

PROSIM: DEVELOPMENT OF A USER-FRIENDLY MOLECULAR MODELLING PACKAGE

HANS HEINDL¹, TAMAS KISS², GABOR TERSTYANSZKY²,
NOAM WEINGARTEN², BABAK AFROUGH³ AND
PAMELA GREENWELL^{1,*}

¹School of Life Sciences, University of Westminster, London, U.K.

²Centre for Parallel Computing, School of Electronics and Computing,
University of Westminster, London, U.K.

³Institute for Women's Health, University College London, London, U.K.

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

Received: 18th March 2010 / Published: 10th December 2010

ABSTRACT

We have developed and tested a user-friendly automated molecular modelling workflow with a web-based interface. The workflow has been tested using protein: drug, enzyme: substrate and lectin: sugar interactions. The work presented here involves studies using 3 glycosidases (a human mannosidase, a viral neuraminidase and a protozoan sialidase). We have illustrated the utility of the workflow using the mannosidase and thio-saccharide inhibitor. We have shown the limitations of such *in silico* technology when working with enzymes like the viral neuraminidase in which dynamic conformation changes take place during the binding or where access to the active site is blocked by a loop or a single residue. The viral neuraminidase inhibitor does not bind to the protozoan sialidase suggesting that such inhibitors would have no use in therapeutic strategies.

BACKGROUND

The work to be described involved the development and validation of a web-interfaced workflow incorporating energy minimisation, docking and molecular dynamics tools for the *in silico* modelling of proteins and their ligands. This worked formed part of a JISC supported project, ProSim, the objectives of which were:

1. To define user requirements and user scenarios of the protein molecule simulation involving energy minimization, checking and validating 3D structures, docking ligand to receptor molecule and refining/*de novo* prediction of ligand-receptor molecule interactions
2. To identify software packages required for protein molecule simulation, test these packages and select those suitable.
3. To automate the protein molecule simulation creating workflow templates and to provide parameter study support.
4. To develop application specific graphical user interfaces (application portlets) to provide seamless access to Grid resources and services.
5. To port and run protein molecule simulation on the National Grid Service (NGS) and make it available for the bioscience research community.

The workflows are generic and work well with glycans, drugs and enzyme substrates. However, although the developed workflow has broad potential, in the context of this paper, we will focus on carbohydrate (glycan) recognition: a phenomenon critical to a number of biological functions in humans including highly specific responses of the immune system [1–3] and interactions of sugars with biosynthetic (glycosyltransferase) and degrading (glycosidase) enzymes [4, 5]. Protein-glycan interactions are keys in cancer cell biology and in recognition of bacteria and viruses [6–8]. Indeed, perturbation of this binding has therapeutic potential. Unlike polypeptides and proteins, oligosaccharides are highly dynamic molecules that may occupy different conformations over time and space. Understanding their conformations should provide clues towards the mechanisms which lead to specific and selective recognition of carbohydrates by a range of proteins [9, 10]. Our current understanding of biological intermolecular responses is dominated by mechanisms involving cell surface proteins, however it is now well documented that complex carbohydrates (often associated with membrane proteins or lipids) are major contributors to the mechanism of specificity in biological recognition processes.

INTRODUCTION

The value of GRID computing

In silico tools that perform docking and molecular dynamics simulations, have significant potential to contribute to biomedical research. However, simulations are computationally intensive and even with relatively small and less complex molecules can take weeks to complete on a single PC. As the “jobs” are repetitive in nature and could be split into “sub-jobs”, the simulations can be run in parallel and the most obvious way to speed up the computation is to use computer clusters or the GRID [11]. In our developed workflow, the jobs are ported to and run on either the University of Westminster cluster or the UK National Grid. The computer scientists in the team have provided seamless access to computer and data resources available on the UK National Grid Service (NGS) using a P-GRADE grid, a web based, service rich environment for the development, execution and monitoring of workflows and workflow based parameter studies on various grid platforms, to run these simulations. Where appropriate, we have also exploited the Grid portal to access additional resources scattered amongst various other Grids, such as the EU Enabling Grid for E-SciencE (EGEE), the US Open Science Grid (OSG) and the US TeraGrid (TG), which also support international co-operation between bioscientists.

In our workflow, we combined GROMACs (a molecular dynamics simulation package) [12] with AutoDock (<http://w3.to/autodock>) which uses a search algorithm (*e.g.* Genetic Algorithm, and Monte Carlo) in combination with a free-energy desolvation function to provide a quick method for identification of putative active/binding sites [13]. The availability of such programs, run on parallel systems with a web-interface, would enable bioscientists to focus their resources and better plan experiments by allowing them to visualise potential interactions and determine the best molecules to investigate in the wet laboratory. This would reduce time and cost and increase the numbers of molecules screened. The ProSim project incorporated both application development and application execution.

Application development

Computer scientists migrated the protein molecule simulation applications as legacy codes to the Grid creating Grid services from these applications [14]. They also developed workflow templates representing different low-level user scenarios and/or applications specific portlets corresponding to high-level user scenarios.

Application execution

Bio-scientists execute the workflow templates through the Grid portal. These workflows are built on workflow templates that allow bio-scientists to modify the templates, create their own workflows, parameterise and run them. Application portlets hide the workflow-level

details and enable users to define their input parameters, run and monitor their experiments, and capture and visualise the results through applications portlets. The Grid portal offers both parameter study and computational workflow support. These are essential to these simulations, considering the wide value range of input parameters and the large number of simulation steps involved.

Molecular Dynamics (MD) simulation packages such as AMBER (<http://ambermd.org/>; [15]) and CHARMM (<http://www.charmm.org/>; [16]) can provide quantified data for time dependant conformational behaviour of dynamic biological molecules such as complex carbohydrates. However these programs run slowly and they are commercial software packages which raise licensing issues on the Grid. As a result in our prototype we used GROMACS [12] which is an open-source software. It gives acceptable speed and allows the use of force fields from other molecular dynamics (e.g. AMBER) packages. AutoDock (<http://autodock.scripps.edu/>) which uses a search algorithm in combination with an empirical binding free energy desolvation function provides a quick method for identification of putative active/binding sites in recognition. Molecular dynamics can then be used to dissect the mechanism of recognition after docking and hence provide insight into the physiochemical properties which provide biological phenomena with remarkable specificity in recognition processes.

Simulation packages which can provide insight into biological recognition processes have significant potential to contribute to biomedical research if the results of the simulation can prove consistent with the outcome of conventional wet laboratory experiments (Figure 1). The aim of the ProSim project was to optimise and develop workflows consisting of command line driven software applications which are executable through simplified web-based interfaces to serve the purpose of either predicting or informing/directing wet laboratory procedures for biologists with limited computing skills.

a) Validation and refinement



b) Utility

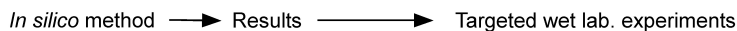


Figure 1. Experimental plan

PROSIM APPROACH

To achieve this aim, we are validating the use of *in silico* modelling in determining how proteins interact with ligands and how to manipulate them to improve or change their specificity. To do this, we have integrated readily available software programs in an optimised workflow to reproduce receptor-ligand complexes with a good degree of accuracy (Figure 2).

