September 30, 2012

Virtual Screening for High Specificity Inhibitors of SSH-2 Against Dual Specificity Phosphatases of Similar Homology

Joshua Wei, Jason Haga

Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA
Cybermedia Center, Osaka University, Osaka 565-0871, Japan

Abstract

Dual Specificity Phosphatases (DSPs) are a subject of interest because of their ability as a protein tyrosine phosphatase group that can dephosphorylate phospho-tyrosine and phospho-serine/threonine. While relatively little is understood about them, their regulatory balance has been linked to the development of cancers, Alzheimer's disease, diabetes, and obesity. In order to rank compounds that bind specifically to a DSP called Slingshot Homolog 2 (SSH2), we use the program DOCK6 to test the docking energies for tens of thousands of ligand-protein combinations against two other proteins in the SSH family. A previously generated list of the top 1% of SSH2 binding ligands was found by screening the ZINC database with DOCK6, reducing number of possible candidates to 20,000. This top 1% is docked against SSH1 and SSH3 so that their binding energies can be compared to that of SSH2, determining which bind well to SSH2 but poorly to SSH1 and SSH3. Using this method across various DSPs, it is possible to discover a compound of medicinal significance. The screening process occurs *in silico* using models of DSP proteins generated in previous PRIME projects [1].

Introduction

The DSP is a subgroup of the type-I cysteine-based Protein Tyrosine Phosphatase (PTP) superfamily. This superfamily all share a characteristic, highly conserved catalytic domain which contains the sequence HCXXXXXR[2]. In this generalized motif, H is histidine, C is cysteine, R is arginine, and X is any amino acid. This group of proteins can be further split into two classes, classical PTPs and dual specificity PTPs, our area of interest. While having similar 3D structure,

the catalytic pocket of DSPs are shallow, but broader than those of classical PTPs, a supposed reason as to how DSPs are thought to be able to accommodate more than one phosphorylated residue [3].

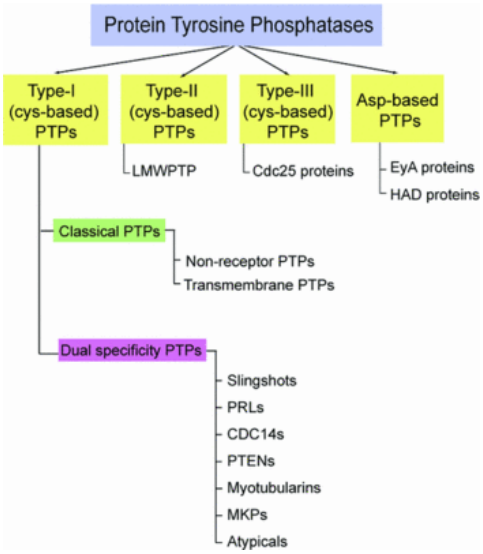

Figure 1: Classification of PTPs [2]

The SSH family is composed of 3 known variants, SSH1, SSH2, and SSH3, as found in human and mouse isoforms. They contain conserved A and B domains, though their functions remain unknown, as well as a general motif of HCXXGXXR at the active site, possibly including a C-terminal F-actin-binding region[4]. SSH2 has been of recent interest because of its role as an actin regulator through the phosphorylation of actin-depolymerizing factor (ADF) and cofilin [5]. As actin makes up the microfilaments of cells, SSH2 plays a large role by indirectly regulating the actin cytoskeletal structure of cells. LIM kinases phophorylate the serine 3 on cofilin, reversibly suspending its actin-severing ability while SSH phosphatases dephosphorylate the cofilin to balance the effects of the kinases [5]. When neurons in the brain undergo stress from peptides like amyloid-β (Aβ) 1–42, over expression of SSH prevents the phosphorylation of cofilin, resulting in rod-shaped actin bundles [6],[7]. Understanding the role of SSH2 in actin suppression grows increasingly necessary as recent efforts have linked Alzheimer's disease to the ectopic expression of these actin rods and cofilin [6]. By finding specific inhibitors of SSH2, dephosphorylation of cofilin can be repressed, making headway in finding the cure for Alzeheimer's. When determining a drug candidate, specificity is key as other DSPs could be affected due to similar homology.

The program MODELLER is a 3D modeling software that uses spatial restraints of a known template x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy to

generate missing DSP models and refine existing one, allowing for the calculation of the binding energies of a ligand database [8]. The DOCK6 program for protein-ligand simulation and scoring was used in conjunction with the 3D DSP structures to find inhibitors of SSH2 that show significantly less affinity to the homologically similar structures like SSH1 and SSH3.

