Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 23 (2013) 68 – 77

Procedia Computer Science

4th International Conference on Computational Systems-Biology and Bioinformatics, CSBio2013

Ligand Path: A Software Tool for Mapping Dynamic Ligand Migration Channel Networks

Tu-Liang Lin and Hong-Yi Chang*

Department of Management Information System, National Chiayi University, 580 Sinmin Rd., Chiayi City 600, Taiwan

Abstract

Proteins are essential compositions of the living organisms and involved in the processes of different life events. Basically proteins are like amazing tiny bio-machines performing the functions in a stable and predictable manner and understanding the underline mechanisms can facilitate the pharmaceutical development. However, protein functions are not carried in a static style, so experimental observations of these dynamic movements of the drugs inside the proteins are difficult, so computational methods have an important and irreplaceable role.

We developed a software tool called LigandPath for mapping the ligand migration channels in a constantly moving protein and this software can function with CADD (Computer aided drug design) software to map the possible migration pathways of candidate drugs inside a protein. Traditionally, biologists use MD (Molecular Dynamics) simulation to locate the ligand migration channels, but it takes long time for them to observe the complete migration paths. In order to overcome the limitations of the trajectory-based MD simulation, we adopt a computational method inspired from robotic motion planning called DyME (Dynamic Map Ensemble) and we develop the software tool LigandPath based on DyME. The software tool has already been successfully applied to map the potential migration channels of drugs candidates of three proteins, PPAR (peroxisome proliferator-activated receptors), UROD (uroporphyrinogen decarboxylase) and Sirt1 (silent information regulator 1) complexes in three publications.

Keywords: ligand migration; MD simulation; motion planning; protein dynamics

* Corresponding author. Tel.: +886-5-2732892; fax: +886-5-2732893.
E-mail address: tuliang@mail.ncyu.edu.tw; hychang@mail.ncyu.edu.tw

doi:10.1016/j.procs.2013.10.010

1. Introduction

Proteins are the basic functional units of all living entities. Understanding the functional mechanisms of the protein can facilitate the bioscience development and medical advancement. Many proteins function via the binding of a ligand, such as the binding of substrate and cofactors in enzyme catalysis process. For many proteins, such as myoglobin, the binding of its ligand site lies inside the protein and there is no obvious route can pass to the internal binding site. It is generally believed that the ligand can go through a path opened due to dynamic disturbance, but exactly how the proteins exert the control is not yet entirely clear. A number of experimental observations through methods such as: time-resolved Laue X-ray diffraction can observe very transient phenomenon. For a longer time observation, molecular dynamics simulation is usually required to observe the full ligand migration path network. Therefore, it is essential to develop new computational tools. The new computational tools will have significant impacts in the CADD (computer aided drug design) research field. When a new drug is filtered out, knowing how the drug enters the protein, and thus reaches the binding sites is also important. Through mapping the drug migration network in a dynamic protein, the researchers can further predict how easy or how difficult it is for the drug to reach the binding site.

In the early 70s, Case and Karplus adopted MD (molecular dynamics) simulation to study the ligand binding of myoglobin[1], and after then, a series of studies have been done, such as Elber and Karplus[2]. So far, MD simulation is still the most widely used research method in finding the ligand migration paths. Nutt and Meuwly utilized classical Classical MD and QM-MD to discover ligand migration paths and the results are accordance with the infrared spectrum observations[3]. Hummer *et al*. used both MD simulations and time-resolved Laue X-ray diffraction to observe ligand migration path, and they found the results of these two methods are consistent[4]. Among the MD simulation studies, the longest simulation in finding the ligand migration path was done by Ruscio *et al*., and the overall simulation time is 7μs[5].

In our previous study, we observed that the ligand migration in a dynamic protein is very much alike the mobile robot navigation in a dynamic environment. Inspired by this observation, we developed a motion planning based computational method called DyME (Dynamic Map Ensemble), this computational method can compute the possible ligand migration paths of a dynamic protein in a very efficient manner[6].