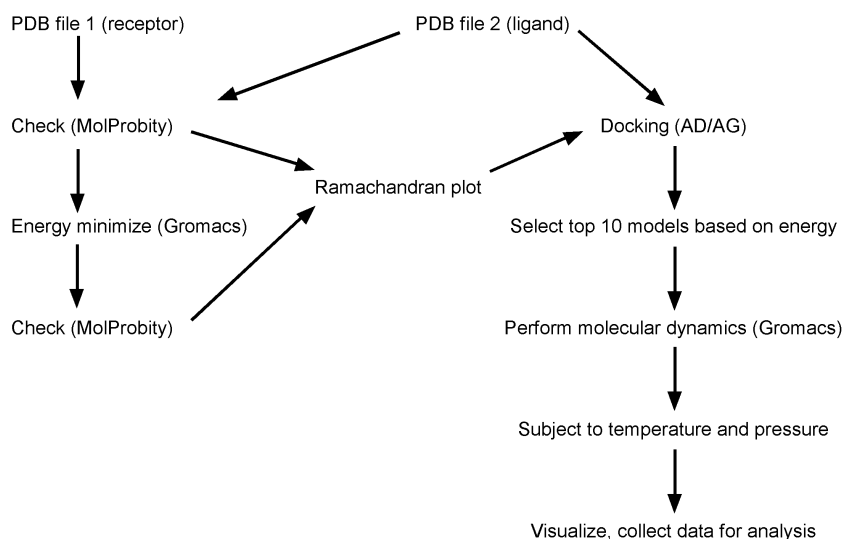


Figure 2. A schematic of the modelling process

In the workflow, the simulation process is divided into four phases: phase 1 involves selecting and preparing the receptor, phase 2 is selecting and preparing the ligand, docking the ligand to the receptor is performed in phase 3 and refining of the ligand-receptor molecule is carried out in phase 4. A schematic, showing how the output of one process is fed automatically into the next is shown in Figure 3. The boxes represent jobs which can take inputs through input ports and produce outputs through output ports (represented by the tiny boxes with the numbers). Application developers can connect programs through these ports to create workflows. The user does not have to deal with storage and program execution issues.

Phase 1 and 2: selecting and preparing receptor and phase ligand

Phase 1 and 2 incorporate the same operation: selection, pre-processing, solvation, energy minimisation and validation of the selected component in order to prepare the components required for docking. The pdb files contain Cartesian atomic coordinates which define 3-D

models of organic and organo-metallic compounds (Bernstein *et al.* 1977). These Cartesian coordinates allow 3-D model visualisation using computer assisted software such as RasMol (<http://rasmol.org/>; [17]) and Swiss PDB viewer deep view (<http://spdbv.vital-it.ch/>).

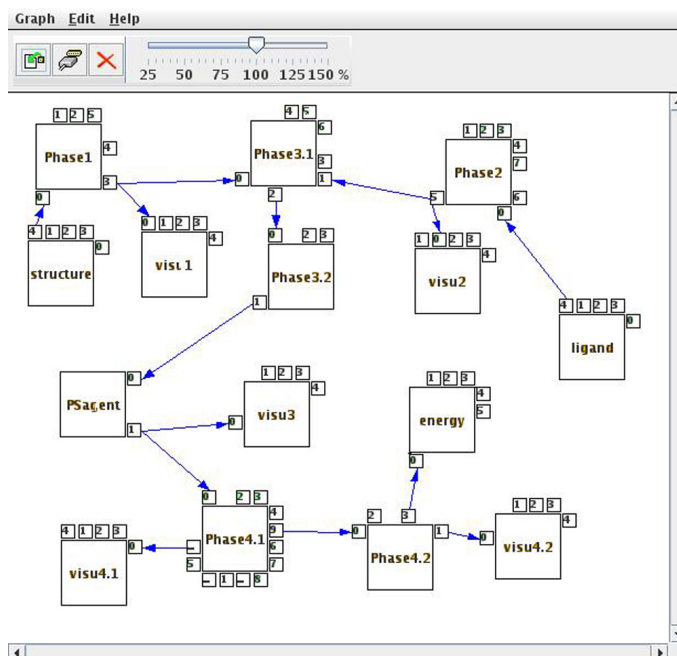


Figure 3. The phases of the workflow.

There are a number of different methods available which can be used to retrieve or generate pdb coordinates for various compounds. For complex biological compounds with thousands of atoms, the pdb coordinates are most frequently derived from data generated using X-ray crystallography and NMR spectroscopy. These pdb coordinates can be downloaded from a central database at <http://www.rcsb.org>. A second method for generating 3-D structures of complex biological compounds is through *in silico* homology modelling or fold recognition for which there are also a number of servers available such as SWISS MODEL [18] and PHYRE [19].

For simpler compounds, 3-D structures in pdb format can be generated using Smiles (http://www.molecular_networks.com/products/corina) coding strings. This method produces a rough structure which then requires further refinement or, through the use of bio-molecular force fields, the introduction of appropriate parameters. The use of the appropriate force field allows optimisation of the potential energy and overall 3-D topology (such as partial atomic charges, bond length and torsions) of the compounds. Force field parameters can either be based on quantum chemistry calculations and/or defined by wet laboratory experimental observations (i.e. X-ray/NMR data).

The downloaded pdb files are not always correct and often require a number of pre-processing steps prior to use in experiments. This may include tasks such as removing coordinates for atoms which are not physiological components of the selected receptor but by-products of reactions used in solving the structure. The pdb coordinates probably do not represent the structure of the receptor at the energy minima and therefore require some form of optimisation. Although structures are often refined prior to deposition in the database, it is important to perform energy minimisation in order to relax the structure. This can be carried out in the presence of the required physiological solvent and ions. Although this step does not make the pdb model ‘more accurate’, it does correct any steric hindrance and repairs distorted geometries in the 3-D structure by releasing internal constraints. This step is particularly useful if the initial model has been manually manipulated (i.e. through the introduction of a mutated residue or addition of new bonds). In effect, the manipulated atoms are repositioned according to the defined force field used in the energy minimisation step and consequently produce a refined structure at the energy minimum. Changes in the position of atoms can be analysed by comparing the root mean squared deviation (RMSD) of the original with the energy minimised structure.

The structure is then checked using programs such as PROCHECK [20] or MolProbity [21]. For complex biological compounds such as proteins, a pdb file containing 3-D coordinates of the selected compound needs to be checked or validated. One way to assess the quality of the receptor structure (for proteins) is to produce a Ramachandran plot which describes the conformations of the individual building units of the protein molecule. The plot can give a good estimate of how “correct” the pdb coordinates are in comparison with that which is already known about protein structure [22].

Phase 3: docking the ligand to the receptor

The docking program used in the protein simulation is AutoDock. It performs the docking in two stages. First, it allows the ligand to move in geometrical space until contact with the receptor is made. Next, AutoDock employs a semi-empirical desolvation force field equation to calculate binding free energy. The lowest binding free energy is identified by a number of iterations (defined by the user) based on either genetic or simulated annealing algorithms. Finally, AutoDock ranks the docked conformations according to their lowest binding free energy. This method has been shown to predict and reproduce X-ray data with a good degree of accuracy [23].

The results generated after AutoDock can be either visualised or analysed using different tools. A visual 3-D model showing hydrogen bond formations between the receptor and the ligand can be produced. Since AutoDock utilises a scoring function, the data can also be clustered to show conformations of the ligand which interacts with the receptor through similar binding free energies.

Phase 4: refining the ligand-receptor molecule

AutoDock often produces a number of different ligand conformations when in contact with a receptor which are found within the same binding free energy. In order to map the atomic interactions between the ligand and the receptor, the molecular dynamics (MD), approach is used downstream of the docking experiment. Through performing MD on different docked models, as defined by their binding free energy data, the most stable complexes can be identified as defined by the MD parameters. This can then be tested in the laboratory for validation through biochemical approaches.

