Beilstein-Institut
Cracking the Sugar Code by Navigating the Glycospace
June 27th – July 1st, 2011, Potsdam, Germany

Development of User-friendly, High throughput Screening for Ligands and Inhibitors of Carbohydrate Modifying Enzymes

Hans Heindl[1], Pamela Greenwell[1]*, Tamas Kiss[2], Chris Reynolds[2], Daniel Farkas[2], Gabor Tersztyanksy[2] and Stephen Winter[2]

[1]School of Life Sciences, University of Westminster, London W1W 6UW, U.K.

[2]Centre for Parallel Computing (CPC), School of Electronics and Computing Science, University of Westminster, London W1W 6UW, U.K.

E-Mail: *greenwp@wmin.ac.uk

Received: 6th December 2011/Published: 11th July 2012

Abstract

This paper describes the impact of cloud computing and the use of GPUs on the performance of Autodock and Gromacs respectively. Cloud computing was applicable to reducing the "tail" seen in running Autodock on desktop grids and the GPU version of Gromacs showed significant improvement over the CPU version. A large (200,000 compounds) library of small molecules, seven sialic acid analogues of the putative substrate and 8000 sugar molecules were converted into pdbqt format and used to interrogate the *Trichomonas vaginalis* neuraminidase using Autodock Vina. Good binding energy was noted for some of the small molecules (~-9 kcal/mol), but the sugars bound with affinity of less than -7.6 kcal/mol. The screening of the sugar library resulted in a "top hit" with α-2,3-sialyllacto-*N*-fucopentaose III, a derivative of the sialyl Lewis^x structure and a known substrate of the enzyme. Indeed in the top 100 hits 8 were related to this structure. A comparison of Autodock Vina and Autodock 4.2 was made for the high affinity small molecules and in some cases the results were superimposable whereas in others, the match was less good. The validation of this work will require extensive "wet lab" work to determine the utility of the workflow in the prediction of potential enzyme inhibitors.

Introduction

At the Glyco-bioinformatics meeting in 2009, we reported the development of ProSim, a user-friendly workflow running on the National Grid Service (NGS), that would enable bioscientists to: model energy minimized proteins and ligands, carry out docking experiments to investigate protein-ligand binding, use molecular dynamics programs to interrogate the stability of the docked products [1 – 3]. Though successful, the methodology involved blind-docking rather than targeted docking: the latter would provide us with more useful data but requires modification of the workflow and currently, manual annotation of the grid parameters. However, such modification does allow investigation into improving the computational speed of the component programs using Graphic Processor Units (GPU), clusters, desktop grids (DG) and cloud based virtual computing (VM). Additionally, we and our collaborators could make use of the component parts of the workflow in scenarios where there is a requirement to dock a number of ligands against a single protein, reducing the need for pre-processing (energy minimisation) the receptor for each experiment. This also highlighted the need for a library of energy minimized proteins and small compounds pre-processed into the relevant pdbqt files ready for docking.

At the conference noted above, we also presented our current work which aims to 1) develop user-friendly interfaces to enable biologists to easily access complex modelling and docking programs whilst still producing high-quality results, 2) increase efficiency of the programs to enable more simulations to give robust data, 3) develop a library of small ligands for high-throughput docking experiments and 4) utilise the tool in research.

Due to progress in information technology, the use of *in silico* methods have gained significant momentum over the last few years. Computer simulations have advantages and shortcomings compared with traditional ‘‘wet lab’’ methods. Due to the increased availability of high computational power, *in silico* methods may be much cheaper than ‘‘wet lab’’ experiments which use costly reagents, may require synthesis of novel chemical compounds and need highly skilled laboratory workers. A simple *in silico* screen used to inform the ‘‘wet lab’’ experiments would significantly reduce the need to purchase, at considerable cost, a full range of compounds many of which would have little further utility. Working *in silico* can also reduce problems of health and safety involved in the handling of potentially dangerous chemicals, cells and organisms. Using web-based interfaces is relatively inexpensive, after the initial cost of development, and the programs are simple to use, although interpretation of the results requires knowledge and experience.

Current docking and validation technology allows the docking of about 50,000 small molecules to a receptor using a 30 standard-CPU cluster within a day. This is at least comparable in speed to ‘‘wet lab’’ high throughput methods. The validity of docking methods as a first screen for promising ligands has been proven in the development of the neuraminidase inhibitors Oseltamivir and Zanamivir, which relied on computer-based docking results [4].

However, the validity of simulations depends on the models used to determine chemical bonds and long range interactions between atoms (electrostatic, van der Waals). Due to limitations in computer technology, the mathematical representations of chemical interactions can only be approximations. Although the equations dominating quantum mechanics cannot be solved precisely for molecular dynamics simulations, the quality of those approximations greatly influences the results obtained. In addition, in the absence of crystal structures, models are based on homology with other known structures, which in turn may introduce bias and compromise the results.

Improving Performance

The original ProSim workflow that comprised of two programs with different computing requirements, Gromacs (energy minimization and molecular dynamics) [5] and Autodock [6] (a receptor-ligand docking), was modified to give separate workflows each with user-friendly web based interfaces. As a GPU version of Gromacs was available this was trialled on a single GPU and on clusters of GPUs [7].

The Autodock program is an ideal candidate for parallelisation as the same task is undertaken thousands or tens of thousands of times and the results co-ordinated to return the best fit molecules. Porting of Autodock to the local desktop grid infrastructure of the University of Westminster, comprising of over 1600 computers, via the P-grade portal resulted in a system that required the user to set the pdbqt files from their pdb molecule files and determine the grid space producing the relevant dpf and gpf files [8]. This was accomplished using MGL tools [9]. A web based interface was then designed to facilitate the use of the Autodock program (Fig. 1).