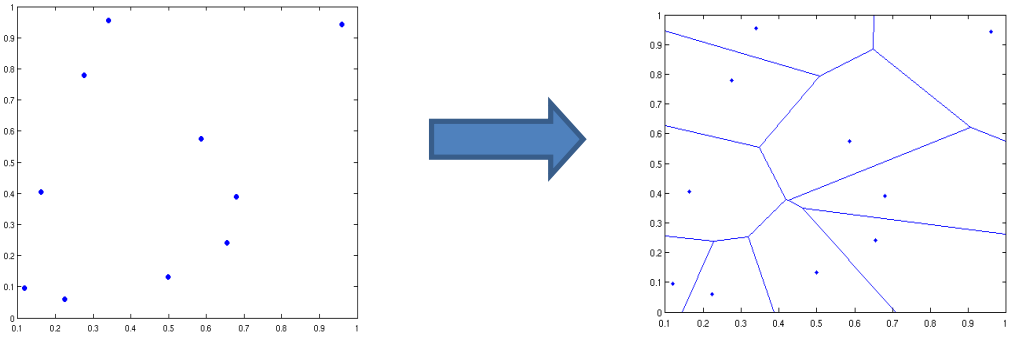

Fig 1: 2D Voronoi diagram

Many robotics methods have been applied to study different biological problems, such as protein loop closure[7, 8], the movement of protein backbone[9], the protein flexibility and structural sampling[10,11] and so on. In particular, robotic motion planning has been successfully applied to protein dynamics[12-14]. Our previous DyME[6] approach can be regarded as one of the methods inspired from robotics.

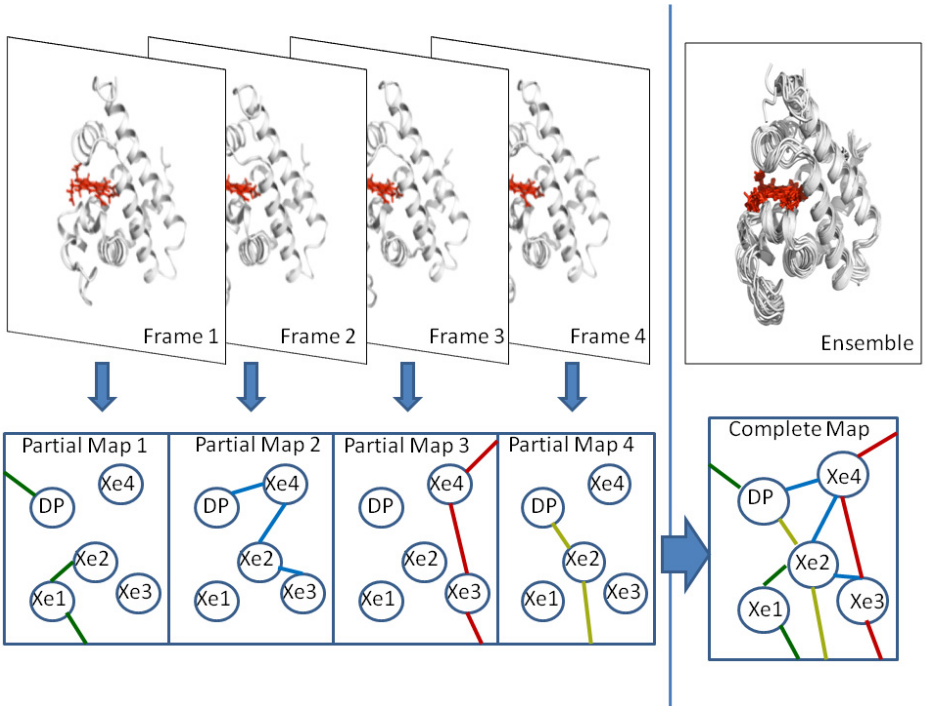

Fig 2: The schematic view of DyME (reproduced from [6])