Method

In the experiment, MODELLER and DOCK6 ran on the supercomputer grid, ranking the binding affinities of potential inhibitors of SSH1 and SSH3. The grid was made possible by PRAGMA, the Pacific Rim Applications and Grid Middleware Assembly. Models generated by previous PRIME projects were evaluated and used in the screening process [1]. Based on the satisfaction of the spatial restraints through a similar DSP template, in this case SSH2, MODELLER employed its homology or comparative modeling method to calculate the 3D model for SSH1 and SSH3 [8]. Evaluations of the models have found them to have sound stereochemistry and as accurate to the crystallographic structures as the template structures if not more so [8]. Discrepancies are often due to misalignment of the template and a known sequence identity or a template structure than simply does not match the correct structure [8]. Identifying unstable high energy loops can also be identified and refined to reduce the energy of the model and stabilizing it [1]. This function on MODELLER is called loopmodel, but is basically a localized version of MODELLER, lowering the energy of the loop and providing a DOPE energy profile for such refinements. Given such functions, MODELLER is a viable solution for generating models to run through DOCK6. Besides loop

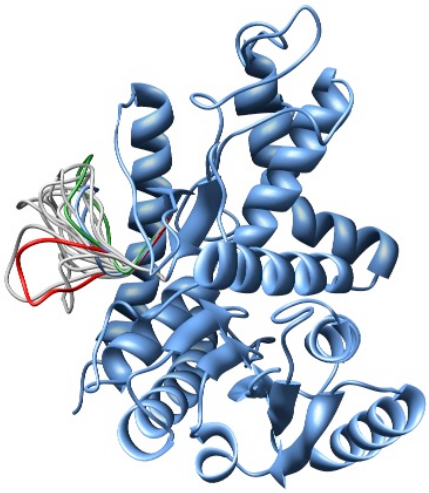

Figure 2: MODELLER scoring different orientations through loopmodel [8]

refinement, the energy minimization tool on the model visual aid and editing program CHIMERA was used. This tool includes two algorithms to calculate the energy minimization, steepest descent and conjugate gradient. While the steepest descent algorithm proved effective at the default 100 steps, we found our models were most stable around 75 steps for conjugate gradient. The accuracy of our models were verified with the multi-purpose online structure validation program, MolProbity.

If the model structures differ from known x-ray crystallography structures, new models can be recalculated on MODELLER with more careful alignment. The models used for SSH1 and SSH3 were actually remodeled entirely as the previous models had loops too large to be fixed by loopmodel. Taking the top scoring 1% of SSH2 inhibitors found by screening the ZINC database, we began running DOCK6 on the clusters using two methods of scoring. The fast method, Grid Based Energy Screening, takes into account Van der Waals and electrostatic forces, but treats both the receptor and the ligand as frozen entities [9]. The AMBER scoring method results in more precise scoring due to simulated "induced fit" of the function, in other words, the program takes into account the changing geometry of the ligand during interaction at the expense of more computing time [9]. The summation of the Grid score and the AMBER score were used to determine the final consensus ranking of the best inhibitors of SSH1 and SSH3. Going back to compare between previously ranked potential inhibitors of SSH2 ascertained which ligands proved the most protein specific.

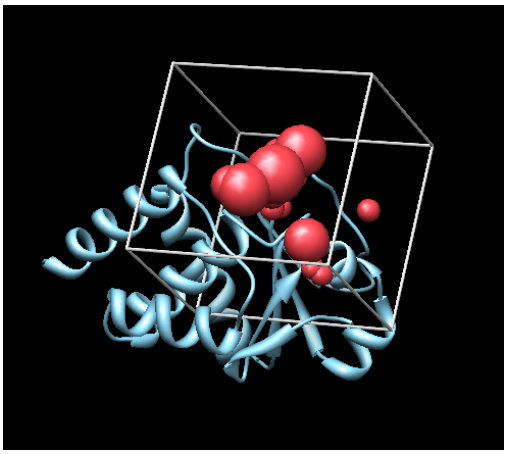

Figure 3:Typical setup of Grid Based scoring. The blue target is SSH3. The red spheres define the atom points and the volume of the active site. The ligand (not shown) will be docked and scored using the grey box, also called the grid.