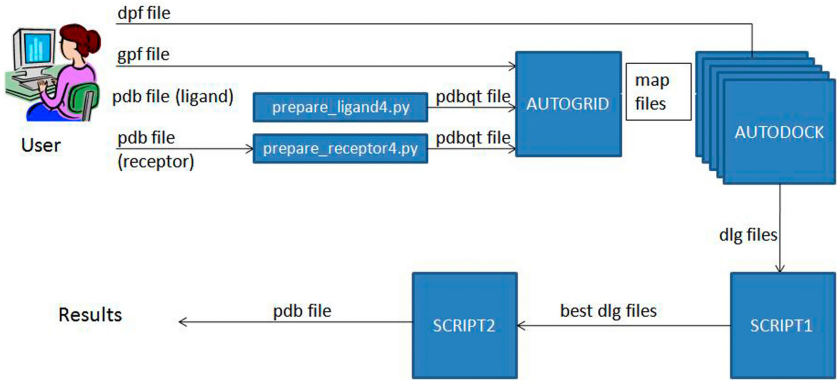


Figure 1. The Autodock workflow

One of the major drawbacks of using desktop grids (DGs) is the problem of tails, unfinished jobs at the end of the job queue that slow down completion of the workflow. To address this problem, a scheduler was constructed that detected when the jobs were entering the tail phase and then ported the remaining jobs to the Cloud using Openstack middleware. The user simply enters jobs via the portal, as previously described. Dealing with the "tail" is automatic and completely hidden from the users. When the tail is detected, jobs are replicated and Cloud jobs are instantiated automatically. The results of both cloud and desktop grid jobs are collated for return to the user (see Fig. 2) [10].

Results

A comparison of the use of a GPU versus 10 nodes of a cluster of CPUs, each with 2 processors, illustrates the increase in performance obtained with a single GPU for different stages of the Gromacs program. The single GPU decreased the workflow time by about 25% (Table 1). The greatest improvement was seen in the Mdrun6 (md1.mdp) stage, the longest of the stages tested, where the run time was reduced by 28.5%.

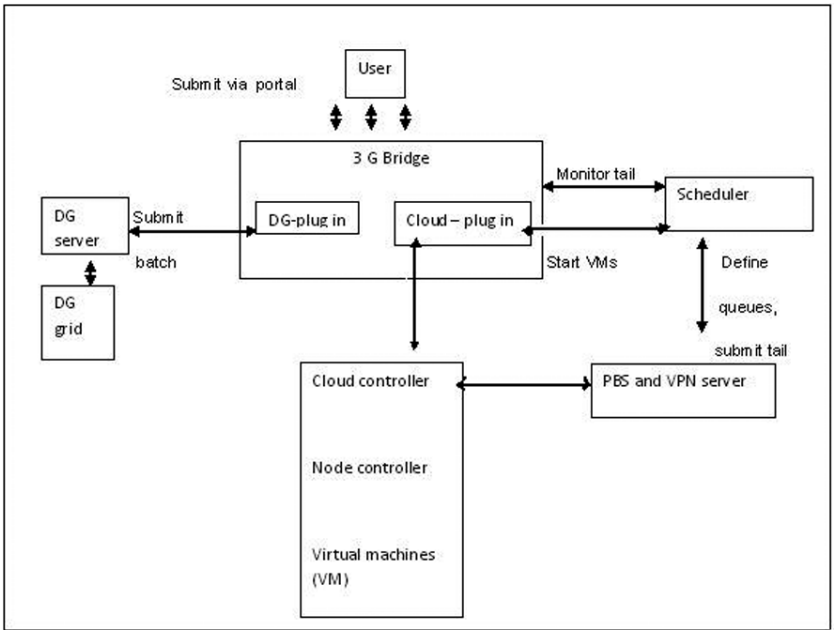


Figure 2. A schematic of the workflow submitted to the DG and VM

Submission of 1000 jobs to Autodock running on the DG resulted in a tail that was still unfinished after 60 minutes. On detection of the tail (last 5% of jobs) the remaining jobs were replicated to the Cloud and the mean percentage decrease in job time achieved using 30 CPU cores of the UoW Openstack Nova Cloud was 38% (Table 2). However, currently the

DG is a free resource whereas the use of the Cloud will impose charges, leading to choices between speed and cost. Improvements in speed themselves enable more simulations to be run in a realistic timespan improving the statistical significance and robustness of the results.

Conclusion

Cloud computing instantiation on detection of the tail significantly improved the performance of the Autodock 4 program, but the cost versus time saved would need to be considered. GPUs were also successful in improving the performance (~25%) of the molecular dynamics simulations using Gromacs using a single GPU versus 10 nodes of a cluster of CPUs, each with 2 processors. Work is now underway to investigate the utility of clusters of GPUs.

Table 1. Computational speed using Gromacs on CPU clusters (10 nodes; 2 processors/node) and a single GPU

Job	Cluster	GPU
MDrun3 (pr.mdp)	67 s	59 s
Mdrun4 (nvt.mdp)	46 s	38 s
Mdrun5 (npt.mdp)	48 s	39 s
Mdrun6 (md1.mdp)	3675 s	2626 s

Table 2. Decreasing the tail problem by utilizing "Cloud bursting". Percentage decrease in makespan using 30 cpu cores of the Cloud to process the tail defined as the last 5% of 1000 jobs.