DyME can be seemed as a general approach to obtain the ligand migration channels of a dynamic protein because it identifies the internal cavities automatically. However, in most of the cases, ligand only binds to a specific site and normally the binding site can be identified. For example, in CADD (computer aided drug design), the drug candidates and their potential binding sites can be determined in the screening process, so the binding site is already known, but how a candidate drug reach the binding site is unknown. Therefore, in this research, we simplified the original DyME and allow the users to put the ligand into the input pdb file, so the new software tool can have the position of the binding site before mapping the free space and we call this new software tool LigandPath. We also make this software tool freely available to the public and the software can be downloaded from http://ligandpath.mis.ncyu.edu.tw.

2. Methods

2.1. DyME

DyME adopt Voronoi diagram to map free space of a single protein structure, and then construct a maximum clearance roadmap. Figure 1 shows a two-dimensional Voronoi diagram example. Given the coordinates of the atoms, Voronoi diagram divided the space into polyhedra and each polyhedron contains an atom so that the distance of any point to the atom in the same polyhedron is smaller than the distance of the point to the atoms in other polyhedra

Figure 2 shows the schematic view of the DyME method, first we will have a protein structure ensemble as the input and a protein structure ensemble is a set of protein structures and these structures can come from different experiments or even computational method, such as: X-Ray, NMR structure or trajectories of MD simulations. DyME then maps the free space of each single structure. Since each single structure is a representation of a stationary state, we call the free space of a single structure partial map. After the free spaces of all structures are

mapped, DyME then integrate all partial maps into a complete map, so that a complete map can truly represent the possible ligand migration paths of a "dynamic" protein.

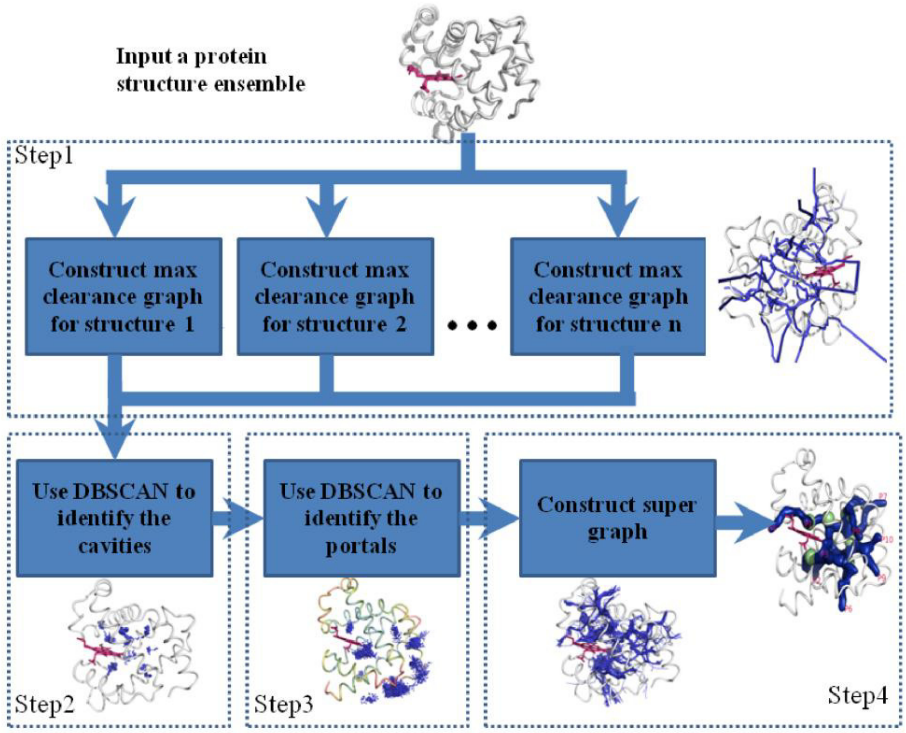

Fig 3: Flow chart of DyME (reproduced from [6])