Further Information

DOCK6 uses 3D protein models, orienting and docking ligands into the active site [9]. After finding the best protein-ligand fit, DOCK6 scores the binding energy and continues the process for the rest of the top ranking SSH2 inhibitors. While not the most exact calculation of binding free energy, DOCK6 ranks the compound library with respect to each other [9]. The loss of accuracy is excusable because the goal is to rank the likeliest compounds for binding specificity. Also, this significantly reduces the time to calculate the scores. Even so, screening tens of thousands of compounds would be overtaxing for a single computational unit. As a result, using the PRAGMA grid of networked supercomputers allows for parallel computing, reducing the calculation time by many times over. Approximately 20,000 ligands were distributed among 176 slices for job submission to Osaka University's Milk Cluster, which housed around 9 nodes for a total of 136 processors.

All the MODELLER and DOCK jobs were sent to the grid through the Opal OP application wrapper. Local scripts for loopmodel, Grid based docking, and AMBER based docking were already written by previous PRIME students. This year, scripts were written to sort through the results folders and package them into manageable tar files. Other scripts included an AMBER error handler and a installation script for all our necessary programs.

The whole project fit within the span of 9 weeks, with the average job time for DOCK preparation being 2 days, assuming the model was loop refined. If no errors occurred, grid based scoring averaged 4-5 days while AMBER calculation took around 1-2 weeks. Variability occurred from shared usage of the supercomputer grid and changing complexities of compound structure, making it hard to pinpoint how long a given job would take. Most of the time went into straightening out various errors as the cluster would often drop jobs or perform slowly. One

cause that was pinpointed retrospectively was that Apache Tomcat server was designed to only have one job submission at a time. Trying to balance between three jobs per cluster lead to slowdowns and failed starts. AMBER also proved to have numerous errors as many protein-ligand combinations proved incompatible with AMBER and would have to be skipped. Overall, we found the limiting factor to the speed of the experiment to be the number of nodes available as only the Milk Cluster was fast enough for the heavy calculations required.

Conclusions

Over the course of the experiment, the priority proteins of SSH1 and SSH3 were successfully docked for both Grid based scoring and AMBER scoring. DSP21 and PRL2 have been Grid docked only while DSP19 is in the process of AMBER scoring. While it was important to find a compound specific for SSH2 against all the DSPs, we focused on SSH1 and SSH3 as they are the most homologically similar. In future studies, it would be important to verify specificity against other DSPs, targeting for either SSH2 specific inhibitors or perhaps specific to the entire SSH family.

zinc id	ssh2 rank	ssh1 rank	difference	ssh3 rank	difference	dsp21 rank (dock only)	difference	dsp19 rank (dock only)	difference
ZINC05260817	1	7289	-7288	5959	-5958	17661	-17660	433	-432
ZINC03869281	2	4039	-4037	1340	-1338	7378	-7376	39	-37
ZINC04543673	3	745	-742	510	-507	4534	-4531	121	-118
ZINC02384698	4	524	-520	609	-605	5217	-5213	956	-952
ZINC03869935	5	5349	-5344	3653	-3648	16822	-16817	25	-20
ZINC04521532	6	8330	-8324	6928	-6922	3676	-3670	737	-731
ZINC04543675	7	665	-658	1233	-1226	5716	-5709	134	-127
ZINC02522549	8	1228	-1220	628	-620	409	-401	135	-127
ZINC04652516	9	44	-35	538	-529	1179	-1170	14381	-14372
ZINC02637978	10	303	-293	364	-354	105	-95	3556	-3546
ZINC01516594	11	9365	-9354	8929	-8918	4365	-4354	13213	-13202
ZINC01325418	12	111	-99	173	-161	168	-156	15801	-15789
ZINC02149821	13	9054	-9041	9473	-9460	1029	-1016	732	-719
ZINC06815633	14	70	-56	390	-376	1766	-1752	11839	-11825
ZINC03276848	15	8674	-8659	3966	-3951	3692	-3677	3266	-3251
ZINC04107594	16	1583	-1567	1451	-1435	1076	-1060	11864	-11848
ZINC06046393	17	14538	-14521	6597	-6580	11881	-11864	6430	-6413
ZINC05444608	18	5368	-5350	923	-905	17534	-17516	18420	-18402
ZINC00411161	19	1250	-1231	2977	-2958	5772	-5753		19
ZINC05373221	20	207	-187	132	-112	474	-454	14480	-14460

Table 1:Ranking of SSH2 inhibitors and their differences for various DSPs.

In the table of results, the three highlighted compounds of interest were ZINC05260817 (Ranked 1st for SSH2), ZINC02384698 (Ranked 4th for SSH2), and ZINC03869935 (Ranked 6th for SSH2). These were noted for their high specificity while maintaining a high rank for SSH2. For compounds past 7, energy scores dropped significantly, so we focused more on the high ranking compounds. It is likely, however, that as more DSPs are added to the list, these compounds will become less and less viable as drug candidates. Still, the information is progress in the right direction.