Run1	Run 2	Run 3	Mean	SD
34.3	39.6	39.7	37.9	3.1

Development of a Small Molecular Library for Use in High-throughput Screening Using Autodock Vina

For the detection of novel ligands binding to a given protein *in silico*, there is the need to develop a library of small molecules suitable for docking experiments. This small molecule library must be in the pdbqt format, a pdb-like format with additional information on rotable bonds and on charges assigned to single atoms. This information is required to compute the energy due to internal conformational changes in the small molecule itself and the interaction between receptor and ligand. Hence we have prepared a ligand library based on the NCI database of small molecules [11]. This file (1.3 GB) was split up into sub-files (~100 files each containing about 2500 compounds). The ligands were translated from the sdf format into pdb using OpenBabel [12, 13] and converted into pdbqt format using the

prepare_ligand4.py script from MGL-Tools package (prepare_ligand4.py) [9]. The topologies of the small molecules were generated using Antechamber (Amber 11.0) and the system was solvated and the appropriate number of ions added to achieve the proper ionic strength and electroneutrality. A short (10 nsec) MD run was performed and the best 6 conformations were then ready to be used for docking.

There are constraints and requirements for the receptor files with respect to receptor size and 3D structure. It is important that the receptor (protein) is no more than 500 amino acids as, although larger receptors may be considered, molecular dynamics experiments carried out in explicit solvent can be computationally costly and the larger the molecule the more computational power required. For the docking process itself, size is not critical. A good 3-D structure of the protein in question must be available either as an X-ray crystal structure, ideally with a substrate/analogue bound to inform on the active centre, or as a structure derived from a good homology model, well informed by available crystal structures from other related proteins. The model was tested using, for example, molprobity [14] which searches for outliers in the Ramachandran plot and for clashes. An *in vacuo* energy minimization was performed before hydrogen atoms could be added using, for example, the MGL-tools package or alternatively leap (Amber) [15]. The model was then solvated in a water volume and appropriate numbers of Na^+ and Cl^- ions added. An initial energy minimization was performed followed by a position restrained molecular dynamics (MD) run which "immobilized" the protein and allowed the solvent to fill the pockets. The model was then heated up and pressurized (nvt and npt runs) and finally a 100 ns MD simulation was performed without restraining the protein. The goal was to demonstrate the stability of the protein and study possible conformational changes of the macromolecule. By performing the MD simulation the stability of the target protein and the substrate binding pocket could be assessed. Samples of this simulation (e. g. in 10 ns steps) could then be used as docking targets. The resulting pdb files were then processed by MGL-Tools into pdbqt files.

In this case Autodock Vina [16] was used to replace the conventional Autodock 4.2 program previously used due to the increased speed and suitability for use in distributed computing such as a DG. The program also incorporates an easy to use scoring function, included in the results file, enabling calculations of binding free energy. Autodock Vina uses a novel docking algorithm and performs at least 100 times faster compared to classical Autodock 4.2 (AD 4.2). The authors [17] also claim a better docking quality but this needs to be confirmed.

For the program itself, no preparational autogrid run is necessary and no gpf and dpf files are needed. This simplifies the use of the tool for novice users. The only information required for the program to commence is the coordinates of the docking space centre, the size of the docking space and the receptor and ligand pdbqt files. The necessary starting files (e. g. 10 – 100 ligand files together with the receptor file) can be wrapped into a package and sent to the appropriate computing infrastructure. In the present case the docking was performed on our local DG. A major advantage of Autodock Vina is that it can be compiled for many

different platforms and a cygwin environment, currently required to run Autodock on Windows machines, is no longer necessary. To date we have only developed this tool to run on a command line, but a web interface is in production. The current library is 200,000 small molecules (NCI) and 8000 glycans (courtesy of Prof. Dr. Thomas Lütteke, Justus-Liebig University Giessen, Germany). After the Autodock simulation, an MD run (100 ns unrestrained molecular dynamics) was performed to ensure that the final molecules were robust and hence likely to exist.

Free energy calculations such as Thermodynamical Integration (TI), Free energy Perturbation (FEP), Potential of Mean Force (PMF) and Umbrella Sampling could then be made. TI is a 'morphing' function (lambda) between two conformational states (e. g. two different bound ligands) and ΔG is found by integration over lambda. This is determined numerically over a number of MD calculations at defined lambda intermediates and allows a relative binding free energy to be computed [18]. The slow 'pullout' of a ligand from a docked position may give an idea of the binding energy by integrating the force acting on the ligand integrated over time. This method would give an easy orientational measure of the binding energy of a ligand. A calibration curve could be constructed for different systems using known (measured) free energies.

With the docking and MD computations, possible ligands or inhibitors may be found. These may then be mutated by changing the side chains on the core and the results re-scored using Autodock Vina. In concept, better 'genes' would be inherited preferentially to the next generations and crossing over and spontaneous mutations could be integrated leading to the production of more potently binding ligands.

Current Research

For the present study an energy minimised homology model of the *Trichomonas vaginalis* (TV) neuraminidase (Accession: EAY07469.1) was used as a macromolecular docking target. TV is a protozoan that infects the human urinogenital tract of 180,000 million women per year and acts as a co-factor for heterosexual transmission of HIV [19]. This neuraminidase is said to be specific for α-2,3 linked sialic acids and unable to liberate α-2,6 linked sialic acid from mucin [20]. In this paper, we present the high throughput result of docking the NCI small molecule library and the 8000 molecule sugar library against the TV neuraminidase [21]. In order to validate the results it would be necessary to run a 100 ns unrestrained molecular dynamics (MD) to assess the stability and possible conformational changes of the macromolecule.

Results

The ten best hits from the screening of the NCI library (200,000 small molecules) against the TV neuraminidase are shown in Figure 3. The figure shows how a terminal COOH perfectly fits into the arginine "bay" and which part of the receptor may accommodate the rest of the natural substrate. The molecules are well clustered around the active site of the neuraminidase and the affinity values obtained were between -8.4 and -9.4 kcal/mol (Fig. 3).