Figure 3 shows the process of DyME method. DyME consists of four steps. In the first step, DyME construct the max clearance graph for each single structure and this part can be processed in parallel. In the second step, DyME employed DBSCAN (Density-Based Spatial Clustering of Applications with Noise[15]) to identify the cavities. In this step, two assumptions were made. The first assumption is that there is a small distance between the cavity center and the solvent which can allow us to separate the internal cavities and surface pockets. The second assumption is that the internal cavities appear more frequently than normal channels in a dynamic protein, so when all max clearance graphs superimposed together, we will have different densities in the channel and cavities such that DBSCAN can be used to identify the internal cavities. The third step employs DBSCAN to find the surface portals. The fourth step integrates the max clearance graphs of the whole protein ensemble based on the identified cavities and portals to become a super graph, so the final super graph can represent the max clearance paths available when the protein dynamics is considered.

2.2. Design of LigandPath

The DyME automatically identifies the possible cavities inside a protein using the internal free space. In order to correctly identify the cavities, two assumptions were made. The first assumption is that a cavity is at least some distance away from the solvent. This helps us to distinguish a cavity from a surface pocket. The second assumption is that the cavities should appear more frequently and with higher clearance than the channels connecting them. The second assumption allows us to distinguish cavities from open channels. These two assumptions were purely for automating the cavity identification process. However, for candidate drugs identified using CADD software, the

binding site information is available, so the cavity identification process in DyME is redundant for CADD related research. Therefore, a simplified DyME for CADD research is developed in this paper. The developed software tool is called LigandPath. LigandPath is also equipped with a better user interface and accept more MD trajectory formats. Figure 4 shows design of the LigandPath.

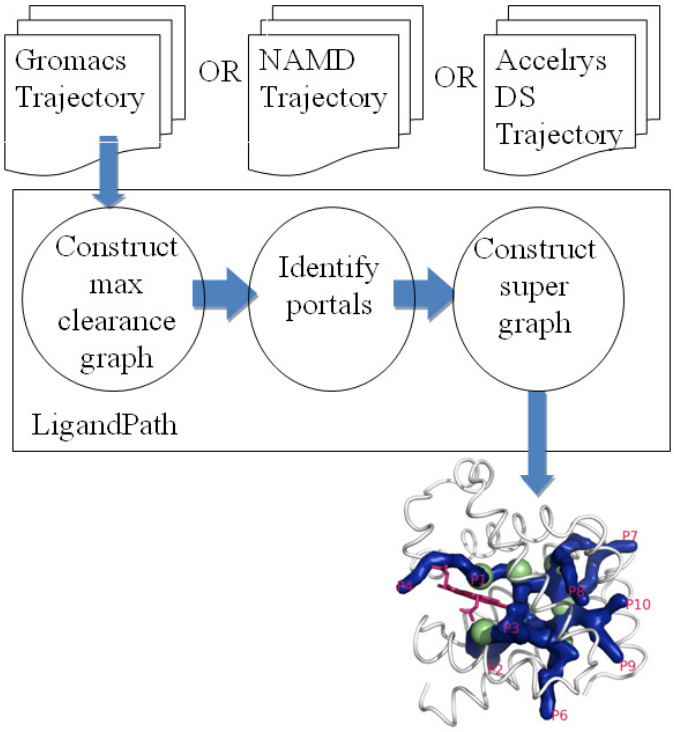

Fig 4 Design of LigandPath

Table 1 shows the comparison between the DyMe and LigandPath. The functionality of auto cavity identification is removed in LigandPath,so LigandPath has better performance than DyME. LigandPath also supports more MD trajectory formats and is public available.