When comparing our results to similar studies, further information can be gleaned. In the screening of highest ranked inhibitors of SSH-2 against the DSP VH-1 done by Pham et al, the results show that ZINC03869935 (Ranked 6th for SSH2) does not actually show as much specificty in relation to other compounds. ZINC02384698 (Ranked 4th for SSH2) actually shows more promise, but again, when compared to other compounds like ZINC06815633, (Ranked 14th for SSH2) the difference becomes marginal. In general, the top ranked SSH-2 inhibitors tended to be smaller and showed less specificity due to their similar molecular motifs [10]. High ranking compound, 2-amino-3-phosphonooxy-propanoic acid (ZINC06815633) listed at 14th based on the energy score rankings and showed the highest specificity because of its carboxyl end and the strong H-bonds that held its end to the activation site [10]. While its binding affinity for SSH2 is certainly lower, its specificity looks sufficient enough to compensate. In our own screenings, the trend remained the same for DSP21 and DSP19, though they lacked a proper AMBER screening.

Table 2: Comparing Consensus Ranking Between Screenings of SSH-2 and VH-1 [10]

ZINC	SSH2	VH1	Difference (VH1-SSH2)
05260817	1	NA	1
03869281	2	610	608
04543673	3	857	854
02384698	4	1227	1223
03869935	5	925	920
04521532	6	403	397
04543675	7	535	528
02522549	8	704	696
04652516	9	1220	1211
02637978	10	17278	17268
01516594	11	3987	3976
01325418	12	6315	6303
02149821	13	813	800
06815633	14	19953	19939
03276848	15	12757	12742

It should be noted that ZINC06815633 (Ranked 14th for SSH2) remained fairly high ranked for SSH1 and SSH3. While this does not represent the SSH2 specific inhibitor that we originally sought, it could still be of importance as the members of the SSH family are so similar, that they play congruous roles in their effects on cofilin. Other DSPs must be AMBER docked before determining which compounds are best to be tested in wet lab as none of the data means anything until it has been proven *in vitro*. In conclusion, there are many more steps to be taken before we find a solid cure for Alzeheimer's disease, but the more knowledge we attain, the better our chances become.

References

1. Tsui, B. (2011, September 1). PRIME. Retrieved January 16, 2012. PRIME Reports 2011: http://prime.ucsd.edu/progress_reports_2011/PRIME_Final_Final/BTsui_Final_2011.pdf
2. Patterson K. I., Brummer T., O'Brien P. M., Daly R. J. (2009) Dual-specificity phosphotases: critical regulators with diverse chemical targets. *Biochem. J.* 418, 475–489. [PubMed]
3. Denu, J and Dixon, J. E.. (1998) Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr. Opin. Chem. Biol*. 2, 633-641.
4. Niwa R, Nagata-Ohashi K, Takeichi M, Mizuno K, Uemura T. (2002) Control of Actin Reorganization by Slingshot, a Family of Phosphatases That Dephosphorylate ADF/Cofilin. *Cell Press.* 108(2):233-246.
5. Soosairajah J, Maiti S, Wiggan O, et al. (2005) Interplay between components of a novel LIM kinase-slingshot phosphatase complex regulates cofilin. *EMBO Journal.* 24(3):473–486.
6. Maloney MT, Bamburg JR. (2007). Cofilin-mediated neurodegeneration in Alzheimer's disease and other amyloidopathies. *Mol Neurobiol.* 35:21–44.
7. Mendoza-Naranjo A, Contreras-Vallejos E, Henriquez DR, Otth C, Bamburg JR, Maccioni RB, and Gonzalez-Billault C. (2012) Fibrillar amyloid-β1-42 modifies actin organization affecting the cofilin phosphorylation state: a role for Rac1/cdc42 effector proteins and the slingshot phosphatase. *J Alzeheimers Dis.* 29(1), 63-77.
8. Sali A, Potterton L, Yuan F, Vlijmen H, Karplus M. (1995) Evaluation of comparative protein structure modeling by MODELLER. *Proteins* 23:318-326.
9. UCSF. (n.d.). What is DOCK? Retrieved August 28, 2011, from The Official UCSF DOCK Web-site: http://dock.compbio.ucsf.edu/Over view_of_DOCK/index.htm
10. Pham, P.D.; Levesque, M.J.; Ichikawa, K.; Date, S.; Haga, J.H. (2008) Identification of a Specific Inhibitor for the Dual-Specificity Enzyme SSH-2 via Docking Experiments on the Grid. *eScience. IEEE Fourth International Conference on* , vol., no., pp.547-554, 7-12.