The individual results from the small molecule docking are seen in Table 3. The affinity (Kcal/mol) for the top 10 docks was between -8.4 and -9.4. Molecules 1 and 2 are highly related, one being a carbonitrile and the other an acetonitrile derivatives of 2,4,5-triphenyl-1 H-pyrrole. These derivatives were better inhibitors than molecule 9, the 2,3,4-triphenyl-1 H-pyrrole (-9.4 and -9.3 versus -8.4 kcal/mol). These molecules are related to those investigated by Mohamed and colleagues in 2009 [22] in a study on antimicrobial activity. Molecule 6, *N*-(2-chloro-4-{[(2,4-diaminoquinazolin-6-yl)methyl]amino}benzoyl) aspartic acid, has been shown to have activity against small cell lung carcinoma cell lines [23]. The RMSD was 0.00 for all ten small molecules.

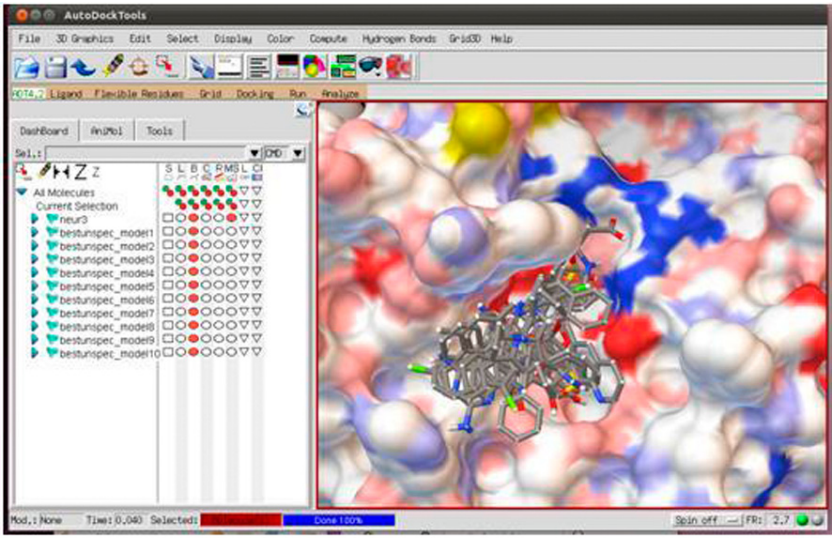


Figure 3. An overlay of the 10 best fit small molecules using TV neuraminidase as target.

The ten best results of this high throughput docking were compared to a classical Autodock 4.2 experiment (100 runs each). Some of the results showed a good agreement with the results obtained using Autodock Vina, some of them were less encouraging. The docking energy values obtained from both experiments showed a good overall correlation (Table 4). To date an equivalent high throughput experiment using Autodock 4.2 is out of reach, but

more parallel implementations of Autodock are in development (using GPUs) which will eventually enable the deployment, using modern graphics cards and the Cuda C language, of such an application on, for example, a desktop grid.

Table 3. The results for the "best fit" small molecules binding to TV neuraminidase

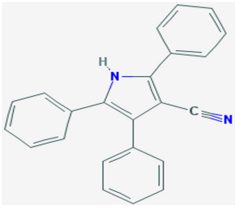
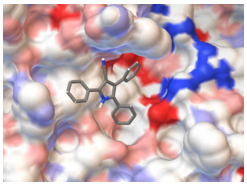
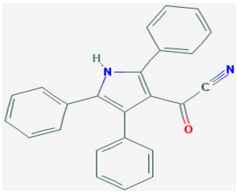
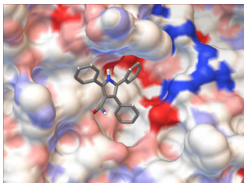
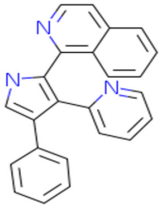
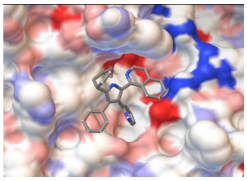
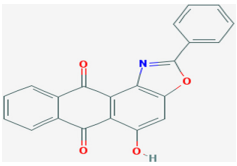
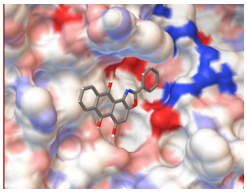
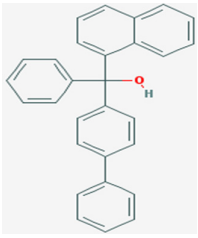
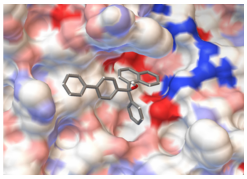
Molecule structure	Autodock result	Affinity kcal/mol	ChemSpider name Canonical SMILES code
1. 		-9.4	2,4,5-triphenyl-1 H-pyrrole-3-carbonitrile[22] N#Cc1c(c2ccccc2)c([nH]c1c1ccccc1)c1ccccc1
2. 		-9.3	oxo(2,4,5-triphenyl-1 H-pyrrol-3-yl)acetonitrile N#CC(=O)c1c(c2ccccc2)c([nH]c1c1ccccc1)-c1ccccc1
3. 		-9.2	None given c1ccc(nc1)c1c(c2ccccc2)c2n(c1c1nccc3c1ccc-c3)
4. 		-9.1	5-hydroxy-2-phenylanthra[1,2-d][1,3]oxazole-6,11-dione Oc1cc2oc(nc2c2c1C(=O)c1ccccc1-C2=O)c1ccccc1
5. 		-9.0	biphenyl-4-yl(naphthalen-1-yl)phenylmethanol O[C@@](c1cccc2c1cccc2)(c1ccccc1)c1ccc(c-c1)c1ccccc1