BENEFITS AND LIMITATIONS OF OUR APPROACH

In addition to developing such workflows, by parallelisation we can take advantage of significant reduction in computational time and generate more simulations which enable statistical analysis of larger data sets over shorter periods of time that should yield better models. An important outcome of our approach would be to enable biologists to identify limitations in computational chemistry software and consequently report back to developers to enable evolution and improvements in software packages in biological contexts.

Using the workflow

The biologist needs to log into the GRID system and then open the ProSim workflow. They can then navigate to a page where they upload their pdb files and set the interaction box size. It is important that the box is not too small as the molecules need to be completely encompassed by it, on the other hand if the box is too big, the molecules will not encounter each other during the experiment. The workflow is then started and the jobs are sent to nodes on the GRID for completion and return of the finished data to the researcher. Throughout the workflow, progress can be monitored and results of workflow jobs can be viewed; this enables jobs with errors to be “spotted” and halted or refined. The researcher then has all the data to create models, analyse interaction parameters and further process the data. For more experienced users, there is the facility to refine the workflow and introduce more specific parameters based on experience and previous data.

Validating the workflow

Analysis of the docking of a thiodisaccharide substrate analogue to a human family 47 glycosylhydrolase (mannosidase)

The structure of a human family 47 glycosylhydrolase (mannosidase) with a bound thio substrate analogue of an α -(1,2)-dimannan has previously been solved using X-ray crystallography (pdb file 1X9D). The thiodisaccharide is a substrate analogue that is cleaved slowly such that the binding of the substrate to the enzyme can be analysed. The natural substrate of the enzyme is not used as it is cleaved so quickly that it would not be seen in the

crystal structure. As a proof of principle of our *in silico* methodology, a *de novo* constructed ligand structure was docked to the protein, using as target the crystal structure from which the analogue had been removed.

Preparation of the 1X9D pdb file

As there were some atoms missing in the coordinate file of the original crystal structure, in order to complete the file for submission the sequence was first submitted to the Swiss homology model server (<http://swissmodel.expasy.org/>; [18]). The molecule is shown as a ribbon representation in Figure 4. The molecule has a central calcium ion and, when viewed in 3D it is apparent that the catalytic site is situated at the bottom of a deep funnel-like channel. The terminal α -1 – 2 linked mannose is thought to be coordinated by the Ca^{2+} ion, bringing the linkage oxygen into the appropriate position for cleavage. Due to energy barriers on the way to the final position, the binding of the cleaved products is much weaker forcing the reaction products out of the catalytic centre of the enzyme.

Preparation of the thiodisaccharide coordinate file

The ligand file was prepared using a molecular editor and subsequent translation of the small molecule file into a pdb coordinate file as previously described. This output was fed directly into the PROSIM workflow.

Docking results for the disaccharide analogue and 1X9D

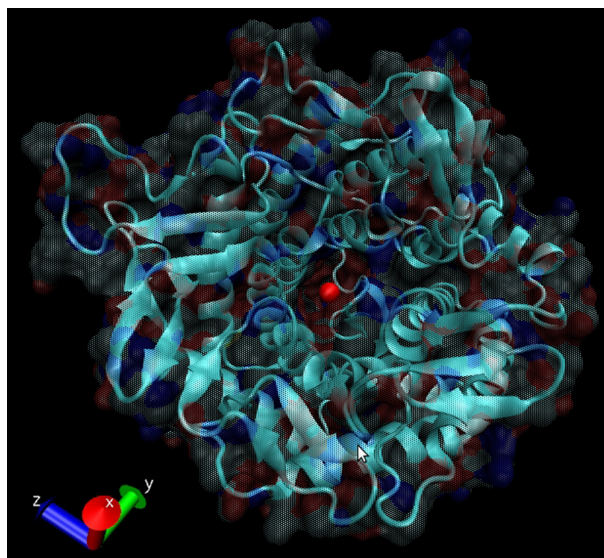


Figure 4. Ribbon representation of the human α -mannosidase 1X9D with the central calcium ion in vdW representation (red) and a transparent surface representation overlaid.

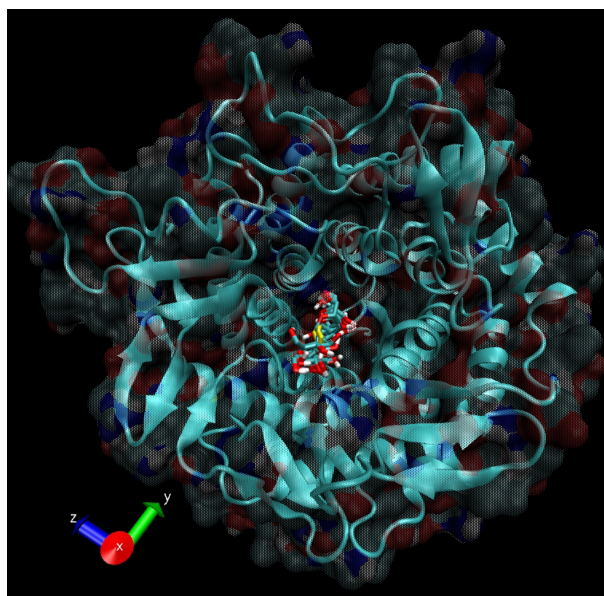


Figure 5. Five docking results shown together with the protein structure. The proper results are found in the third clustering group near the energy minimum.

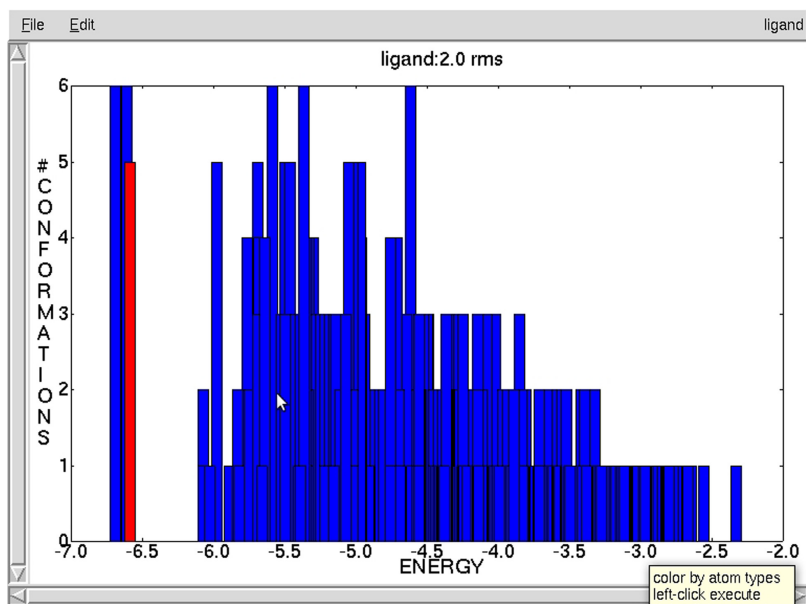


Figure 6. Clustering of the 1000 docking runs performed by the PROSIM workflow. The cluster containing the conformations in the proper position is marked red.

