., P. Koehl, M. Levitt, R. V. Pappu, and J. W. Ponder, "Accuracy of Side-chain Prediction upon Near-native Protein Backbones Generated by Ab Initio Folding Methods," *Proteins*, **33**, 204 (1998).

Im, W., M. Lee, and C. Brooks, "Generalized Born Model with a Simple Smoothing Function," *J. Comput. Chem.*, **24**, 1691 (2003).

Josa, D., E. F. da Cunha, T. C. Ramalho, T. C. Souza, and M. S. Caetano, "Homology Modeling of Wild-type, D516v, and H526l Mycobacterium Tuberculosis RNA Polymerase and Their Molecular Docking Study with Inhibitors," *J. Biomol. Struct. Dyn.*, **25**, 373 (2008).

Klebe, G., U. Abraham, and T. Mietzner, "Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity," *J. Med. Chem.*, **37**, 4130 (1994).

Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, "Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models," *Acc. Chem. Res.*, **33**, 889 (2000).

Kontoyianni, M., G. S. Sokol, and L. M. McClellan, "Evaluation of Library Ranking Efficacy in Virtual Screening," *J. Comput. Chem.*, **26**, 11 (2005).

Koska, J., V. Z. Spassov, A. J. Maynard, L. Yan, N. Austin, P. K. Flook, and C. M. Venkatachalam, "Fully Automated Molecular Mechanics Based Induced Fit Protein–Ligand Docking Method," *J. Chem. Inf. Model.*, **48**, 1965 (2008).

Landon, M. R., R. L. Lieberman, Q. Q. Hoang, S. Ju, J. M. Caaveiro, S. D. Orwig, D. Kozakov, R. Brenke, G. Y. Chuang, D. Beglov, S. Vajda, G. A. Petsko, and D. Ringe, "Detection of Ligand Binding Hot Spots on Protein Surfaces via Fragment-based Methods: Application to DJ-1 and Glucocerebrosidase," *J. Comput. Aided Mol. Des.*, **23**, 491 (2009).

Lin, Y. J., Y. C. Hou, C. H. Lin, Y. A. Hsu, J. C. Sheu, C. H. Lai, B. H. Chen, P. D. Lee, L. Wan, and F. J. Tsai, "Puerariae radix isoflavones and their metabolites inhibit growth and induce apoptosis inbreast cancer cells," *Biochem. Biophys. Res. Commun.*, **378**, 683 (2009).

Luthy, R., J. U. Bowie, and D. Eisenberg, "Assessment of Protein Models with Three-dimensional Profiles," *Nature*, **356**, 83 (1992).

Marti-Renom, M. A., A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali, "Comparative Protein Structure Modeling of Genes and Genomes," *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291 (2000).

Mattos, C., C. R. Bellamacina, E. Peisach, A. Pereira, D. Vitkup, G. A. Petsko, and D. Ringe, "Multiple Solvent Crystal Structures: Probing Binding Sites, Plasticity and Hydration," *J. Mol. Biol.*, **357**, 1471 (2006).

Mccammon, J. A., B. R. Gelin, and M. Karplus, "Dynamics of Folded Proteins," *Nature*, **267**, 585 (1977).

Mishra, S., "Function Prediction of Rv0079, a Hypothetical Mycobacterium Tuberculosis Dosr Regulon Protein," *J. Biomol. Struct. Dyn.*, **27**, 283 (2009).

Mohan, S. S., J. J. Perry, N. Poulose, B. G. Nair, and G. Anilkumar, "Homology Modeling of Glut4, an Insulin Regulated Facilitated Glucose Transporter and Docking Studies with ATP and Its Inhibitors," *J. Biomol. Struct. Dyn.*, **26**, 455 (2009).

Payne, M. C., M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, "Iterative Minimization Techniques for Ab Initio Total-energy Calculations: Molecular Dynamics and Conjugate Gradients," *Rev. Mod. Phys.*, **64**, 1045 (1992).

Rajakrishnan, V., V. R. Manoj, and G. Subba Rao, "Computer-aided, Rational Design of a Potent and Selective Small Peptide Inhibitor of Cyclooxygenase 2 (Cox2)," *J. Biomol. Struct. Dyn.*, **25**, 535 (2008).

Ramalho, T. C., M. S. Caetano, E. F. da Cunha, T. C. Souza, and M. V. Rocha, "Construction and Assessment of Reaction Models of Class I Epsp Synthase: Molecular Docking and Density Functional Theoretical Calculations," *J. Biomol. Struct. Dyn.*, **27**, 195 (2009).

Roy, S. and A. R. Thakur, "20ns Molecular Dynamics Simulation of the Antennapedia Homeodomain–DNA Complex: Water Interaction and DNA Structure Analysis," *J. Biomol. Struct. Dyn.*, **27**, 443 (2010).

Sheu, J. C., C. H. Hua, L. Wan, Y. J. Lin, H. C. Tseng, N. Jinawath, M. H. Tsai, N. W. Chang, C. F. Lin, C. C. Lin, L. J. Hsieh, T. L. Wang, I. M. Shih, and F. J. Tsai, "Functional genomic analysis identified EGFR activation as the most common genetic event in oral squamous cell carcinoma," *Cancer Res.*, **69**, 2568 (2009).

Singer, J. A. and P. William, "Purcell Relationships among Current Quantitative Structure–Activity Models," *J. Med. Chem.*, **10**, 1000 (1967).

Sippl, M. J., "Recognition of Errors in Three-dimensional Structures of Proteins," *Proteins*, **17**, 355 (1993).

Spassov, V. Z., L. Yan, and S. Szalma, "Introducing an Implicit Membrane in Generalized Born/Solvent Accessibility Continuum Solvent Models," *J. Phys. Chem. B*, **106**, 8726 (2002).

Sujatha, K., A. Mahalakshmi, D. K. Solaiman, and R. Shenbagarathai, "Sequence Analysis, Structure Prediction, and Functional Validation of Phac1/Phac2 Genes of *Pseudomonas* sp. Ldc-25 and Its Importance in Polyhydroxyalkanoate Accumulation," *J. Biomol. Struct. Dyn.*, **26**, 771 (2009).

Venkatachalam, C. M., X. Jiang, T. Oldfield, and M. Waldman, "Ligandfit: A Novel Method for the Shape-directed Rapid Docking of Ligands to Protein Active Sites," *J. Mol. Graph. Modell.*, **21**, 289 (2003).

Warren, G. L., C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head, "A Critical Assessment of Docking Programs and Scoring Functions," *J. Med. Chem.*, **49**, 5912 (2006).

Wu, G. S., D. H. Robertson, C. L. Brooks, and M. Vieth, "Detailed Analysis of Grid-based Molecular Docking: A Case Study of CDOCKER—A CHARMM-based MD Docking Algorithm," *J. Comput. Chem.*, **24**, 1549 (2003).