Table 3. continued

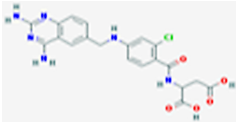
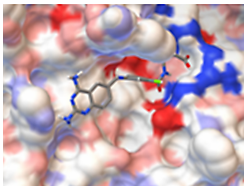
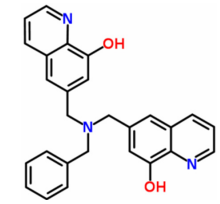
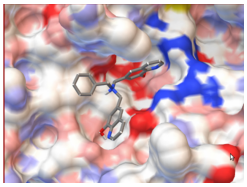
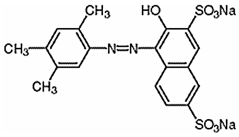
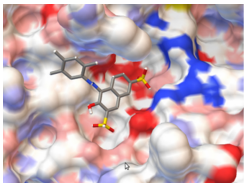
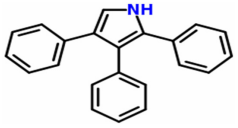
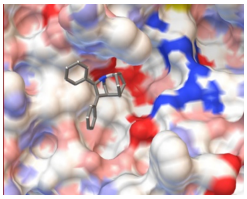
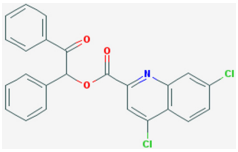
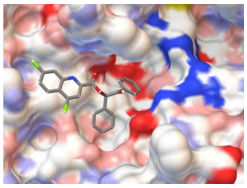
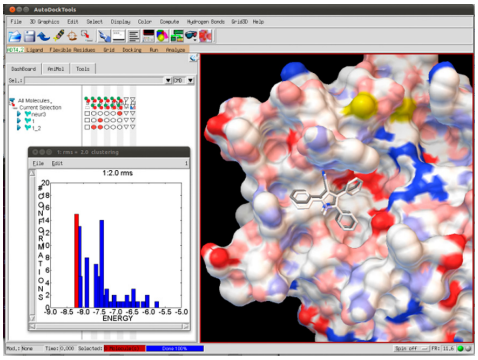
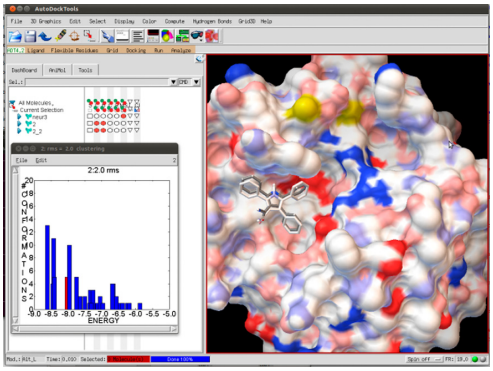
Molecule structure	Autodock result	Affinity kcal/mol	ChemSpider name Canonical SMILES code
6. 		-8.9	N-(2-chloro-4-{[(2,4-diaminoquinazolin-6-yl) methyl]amino}benzoyl)aspartic acid[23] OC(= O)C[C@@H](C(= O)O)N-C(= O)c1ccc(cc1Cl)NCc1ccc2c(c1)c(N)nc(-n2)N
7. 		-8.7	benzyl-bis[(8-hydroxy-6-quinolyl)methyl]ammonium Oc1cc(C[NH+](Cc2c-c(O)c3c(c2)cccn3)Cc2ccccc2)cc2c1nccc2
8. 		-8.6	disodium 3-oxidanylidene -4-[(2,4,5-trimethyl-phenyl) hydrazinylidene]naphthalene-2,7-disulfonate CC 1 = CC(= C(C = C 1C)NN = C 2C 3 = C(-C = C(C = C 3)S(= O)(= O)[O-])C = C(C 2 = O)S (= O)(= O)[O-)C.[Na+].[Na+]
9. 		-8.4	2,3,4-triphenyl-1 H-pyrrole c1ccc(cc1)[C@@H]1[C@H]2CN([C@@H]1-c1ccccc1)[H]1C 3C[C@H]2C = CC 13
10. 		-8.3	2-oxo-1,2-diphenylethyl 4,7-dichloroquinoline-2-carboxylate Clc1ccc2c(c1)nc(cc2Cl)C(= O)O[C@@H](-C(= O)c1ccccc1)c1ccccc1

Table 4. Overlay of the Autodock (dark coloured) and Autodock Vina (pale coloured) results for small molecules 1 – 10