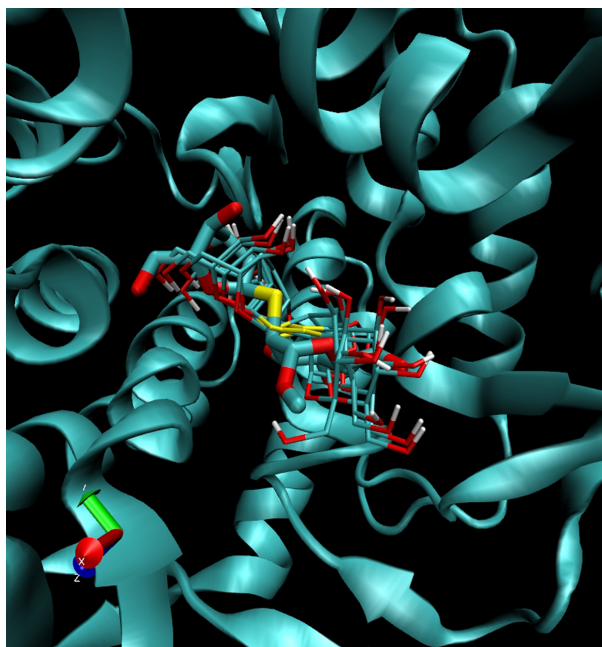


Figure 7. Superposition of the five Autodock results (in thin stick representation) and the thiodisaccharide from the original crystal structure.

The results of docking 5 analogues to the enzyme are shown in Figure 5. The automated workflow could not find the proper docking site using the blind dock option; twelve results showed an energetically more favourable result, but the difference was very small. It is apparent from the docking results (Figure 6) that due to lectin-like binding at alternative sites on the receptor surface further favourable docking sites may be found (especially in blind docking experiments). In blind docking wider grid spacing is used and hence the final positioning may not be as exact as if an approximate docking site was chosen on which docking could be focussed.

In Figure 7, the sugar ring at the position 1 in the catalytic centre (left) is very well aligned. The mannose at the position 2 (right) shows more variation. The position of the sulphur atom that is important for the catalytic mechanism is very well conserved. This shows that:

1. the *de novo* docking experiments reproduce the final position of the ligand found in the crystal structure
 2. the central sulphur atom which marks the position of the linkage oxygen is well placed
 3. there is much less conformational freedom for the central mannose residue than for the more distal one.
-

Comparison of the docking of oseltamivir to a H5N1 influenza virus neuraminidase and a *Trichomonas vaginalis* sialidase homology model

This trial aimed to demonstrate the potential binding of the antiviral drug oseltamivir (Tamiflu®) to the influenza neuraminidase and a homology model of a putative sialidase (neuraminidase) of the protozoan *Trichomonas vaginalis*. The binding of the drug to the influenza neuraminidase is well documented [24]. The aim was to validate the workflow using a known enzyme: inhibitor pair and to determine whether the protozoan enzyme bound the inhibitor and if so, what residues were responsible for that binding.

Preparation of the ligand pdb file

The ligand file was constructed by drawing a 2D sketch of the small molecule using JME a java based molecular editor (<http://www.molinspiration.com/jme/index.html>). The 2D construct was then translated into a 3D coordinate file using e.g. a SMILES translator or the program CORINA which is available on the molecular networks homepage (<http://www.molecular-networks.com/products/corina>). The ligand coordinate file generated by the SMILES translator was used as an input file for the ProSim workflow. A blind dock was performed to scan the receptor for possible low energy binding sites.

Preparation of the viral neuraminidase enzyme .pdb file

The protein coordinates were taken from the RCSB server (2HT7). As before missing residues and atoms were added by submission to a homology model server and the resulting pdb file submitted to the PROSIM workflow.

Preparation of the *T. vaginalis* enzyme pdb file

Using known eukaryotic sialidase sequences the *T. vaginalis* genome (www.trichdb.org) was searched for putative sialic acid cleaving enzymes. The applied domain recognition software (e.g. Pfam) suggested the existence of a six blade beta propeller fold which is typical for sialidases. In a next step the most promising sequences were fed into an automated homology model server and the returning results were critically reviewed with respect to structure probability and completeness. The arginine triad typical for sialidases was taken as a second “identifier” of the sialidase structure. Since this work was begun, the selected genes have been annotated out of the genome and classified as sialidases. The enzyme pdb file as it was generated by the Swiss homology model server was the input into the ProSim workflow.

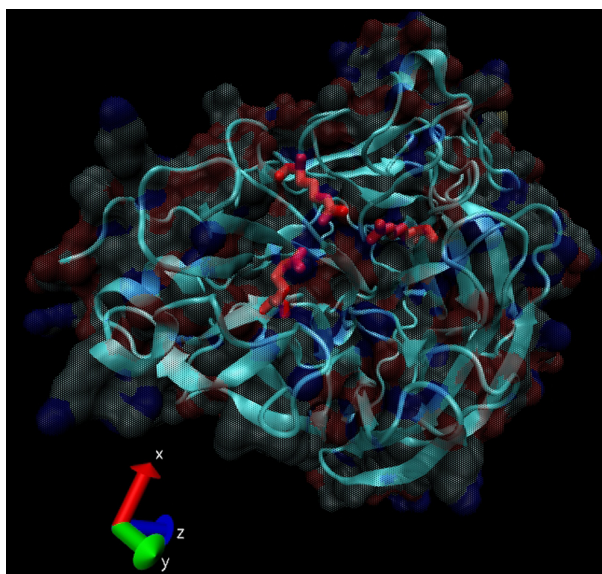


Figure 8. Ribbon representation of the influenza viral (H5N1) neuraminidase 2HT7. The characteristic arginine triad is shown in red stick representation and the molecular surface in transparent representation.

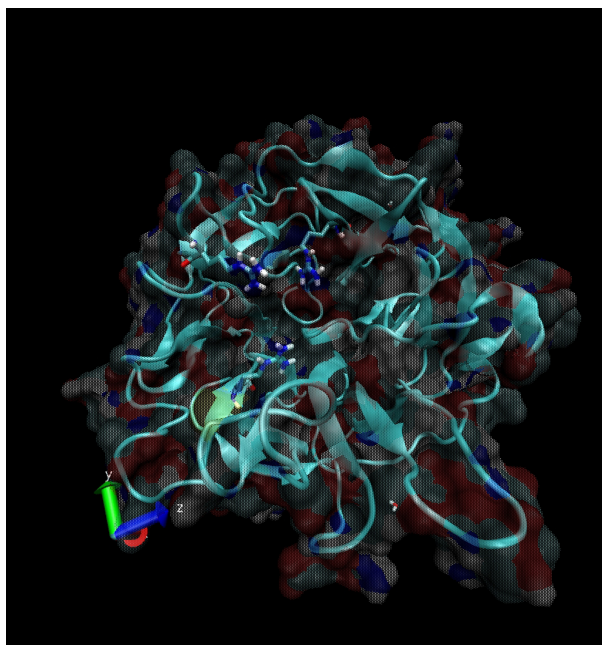


Figure 9. Visualization of the putative *T. vaginalis* sialidase showing a transparent molecular surface rendering the underlying secondary protein structure in cartoon mode and the arginine triad (blue) typical for this class of enzymes. The six-fold symmetry of the propeller structure is also seen.

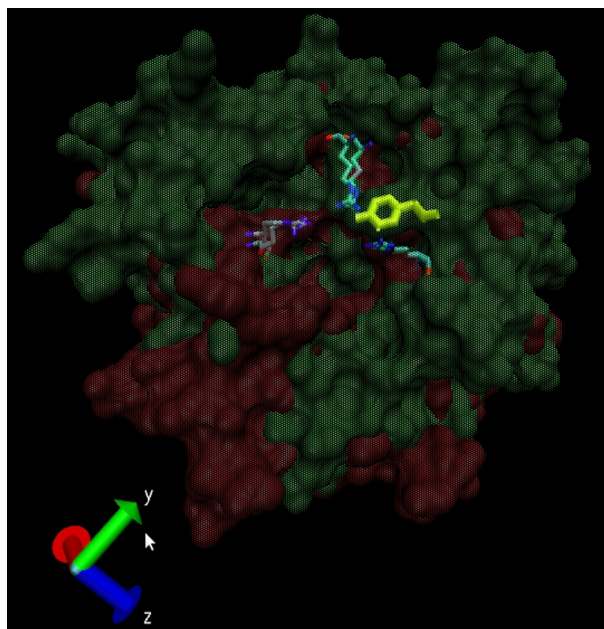


Figure 10. An alignment of the *T.vaginalis* sialidase and the H5N1 viral neuraminidase with the arginine triad in the proper position and the tyrosine residue (Y264) which blocks the access to the active centre and does not exist in the protozoan enzyme coloured in yellow.