Table 1 Comparison between DyME and LigandPath

	DyME	LigandPath
Auto cavity identification	Yes	No
Better performance	No	Yes
Supporting MD trajectory formats	NAMD	NAMD、Gromacs、Accelrys DS
Integrated with CADD	No	Yes
Public available	No	Yes

2.3. The Efficiency of DyME and LigandPath

When comparing trajectory-based MD methods with DyME/LigandPath, most trajectory-based MD methods explicitly simulate the ligand and each simulation runs tens or perhaps even hundreds of nanoseconds to find one ligand migration pathway. Multiple simulations have to be executed in order to find more pathways. However,

DyME/LigandPath does not explicitly follow the ligand movements, so the requirement for the coordination between the ligand motion and the protein fluctuation through random diffusion is removed. In DyME/LigandPath, we only need to examine the dynamics of protein long enough to see the transient channels open once and this is many times shorter than the simulation time needed to see a ligand actually traversing through these channels. In previous research[6], we use a 10-ns MD simulation to generate a ligand migration channel graph that is comparable to a cumulative 7-μs trajectory-based MD simulation.

3. Results and Discussions

The LigandPath program has been employed to identify the potential drug migration channel networks of peroxisome proliferator-activated receptors (PPAR) in [16], uroporphyrinogen decarboxylase (UROD) in [17] and silent information regulator 1 (Sirt1) complexes in [18]. Conformations from the MD trajectories form the protein structural ensembles which encoded the dynamic information over a given period of time. Pathways computed using the first 5 ns conformations were assigned as entries and those computed from the last 5 ns conformations were designated as exits.

3.1. Ligand pathways of PPARs

LigandPath has been successfully used to obtain the potential ligand migration pathways of Peroxisome proliferator-activated receptors (PPARs) in [16]. PPARs are essential regulators of adipogenesis, lipid and carbohydrate metabolism, and are potential drug targets for metabolic syndromes. There are three isoforms: PPAR-α, PPAR-γ, and PPAR-δ. Two ligands, (S)-Tryptophan-betaxanthin and berberrubine, have higher docking scores than the control compound and also show stable interactions during MD simulations[16].

The MD simulation trajectories were used to map the possible migration pathways. Figure 5 shows the pathways of the control compound and the screened ligands in three isoforms, PPAR-α, PPAR-δ, and PPAR-γ, respectively. LigandPath maps the free space inside the protein at every conformation during the first 5 ns and the last 5 ns of MD simulations, respectively. Within each PPAR isoforms, similar portals and pathways are identified for different ligands. The different paths in the first 5 ns and the last 5 ns of MD simulations may be due to the possible effect of each ligand docking.

Table 2 shows the running time, memory requirement and runtime environment of LigandPath for a 5ns MD trajectory (2000 frames) of PPAR. It takes 6 hours to complete the whole run on a PC with Intel Core i5 CPU and 4 GB memory. The running time can be affected by the size of the protein and large proteins normally take more time to compute. Since the programs are written in Matlab, the Matlab Compiler Runtime environment is needed for executing the program.