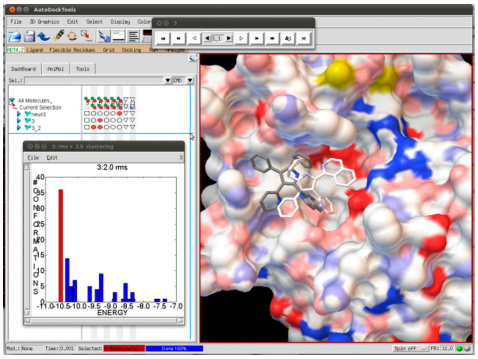
1



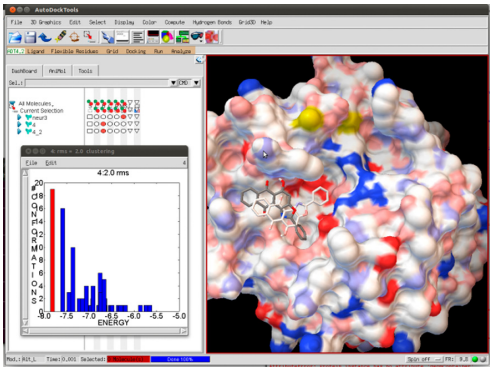
2



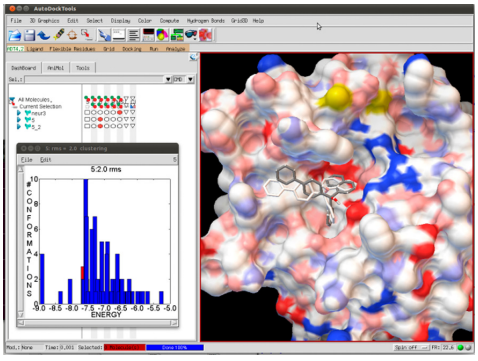
3



4



5



6

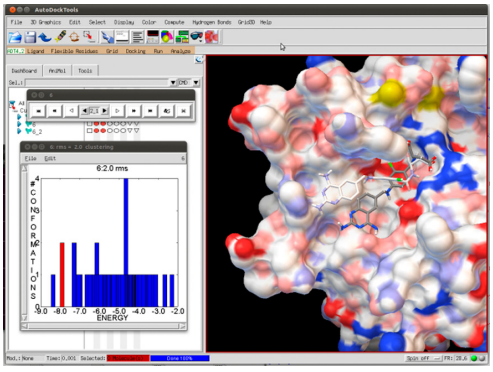
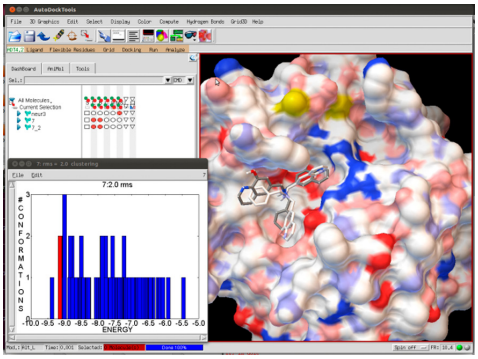
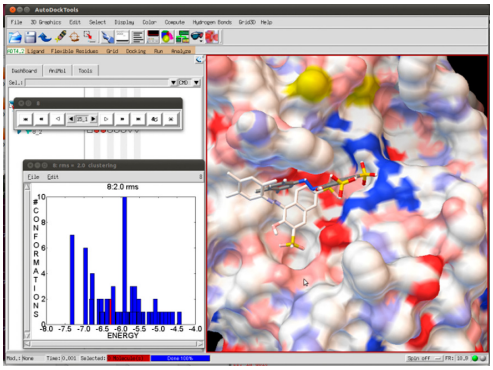


Table 4. continued

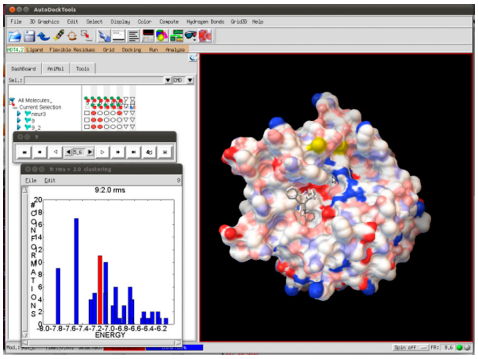
7



8



9



10

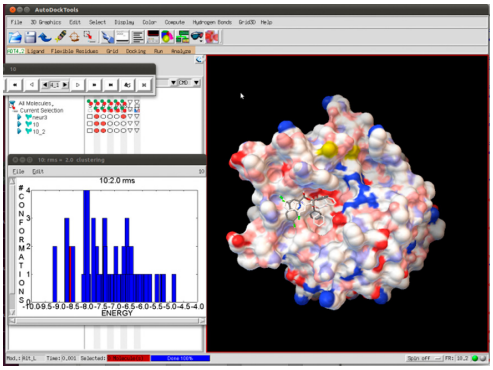


Table 5 shows the same experiment run with a selection of selected sialyl derivatives (S 1-S 7). The sugar molecules were made using Glycam Biomolecular Builder (http://glycam.ccrc.uga.edu/CCRC/biombuilder/biomb_index) translated into pdbqt using prepare_ligand4.py and then docked using Autodock Vina. It is noteworthy that, overall, the binding energy was significantly lower than the binding energies of the best small molecules. The RMSD for all but α-D-Neu5Ac-(2 – 3)-β-D-Galp-(1 – 4)-α-D-Glcp-OME (3.014 and 6.447) were 0. Interestingly the beta methylated sialyllactose bound at lower energy than the alpha conformer. It is surprising that a change in the conformation that seems to be so far from the active centre of the enzyme made such a difference (1.1 kcal/ml) to the affinity.