Docking results for the viral Neuraminidase

The execution of the workflow typically takes less than 24 h. 1000 docking trials were performed and the 10 best conformations conserved. Of these ten conformations consecutive short molecular dynamic runs at room temperature 1 bar pressure, explicit solvent and ion concentration of 150 mM were performed.

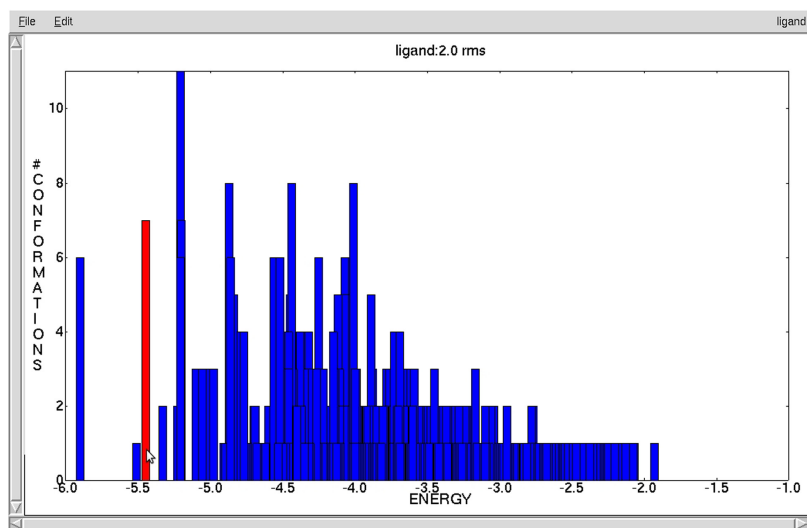


Figure 11. Clustering of the results of the oseltamivir viral neuraminidase docking. The red bar and the two neighbouring bars give reasonable results near the active centre of the enzyme.

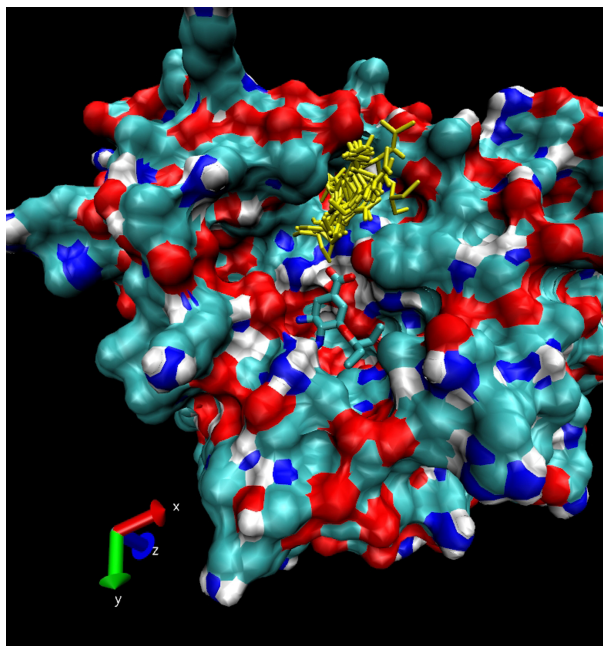


Figure 12. H5N1 viral neuraminidase with the substrate as localized in the crystal structure (stick representation near the centre of the molecule and with the ten docked results which are found near the active centre which is formed by the arginine triad.

The results of the blind docking of oseltamivir to the viral neuraminidase surprisingly do not agree with those of the crystal structure (see Figures 11 and 12). This is difficult to explain as it has been claimed that the antivirals oseltamivir and zanamivir have been designed using *in silico* methods. Considering the results of another group (Wang *et al.* 2010) the ligand seems to have docked in the region of the second substrate sugar whilst the active centre seems to be “unreachable” for the docking algorithm

A possible reason is a flexible tyrosine (Y264) that may, *in silico* impede the access to the active centre but may in reality be part of a flexible receptor system (induced fit). This residue is not found in the *T. vaginalis* enzyme. To prove this, the tyrosine in the viral neuraminidase was mutated to an alanine and the submission to the PROSIM workflow repeated. A run with the enzyme with no added inhibitor was conducted in order to study possible spontaneous movements and flexibility of the tyrosine residue

Docking results for the mutated receptor structure

As expected when the blocking tyrosine residue was replaced by the less bulky and apolar alanine some of the docking results with a low energy level were in the expected site (Figures 13, 14, 15 and 16).

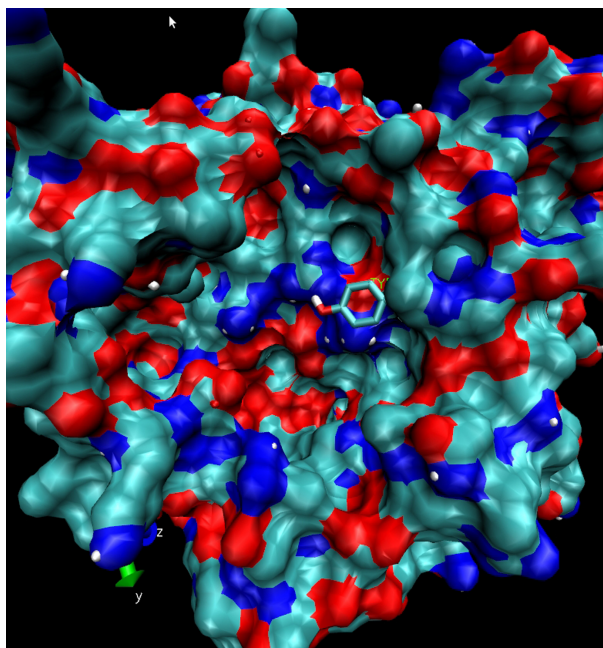


Figure 13. Critical tyrosine residue which may impede the docking if the receptor molecule is viewed as completely rigid.

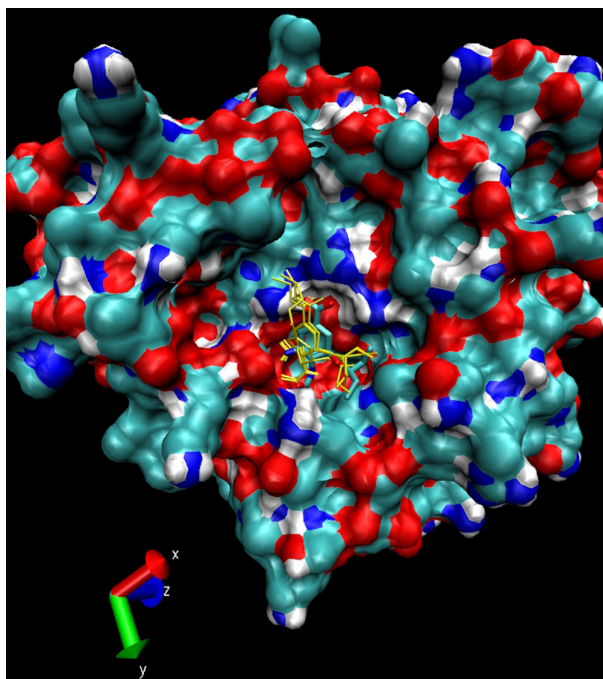


Figure 14. Four docking results placed properly by the Autodock algorithm (in yellow stick representation). The small molecule from the original crystal structure is shown in stick representation, the receptor in ribbon representation with a transparent surface.