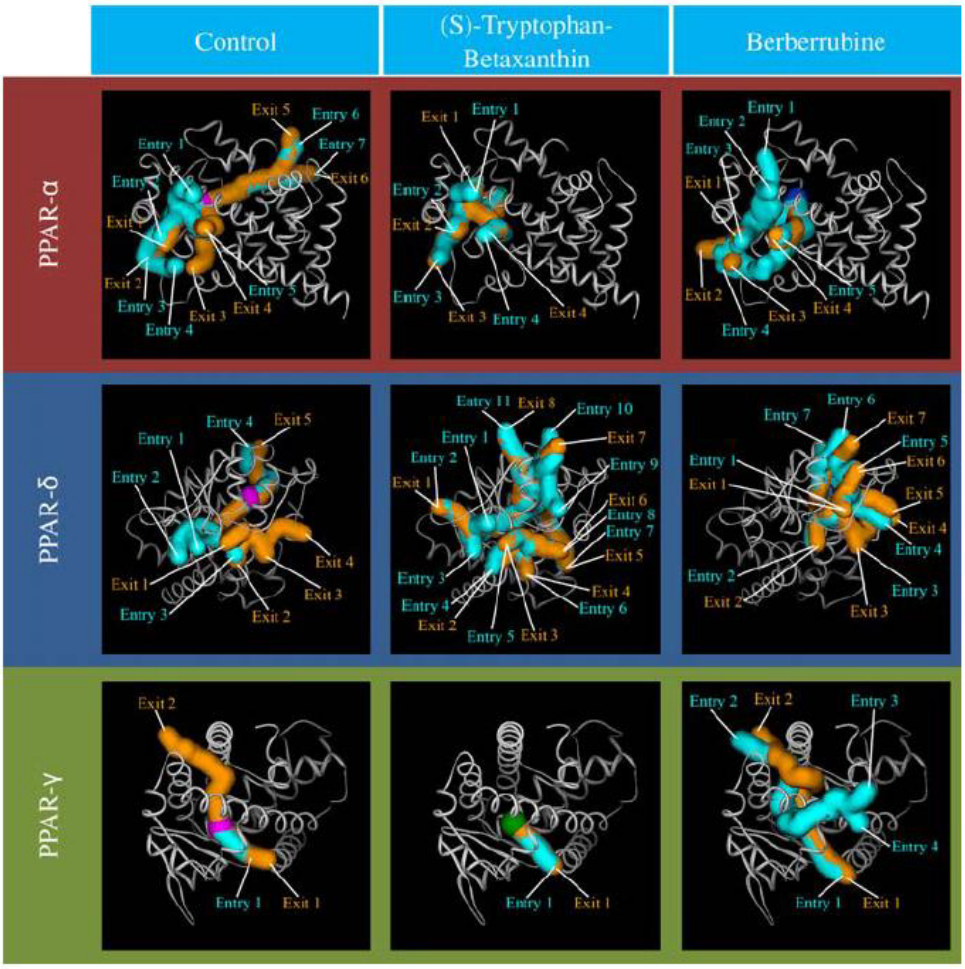

Fig 5 Ligand pathways for PPAR-α, PPAR-δ, and PPAR-γ. Paths and portals are shown in cyan and orange for first 5 ns and last 5 ns, respectively (reproduced from [16])

Table 2 Running time, memory requirement and runtime environment of LigandPath for a 5ns MD trajectory (2000 frames) of PPAR

	LigandPath
Running time	6 hrs 45 mins
Memory requirement	4 GB
Runtime environment	Matlab Compiler Runtime V7.14

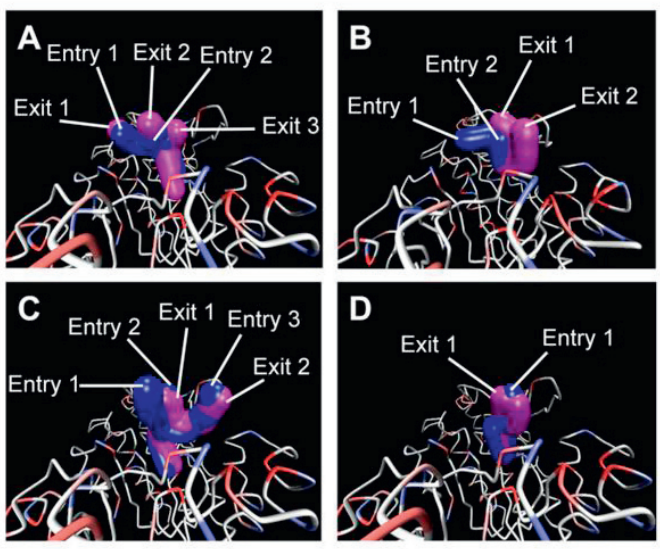

Fig 6. Snapshots of simulated channels for selected ligands to the binding site in UROD. Liagnd migration channels for (A) Coproporphyrinogen III (B) isopraeroside IV (C) scopolin (D) nodakenin are showed. Pathways calculated from MD confirmations during 0–5 ns are marked in blue and assigned as entry pathways. Pathways calculated from MD confirmations during the last 5 ns are marked in magenta and assigned as exit pathways (reproduced from [17])