Table 5. The results for the sialyl derivatives binding to TV neuraminidase

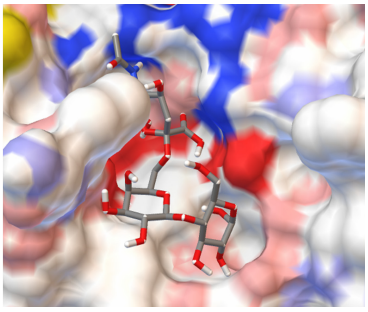
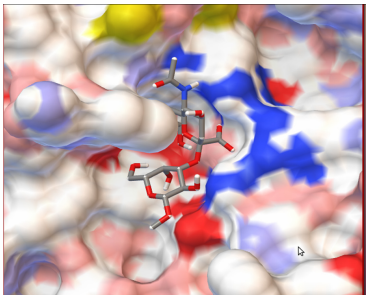
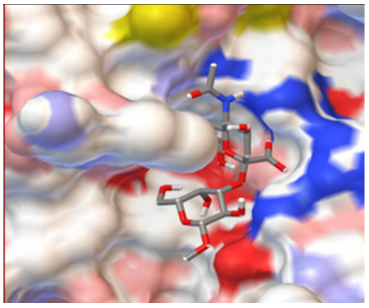
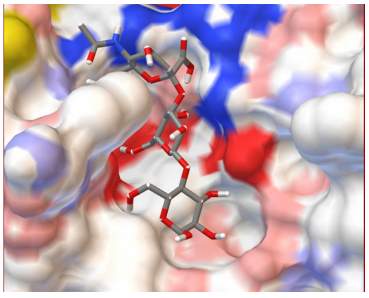
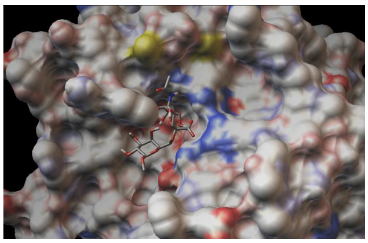
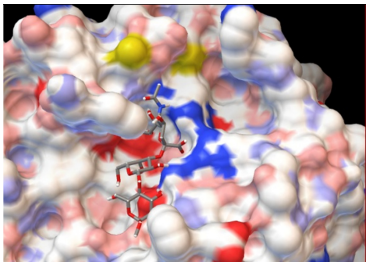
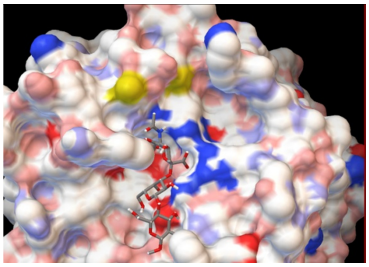
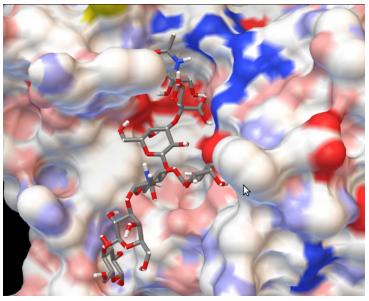
	Docking	Affinity kcal/mol	Structure/derivative
S1		-6.8	α-D-Neu5Ac-(2 – 6)-β- D-Galp-(1 – 4)- D-Glcp α2,6 sialyllactose
S2		-6.5	α-D-Neu5Ac-(2 – 6)-α-D-Glcp-(1 – 4)-OME α-methylated α2,6 sialylglucose
S3		-6.6	α-D-Neu5Ac-(2 – 6)-β-D-Galp-(1 – 4)-OME β-methylated α2,6 sialylgalactose
S4		-6.8	α-D-Neu5Ac-(2 – 3)-β-D-Galp-(1 – 4)-D-Glcp α2,3 sialyllactose

Table 5. continued

	Docking	Affinity kcal/mol	Structure/derivative
S5		-6.7	α-D-Neu5Ac-(2 – 3)-β-D-Galp-(1 – 4)-alpha -D-Glcp-OME α-methylated α2,3 sialyllactose
S6		-7.6	α-D-Neu5Ac-(2 – 3)-β-D-Galp-(1 – 4)-β-D-Glcp-OME β-methylated α2,3 sialyllactose
S7		-7.2	α-D-Neu5Ac-(2 – 3)-β-D-Galp-(1 – 4)-β-D-Galp-OME β-methylated α2,3 sialyl β galactobiose
S8		-7.4	α-D-Neu5Ac-(2 – 3)-β-D-Galp-(1 – 4) – (α – L-Fuc [1 – 3])- β-D-GlcNAc- β-D-Galp(1 – 4)- β – D-Glcp α2,3 sialyllacto-N-fucopentaose III

The *T. vaginalis* neuraminidase is reported to be specific for α-2,3 sialyl linkages yet the α-2,3 and α-2,6-sialyllactoses bound with a similar affinity although there was an obvious steric hindrance where the carboxyl group was not properly positioned in the α-2,6 linked compound. The assumption must be that the α-2,6 conformer binds but is not cleaved whereas the α-2,3 is both bound and cleaved. Of course, the protein may change conforma-

tion and adapt to the ligand. A molecular dynamics investigation would provide support for the answer. Screening the sugar library provided by Thomas Lütteke revealed that a sialyl Lewis^x derivative, α-2,3 sialyllacto-*N*-fucopentaose III (S 8) bound with the lowest energy (-7.4 kcal/mol). Indeed of the top 100 hits returned, eight were related to α-2,3-sialyl Lewis^x.

Discussion

In our Autodock Vina workflow, the small molecules bound significantly better than the ''natural'' sugar ligands. Although focused on the interactions of sialic acid binding lectins, the paper by Kiefel and von Itzstein [25] addresses the possibilities of using other sialic acid mimetics therapeutically. In theory such mimetics are more stable than complex sugars and easier to produce. Having determined that a mimetic has good characteristics, ''mutations'' may be made to improve the binding. The potential of producing better molecules for the inhibition of neuraminidase has been shown in the development of Relenza [24] (Zanamivir), a viral neuraminidase inhibitor. The parent sialic acid analogue, 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid (DANA), is 100 times less effective than the engineered molecule.

The screening of the *T.vaginalis* against the 8,000 sugars in the library supplied by Thomas Lüttecke revealed, as anticipated a sialyl compound, α-2,3-sialyllacto-*N*-fucopentaose 1 (a sialyl Lewis ^{x} containing molecule), as the best ligand. Indeed, eight of the top 100 binding molecules contained both fucose and sialic acid. Such results increase confidence in the approach. The next task is to carry out molecular dynamics on the best compounds identified and then to test the most robust on the purified enzyme *in vitro* to compare to the *in silico* results. A preliminary molecular dynamics run with the sialyl Lewis^x derivative reveals a robust conformation that appears to be stabilized by the presence of the α3-fucosylated residue.