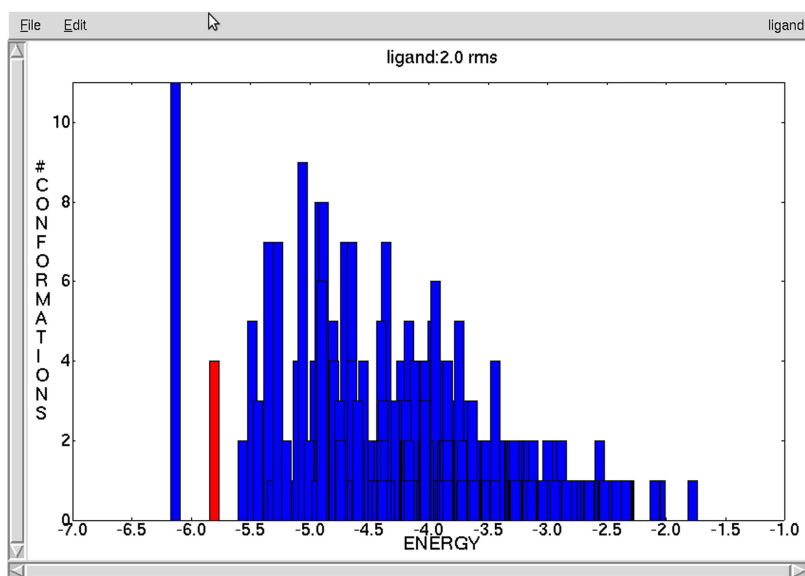


Figure 15. Clustering of the energy levels of the docking results for the mutated enzyme.

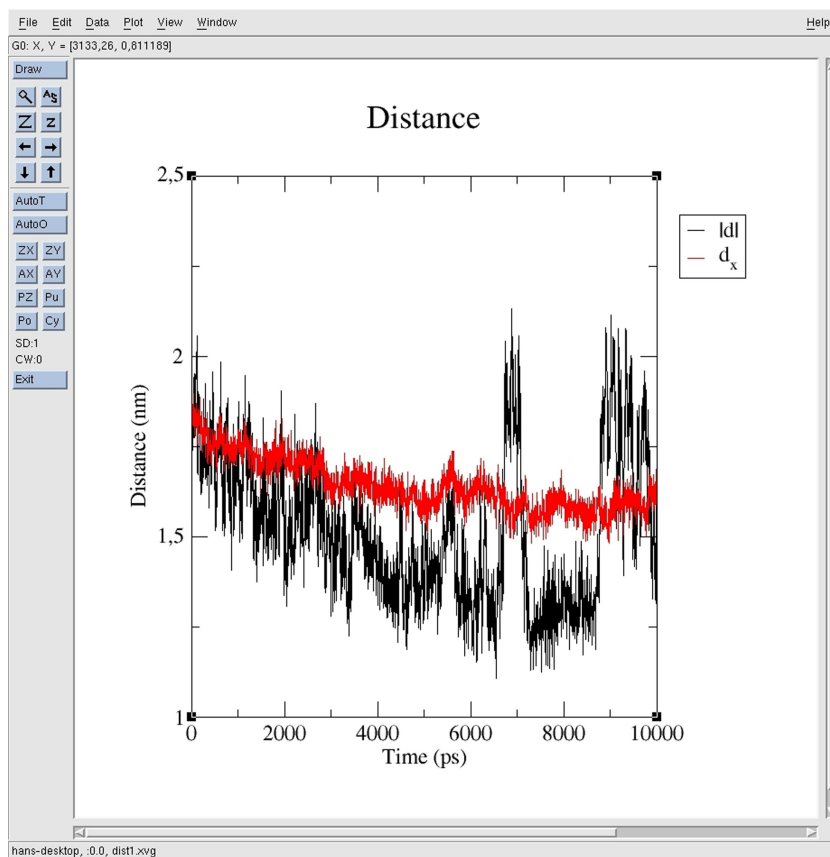


Figure 16. Distance between an atom near the centre of gravity of the receptor and the Y264 C-alpha (red) and the Y264 hydroxyl oxygen (black) showing the high mobility of the "lid".

The docking of oseltamivir to the protozoan sialidase

Sialidases of the sexually transmitted protozoon *T. vaginalis* facilitate the cleavage of terminal sialic acids which are frequently found in the mucus produced by epithelial cells and glands in the genitourinary tract of the host organism and are thought to play a role in cytotoxicity [25, 26]. The homology model of the sialidase was docked with the oseltamivir inhibitor and the results are shown in Figures 16 and 17. The ligand does not really enter the catalytic cavity but seems to block the putative half pipe like entrance that could impair substrate binding.

ProSim: Development of a User-friendly Molecular Modelling Package

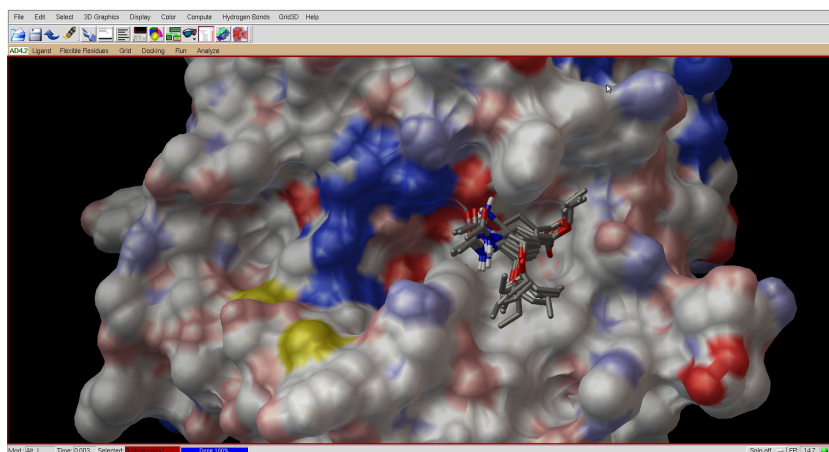


Figure 17. Six of the ten best docking results show the ligand in stick representation with the atoms coloured according to type and the receptor shown in surface representation. Note the arginine triad in blue left of the ligand.

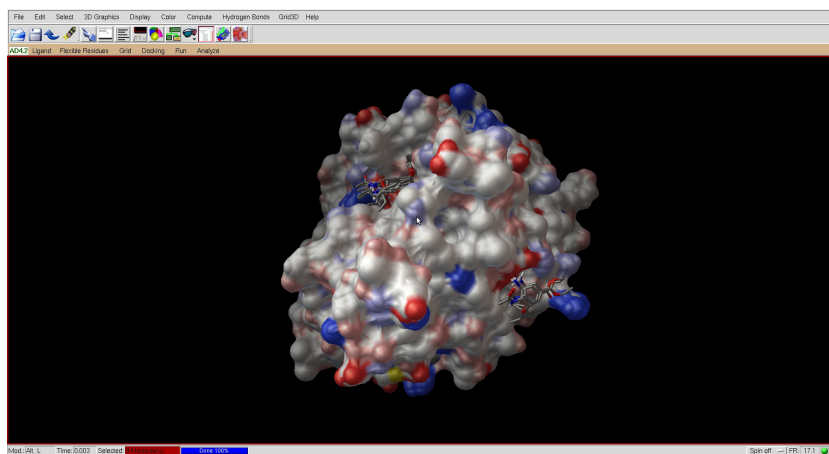


Figure 18. This figure shows a slightly turned version of Figure 17. The two ligand conformations visible in this picture represent the 10 best docking results from 1000 trials.