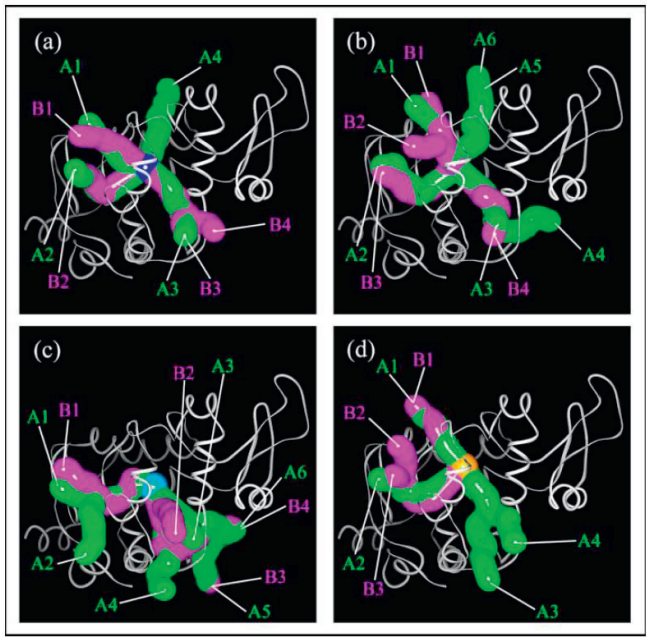

Fig 7 Ligand pathways for Sirt1 complexes with (a) cambinol, (b) (S)-tryptophan-betaxanthin, (c) 5-O-feruloylquinic acid, and (d) RosA. Notes: Paths and portals are shown in green and magenta for first 5 ns (A) and last 5 ns (B), respectively (reproduced from [18])

3.2. Ligand pathways of UROD

LigandPath has been successfully used to obtain the potential ligand migration pathways of Uroporphyrinogen decarboxylase (UROD) in [17]. UROD has been suggested as a tumor-selective protectant for head and neck cancer (HNC) against radiation. Four ligands, Coproporphyrinogen III, isopraeroside IV, scopolin and nodakenin are sceened to have possible inhibitory effects on UROD [17]. Figure 6 illustrates that all CADD screened ligands might be able to access to the designated binding site through "entry" pathways and "exit" pathways based on UROD conformations obtained from the initial and final 5 ns of MD simulation.

3.3. Ligand pathways of Sirt1 complexes

LigandPath has been successfully used to obtain the potential ligand migration pathways of Silent information regulator 1 (Sirt1) complexes in [18]. Sirt1, one member of the sirtuin family, play an Important role in metabolic disease and diseases associated with aging. (S)-tryptophan-betaxanthin, cambinol, 5-O-feruloylquinic acid, and RosA were identified as potential drugs to bind Sirt1 protein [18]. Figure 7 shows the possible migration channels for each protein–ligand complex. The free space inside the protein from LigandPath at the first 5 ns and last 5 ns of MD simulation illustrates the possible docking effect of screened ligands and control compound on Sirt1 protein. For Cambinol and (S)-tryptophan-betaxanthin, the similar trajectories of A4 for Cambinol and A5 and A6 for (S)-tryptophan-betaxanthin disappeared after MD simulation. Paths A3 and A4 in RosA also disappeared after MD simulation. Cross examinations of all possible migration channels show that A2 is the stable pathway for all ligands in Sirt1 protein and B2 for Cambinol, A2 and B3 for (S)-tryptophan-betaxanthin, A1 and B1 for 5-O-feruloylquinic acid, and A2 and B3 for RosA.