Generally the binding prediction of the much more flexible sugar molecules, especially those with more than 3 to 4 rings is much more difficult using the classical Autodock 4.2 approach since thousands of docking runs are necessary to achieve results in agreement with X-ray crystallography conformations. The difference in docking energy between the small molecules and the sugars seen in the experiments may be attributable to the bigger conformational space probed by the more than 200,000 docked small molecules, electrostatic and van der Waals interactions not found in sugars (most of the docked compounds contain aromatic rings) or simply the lack of sufficient trials to probe all the possible conformations.

To check for possible conformation changes in the receptor molecule it will be necessary to complement the docking experiments with consecutive molecular dynamics runs using Gromacs or Amber11. GPU implementations of both already exist, with the latest version of pmemd.cuda (Amber11) clearly outperforming the current GPU implementation of Gromacs [26].

Conclusion

Results from the preliminary screening of the sugar libraries are in agreement with the known specificity of the neuraminidase. The utility of the results from the screening of the small molecule library will be determined when the relevant "wet lab" work is completed. Nevertheless, we now have a user-friendly system that enables researchers to interrogate their proteins with more than 200,000 ligands generating potential targets for *in vitro* work.

References

[1] Kiss, T., Szmetanko, G., Terstyanszky, G., Greenwell, P., Heindl, H. (2010) Molecular docking simulations on a combined desktop and service grid infrastructure. In: *Proceedings: Third AlmereGrid Desktop Experience workshop: desktop grid applications for eScience and eBusiness*, 30 March 2010, Almere, The Netherlands. EnterTheGrid, pp. 23 – 27.

[2] Heindl, H., Kiss, T., Terstyanszky, G., Weingarten, N., Afrough, B., Greenwell, P. (2010) ProSim: development of a user-friendly molecular modelling package. In: Hicks, Martin G. and Kettner, Carsten, (eds.) *Proceedings of the International Beilstein Symposium on Glyco-Bioinformatics: Bits 'n' Bytes of Sugars*, October 4th – 8th, 2009, Potsdam, Germany. Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Logos-Verlag Berlin, pp. 61 – 85.

[3] Kiss, T., Greenwell, P., Heindl, H., Terstyanszky, G., Weingarten, N.(2010) Parameter sweep workflows for modelling carbohydrate recognition. *J. Grid Comput.* **8:**587 – 601.
doi: http://dx.doi.org/10.1007/s10723-010-9166-8.

[4] Varghese, J.N. (1999) Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug Dev. Res.* **46:**176 – 196.
doi: http://dx.doi.org/10.1002/(SICI)1098-2299(199903/04)46:3/4<176::AID-DDR4>3.0.CO;2-6.

[5] http://www.gromacs.org

[6] http://autodock.scripps.edu/resources/adt

[7] http://www.nvidia.co.uk/object/gromacs_on_tesla.html

[8] http://autodock.scripps.edu/faqs-help/manual/autodock-4 – 2-user-guide/AutoDock4.2_UserGuide.pdf

[9] http://mgltools.scripps.edu/

[10] Reynolds, C.J., Winter, S., Terstyanszky, G.Z., Kiss, T., Greenwell, P., Acs, S., Kacsuk, P. (2011) Scientific workflow makespan reduction through cloud augmented desktop grids. *CloudCom.* In press.

[11] http://dtp.nci.nih.gov/index.html

[12] Dalby, A., Nourse, J.G., Hounshell, W.D., Gushurst, A.K.I., Grier, D.L. *et al.* (1992) Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comp. Sci.* **32:**244 – 255. doi: http://dx.doi.org/10.1021/ci00007a012.

[13] http://openbabel.org

[14] http://molprobity.biochem.duke.edu/

[15] http://www.ambermd.org

[16] http://vina.scripps.edu/

[17] Trot, A., Olsen, J. (2010) Autodock Vina: improving speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comp. Chem.* **31:**455 – 461. doi: 10.1002/jcc.21334.

[18] Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comp. Chem.* **19:**1639 – 1662. doi: http://dx.doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

[19] Thurman, A.R., Doncel, G.F. (2011) Innate immunity and inflammatory response to *Trichomonas vaginalis* and bacterial vaginosis: relationship to HIV acquisition. *Am. J. Reprod. Immunol.* **65:**89 – 98. doi: http://dx.doi.org/10.1111/j.1600-0897.2010.00902.x.

[20] Padilla-Vaca, F., Anaya-Velazquez, F. (1997) Biochemical properties of a neuraminidase of *Trichomonas vaginalis. J. Parasitol.* **83:**1001 – 6. doi: http://dx.doi.org/10.2307/3284352.

[21] http://www.trichdb.org

[22] Mohamed, M.S., El-Domany, R.A., El-Hameed, R.H.A.(2009) Synthesis of certain pyrrole derivatives as antimicrobial agents. *Acta Pharm.* **59:** 145 – 15. doi: http://dx.doi.org/10.2478/v10007-009-0016-9.

[23] http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid = 00421455

[24] von Itzstein, M., Wu, W.Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Van Phan, T., Smythe ,M.L., White, H.F., Oliver, S.W., *et al*. (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363:**418 – 23. doi: http://dx.doi.org/10.1038/363418a0.

[25] Kiefel, M.J., von Itzstein, M. (2002) Recent advances in the synthesis of sialic acid derivatives and sialylmimetics as biological probes. *Chem. Rev.* **102:**471 – 90. doi: http://dx.doi.org/10.1021/cr000414a.

[26] http://ambermd.org/gpus/

*All internet sites access 22/11/2011