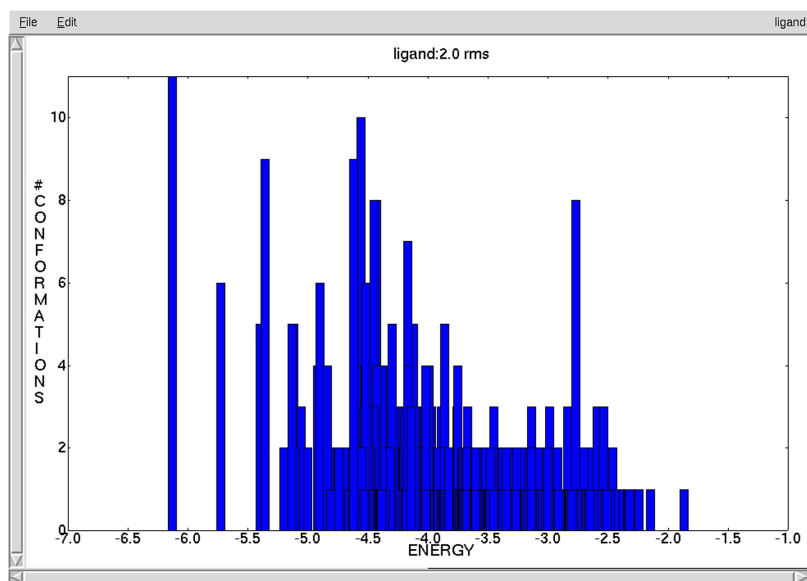


Figure 19. The histogram showing the “docking energy” (as named by the autodock convention) clustered in groups of max 2.0 rms. The group at the left represents the above shown conformations and can be well discerned from the bulk of results.

Analysis of the molecular dynamics trajectory

As a prerequisite the temperature and pressure profile of the molecular dynamics run is plotted to look for irregularities. For that purpose the GROMACS program *g_energy* was used and visualization performed using *xmgrace*.

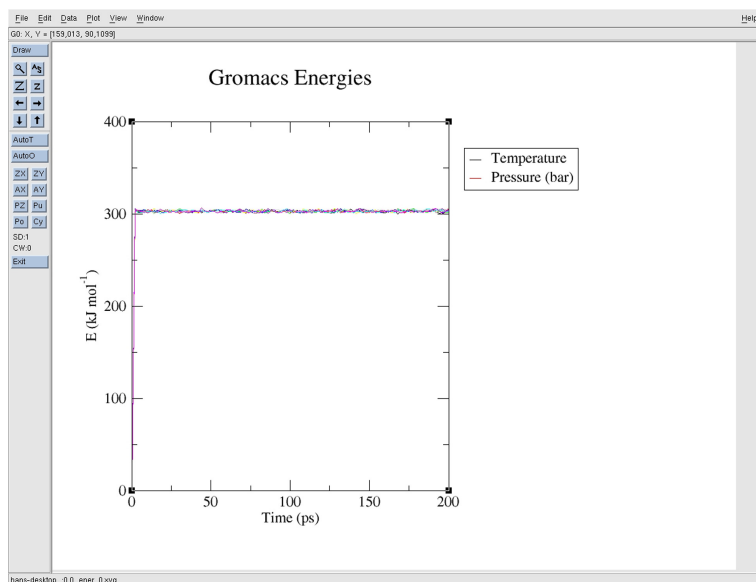


Figure 20. Temperature plot (overlay of 10 trajectories) showing constant temperature over the whole trajectory.

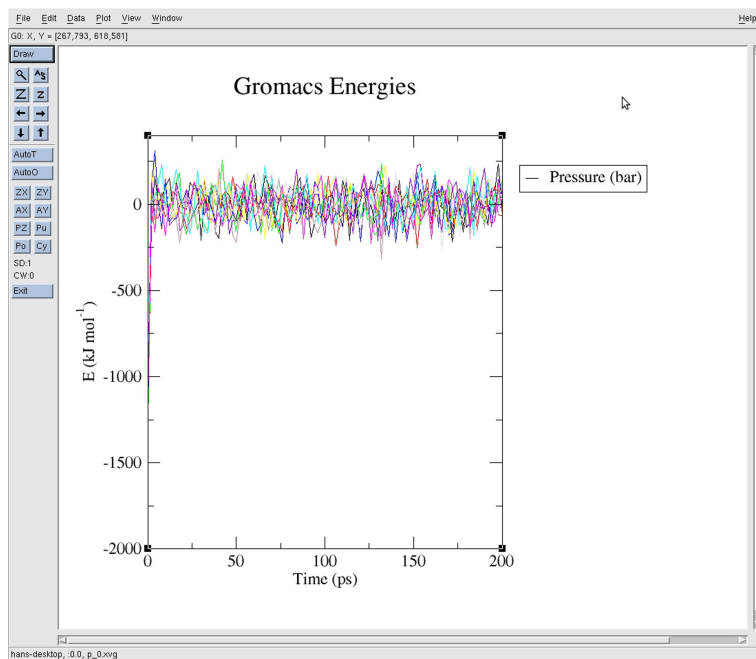


Figure 21. The pressure plot (overlay of all 10 trajectories) shows considerably more variation.

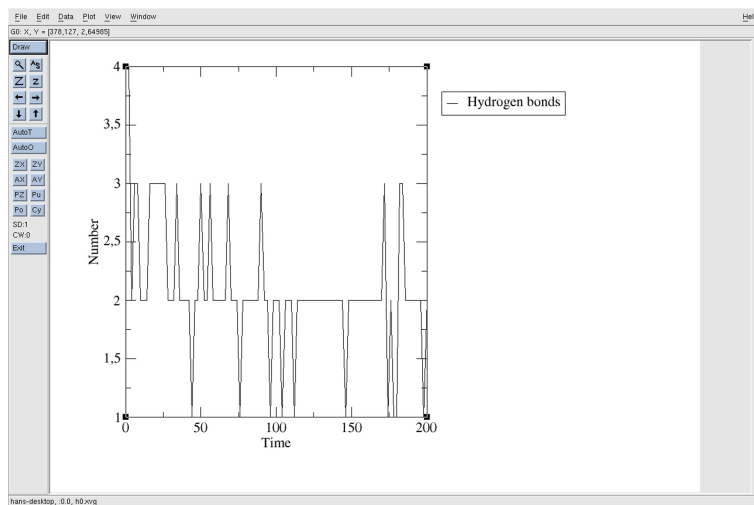


Figure 22. Number of H-bonds between the ligand and the enzyme as a function of time, generated with the function *g_hbonds* of GROMACS and visualized using *xmgrace*.