4. Conclusion

In this paper, we develop a software tool called LigandPath for mapping the ligand migration pathways in a dynamic protein. LigandPath can cooperate with the CADD software to map the possible migration pathways of candidate drugs. LigandPath has been successfully applied to map the potential drug migration channels of three proteins, peroxisome proliferator-activated receptors (PPAR) [16], uroporphyrinogen decarboxylase (UROD) [17] and silent information regulator 1 (Sirt1) complexes [18], in three publications. Potential drug migration pathways of the three proteins were identified and some of the potential pathways exhibit stable behaviours which can be further verified. Understanding the drug migration channel is important because it might affect the percentage of the drugs that actually reach the binding site and the percentage will indirectly influence the effectiveness of the potential drugs.

Acknowledgements

This research was supported by grants from the National Science Council of Taiwan (NSC101-2218-E-415-002-MY2 and NSC102-2221-E-415-012-).

References

1. Case DA and Karplus M. Dynamics of ligand binding to heme proteins. *J Mol Biol* 1979; **132**:343-68.
2. Elber R and Karplus M. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J Am Chem Soc* 1990; **112**:9161-9175.
3. Nutt DR and Meuwly M. CO migration in native and mutant myoglobin: atomistic simulations for the understanding of protein function. *Proc Natl Acad Sci U S A* 2004; **101**:5998-6002.
4. Hummer G, Schotte F, and Anfinrud PA. Unveiling functional protein motions with picosecond x-ray crystallography and molecular dynamics simulations. *Proc Natl Acad Sci U S A* 2004; **101**:15330-4.
5. Ruscio JZ, Kumar D, Shukla M, Prisant MG, Murali TM, and Onufriev AV. Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *Proc Natl Acad Sci U S A* 2008; **105**:9204-9.

6. Lin TL and Song G. Efficient mapping of ligand migration channel networks in dynamic proteins. *Proteins* 2011; **79**:2475-90.
7. Canutescu AA and Dunbrack RL. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Science* 2003; **12**:963-972.
8. Kolodny R, Guibas L, Levitt M, and Koehl P. Inverse kinematics in biology: The protein loop closure problem. *International Journal of Robotics Research* 2005; **24**:151-163.
9. Noonan K, O'Brien D, and Snoeyink J. Probik: Protein backbone motion by inverse kinematics. *International Journal of Robotics Research* 2005; **24**:971-982.
10. Shehu A, Clementi C, and Kavraki LE. Modeling protein conformational ensembles: from missing loops to equilibrium fluctuations. *Proteins* 2006; **65**:164-79.
11. Shehu A, Kavraki LE, and Clementi C. Multiscale characterization of protein conformational ensembles. *Proteins* 2009; **76**:837-51.
12. Amato NM and Song G. Using motion planning to study protein folding pathways. *Journal of Computational Biology* 2002; **9**:149-168.
13. Apaydin MS. *Stochastic roadmap simulation: an efficient representation and algorithm for analyzing molecular motion*, PhD, Stanford University, Stanford, CA, 2004.
14. Song G. *A Motion Planning Approach to Protein Folding*, PhD, Texas A&M University, College Station, TX, 2003.
15. Ester M, Kriegel HP, Sander J, and Xu XW. "A density-based algorithm for discovering clusters in large spatial databases with noise," in *the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996, pp. 226-231.
16. Chen KC, Chang SS, Huang HJ, Lin TL, Wu YJ, and Chen CY. Three-in-one agonists for PPAR-alpha, PPAR-gamma, and PPAR-delta from traditional Chinese medicine. *J Biomol Struct Dyn* 2012; **30**:662-83.
17. Tsou YA, Chen KC, Lin HC, Chang SS, and Chen CY. Uroporphyrinogen decarboxylase as a potential target for specific components of traditional Chinese medicine: a virtual screening and molecular dynamics study. *PLoS One* 2012; **7**:e50087.
18. Chen KC, Jian YR, Sun MF, Chang TT, Lee CC, and Chen CY. Investigation of silent information regulator 1 (Sirt1) agonists from Traditional Chinese Medicine. *J Biomol Struct Dyn* 2012.