CONCLUSIONS

The PROSIM workflow is able to reproduce the docking of a sugar-like substrate very well as shown with the mannosidase. This example shows that sugars may also bind to other sites of the molecule with high affinity. A major improvement of the workflow would involve the user picking the most likely results and to exclude dockings with high affinity to areas other than the expected binding sites before starting the molecular dynamics run. An alternative would be to restrict the docking to regions of high interest. On the other hand the blind docking approach leads to more unbiased results. The second example involving the viral neuraminidase shows that AutoDock may fail if dynamic conformation changes take place during the binding or the access to the active site is blocked by a loop or a single residue. Interestingly although the *T. vaginalis* neuraminidase does not have an active site with a “lid”, the docking into the active site fails, unlike the mutated viral neuraminidase. Here a molecular dynamics simulation exerting a force (push or pull) on the ligand could help to understand the conformational changes during the binding process. These results could also represent the differences between a eukaryotic and viral enzyme with respect to the artificial ligand oseltamivir which may be the basis of the low toxicity of the drug.

REFERENCES

- [1] Zhang, X.L (2006) Roles of glycans and glycopeptides in immune system and immune-related diseases. *Curr. Med. Chem.* **13**(10):1141 – 7.
doi: <http://dx.doi.org/10.1016/j.ejmech.2009.10.040>.
 - [2] Crocker, P.R., Paulson, J.C., Varki, A. (2007) Siglecs and their roles in the immune system. *Nat. Rev. Immunol.* **7**(4):255 – 66.
doi: <http://dx.doi.org/10.1038/nri2056>.
 - [3] Erbacher, A., Gieseke, F., Handgretinger, R., Müller, I. (2009) Dendritic cells: functional aspects of glycosylation and lectins. *Hum. Immunol.* **70**(5):308 – 12.
doi: <http://dx.doi.org/10.1016/j.humimm.2009.02.005>.
 - [4] Henrissat, B., Sulzenbacher, G., Bourne Y. (2008) Glycosyltransferases, glycoside hydrolases: surprise, surprise! *Curr. Opin. Struct. Biol.* **18**(5):527 – 33.
doi: <http://dx.doi.org/10.1016/j.sbi.2008.09.003>.
 - [5] Hinou, H., Nishimura, S. (2009) Mechanism-based probing, characterization, and inhibitor design of glycosidases and glycosyltransferases. *Curr. Top. Med. Chem.* **9**(1):106 – 16.
doi: <http://dx.doi.org/10.2174/156802609787354298>.
 - [6] An, H.J., Kronewitter, S.R., de Leoz, M.L., Lebrilla, C.B. (2009) Glycomics and disease markers. *Curr. Opin. Chem. Biol.* **13**(5 – 6): 601 – 7.
doi: <http://dx.doi.org/10.1016/j.cbpa.2009.08.015>.
 - [7] Narimatsu, H., Sawaki, H., Kuno, A., Kaji, H., Ito, H., Ikehara, Y. (2010) A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics. *FEBS J.* **277**(1):95 – 105.
doi: <http://dx.doi.org/10.1111/j.1742-4658.2009.07430.x>.
 - [8] Reading, P.C., Tate, M.D., Pickett, D.L., Brooks, A.G. (2007) Glycosylation as a target for recognition of influenza viruses by the innate immune system. *Adv. Exp. Med. Biol.* **598**:279 – 92.
doi: <http://dx.doi.org/10.1007/978-0-387-71767-8>.
 - [9] Shaikh, F.A., Withers, S.G., (2008) Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis. *Biochem. Cell Biol.* **86**(2):169 – 77.
doi: <http://dx.doi.org/10.1139/007-149>.
 - [10] Woods, R.J. (1998) Computational carbohydrate chemistry: what theoretical methods can tell us. *Glycoconj. J.* **15**(3):209 – 16.
doi: <http://dx.doi.org/10.1023/A:1006984709892>.
-

- [11] Kacsuk, P., Sipos, G. (2005) Multi-Grid, multi-user workflows in the P-GRADE Grid Portal, *Journal of Grid Computing* **3**(3–4):221–238.
doi: <http://dx.doi.org/10.1016/j.future.2008.02.008>.
 - [12] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G. Mark, A.E., Berendsen, H.J. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**(16):1701–18.
doi: <http://dx.doi.org/10.1002/jcc.20291>.
 - [13] 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. Computational Chemistry* **19**:1639–1662.
doi: [http://dx.doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639](http://dx.doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639).
 - [14] Delaitre, T., Kiss, T., Goyeneche, A., Terstysanzky, G., Winter, S., Kacsuk, P. (2005) GEMLCA: Running legacy code applications as grid services. *Journal of Grid Computing* **3**:75–90.
doi: <http://dx.doi.org/10.1007/s10723-005-9002-8>.
 - [15] Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R.J. (2005) The Amber bio-molecular simulation programs. *Journal of Computational Chemistry* **26**:1668–1688.
doi: <http://dx.doi.org/10.1002/jcc.20290>.
 - [16] MacKerell, A.D., Brooks, C., Brooks, L., Nilsson, L., Roux, B., Won, Y., Karplus, M. (1998) CHARMM: *The Energy Function and Its Parameterization with an Overview of the Program*. In *The Encyclopedia of Computational Chemistry* **1**:271–277. Schleyer, P.v.R. *et al.*, editors. John Wiley & Sons: Chichester. ISBN: 0–471–96588-X
 - [17] Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.* **112**:535–42.
doi: [http://dx.doi.org/10.1016/S0022-2836\(77\)80200-3](http://dx.doi.org/10.1016/S0022-2836(77)80200-3).
 - [18] Arnold, K., Bordoli, L., Kopp, J., Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**(2):195–201.
doi: <http://dx.doi.org/10.1093/bioinformatics/bti770>.
 - [19] Bennett-Lovsey, R.M., Herbert, A.D., Sternberg, M.J., Kelley, L.A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* **70**(3):611–25.
doi: <http://dx.doi.org/10.1002/prot.21688>.
-

- [20] Morris, A.L., MacArthur, M.W., Hutchinson, E.G., Thornton, J.M. (1992) Stereochemical quality of protein structure coordinates. *Proteins* **12**(4):345 – 64.
doi: <http://dx.doi.org/10.1002/prot.340120407>.
 - [21] Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C. (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* **50**(3):437 – 50.
doi: <http://dx.doi.org/10.1002/prot.10286>.
 - [22] Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V. (1963): Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**:95–9.
doi: [http://dx.doi.org/10.1016/S0022-2836\(63\)80023-6](http://dx.doi.org/10.1016/S0022-2836(63)80023-6).
 - [23] Englebienne, P., Fiaux, H., Kuntz, D.A., Corbeil, C.R., Gerber-Lemaire, S., Rose, D.R., Moitessier, N. (2007) Evaluation of docking programs for predicting binding of Golgi alpha-mannosidase II inhibitors: a comparison with crystallography. *Proteins* **69**(1):160 – 76.
doi: <http://dx.doi.org/10.1002/prot.21479>.
 - [24] Collins, P.J., Haire, L.F., Lin, Y.P., Liu, J., Russell, R.J., Walker, A., Skehel, J.J., Martin, S.R., Hay, A.J., Gamblin, S.J. (2008) Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature* **453**:1258 – 1261.
doi: <http://dx.doi.org/10.1038/nature06956>.
 - [25] Padilla-Vaca, F., Anaya-Velázquez, F. (1997) Biochemical properties of a neuraminidase of *Trichomonas vaginalis*. *J. Parasitol.* **83**(6):1001 – 6.
doi: <http://dx.doi.org/10.2307/3284352>.
 - [26] Wiggins, R., Hicks, S.J., Soothill, P.W., Millar, M.R., Corfield, A.P. (2001) Mucins and sialidases: their role in the pathogenesis of sexually transmitted infections in the female genital tract. *Sex. Transm. Infect.* **77**(6):402 – 8.
doi: <http://dx.doi.org/10.1136/sti.77.6.402>.
-

