



**HAL**  
open science

## Atomistic Modeling of the Membrane-Embedded Synaptic Fusion Complex: a Grand Challenge Project on the DEISA HPC Infrastructure

Elmar Krieger, Laurent Leger, Marie-Pierre Durrieu, Nada Taib, Peter Bond, Michel Laguerre, Richard Lavery, Mark Sp Sansom, Marc Baaden

► **To cite this version:**

Elmar Krieger, Laurent Leger, Marie-Pierre Durrieu, Nada Taib, Peter Bond, et al.. Atomistic Modeling of the Membrane-Embedded Synaptic Fusion Complex: a Grand Challenge Project on the DEISA HPC Infrastructure. *Parallel Computing: Architectures, Algorithms and Applications*, Sep 2007, Julich, Germany. pp.729-736. hal-03907311

**HAL Id: hal-03907311**

**<https://hal.science/hal-03907311v1>**

Submitted on 6 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Atomistic modeling of the membrane-embedded synaptic fusion complex: a grand challenge project on the DEISA HPC infrastructure

Elmar Krieger<sup>1</sup>, Laurent Leger<sup>2</sup>, Marie-Pierre Durrieu<sup>3</sup>, Nada Taib<sup>4</sup>, Peter Bond<sup>5</sup>, Michel Laguerre<sup>4</sup>, Richard Lavery<sup>6</sup>, Mark S. P. Sansom<sup>5</sup>, and Marc Baaden<sup>3</sup>

<sup>1</sup> Center for Molecular and Biomolecular Informatics, Radboud University Nijmegen, Toernooiveld 1, 6525ED Nijmegen, The Netherlands

<sup>2</sup> IDRIS, Bâtiment 506, F-91403 Orsay cedex, France

<sup>3</sup> Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, CNRS UPR 9080, 13, rue Pierre et Marie Curie, F-75005 Paris, France, *E-mail: baaden@smplinux.de*

<sup>4</sup> UMR 5144 CNRS Mobios, Institut Européen de Chimie et Biologie, Université de Bordeaux 1, 2 Av. Robert Escarpit, F-33607 Pessac, France

<sup>5</sup> Department of Biochemistry, University of Oxford, South Parks Road, Oxford 1 3QU, United Kingdom

<sup>6</sup> Institut de Biologie et Chimie des Protéines, Département des Biostructures Moléculaires, CNRS UMR 5086 / Université de Lyon, 7, passage du Vercors, F-69367 Lyon, France

The SNARE protein complex is central to membrane fusion, a ubiquitous process in biology. Modeling this system in order to better understand its guiding principles is a challenging task. This is mainly due to the complexity of the environment: two adjacent membranes and a central bundle of four helices formed by vesicular and plasma membrane proteins. Not only the size of the actual system, but also the computing time required to equilibrate it render this a demanding task requiring exceptional computing resources. Within the DEISA Extreme Computing Initiative (DECI), we have performed 40 ns of atomistic molecular dynamics simulations with an average performance of 81.5 GFlops on 96 processors using 218 000 CPU hours. Here we describe the setup of the simulation system and the computational performance characteristics.

## 1 Introduction

Exocytosis involves the transport of molecules stored within lipid vesicles from the inside of a cell to its environment. The final step of this process requires fusion of the vesicles with the plasma membrane and is mediated via SNARE (soluble N-ethylmaleimide sensitive factor attachment protein receptor) fusion proteins. Figure 1A shows an atomic model of the SNARE complex, a bundle of four protein helices, that supposedly brings and holds the two biological membranes together. This model was built upon previous studies and used as starting point for molecular dynamics simulations. The SNARE complex is a target for studying several pathologies such as botulism and tetanus. The purpose of our simulations is to obtain a detailed atomic picture of the structure, conformational dynamics and interactions in this system in order to improve our understanding of membrane fusion and related molecular processes like the diseases mentioned before. For further details and biological background information, see Ref. <sup>1</sup> and the references cited therein, as these

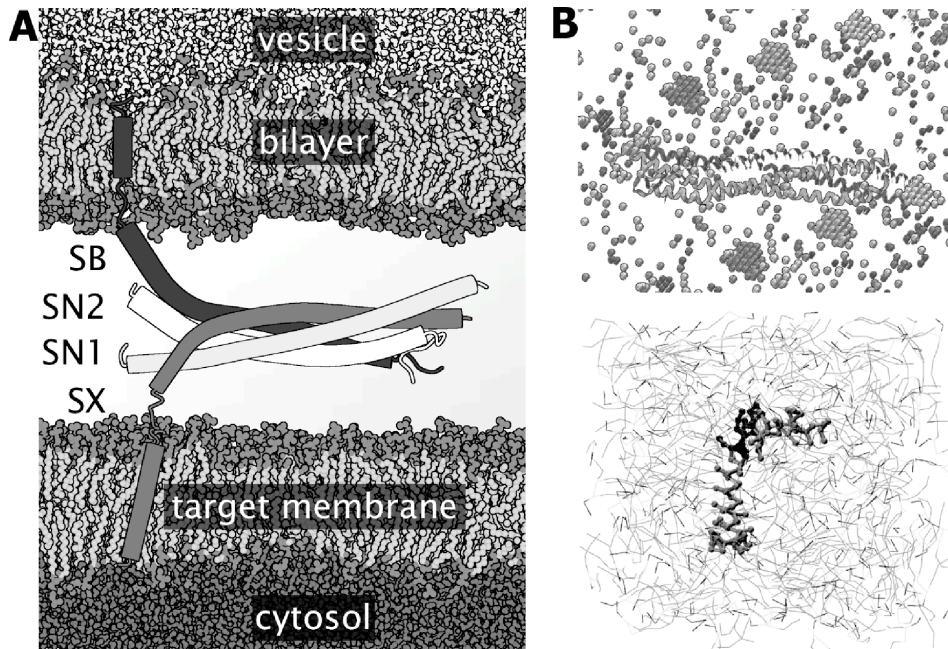


Figure 1. A: Illustration of the simulation system. The atomistic model comprises 339 792 atoms consisting of four proteins (346 amino acids), two charged POPC:POPS lipid bilayers (1008:123 lipids), 296  $\text{Na}^+$  ions, 166  $\text{Cl}^-$  ions and 92 217 water molecules. The simulation production runs were carried out with the Gromacs software<sup>3</sup>. B: Other simulation approaches<sup>2a,b</sup>: all-atom molecular dynamics simulation of the cytosolic domain (top) and coarse-grained simulation of the transmembrane domain of Synaptobrevin (bottom).

aspects will not be discussed in the present paper. Here we will focus on the technical and computational aspects of the computer simulations that were carried out.

Several challenges exist from a computational point of view. First of all, no single simulation will suffice to fully understand such a complex biological system as exists around the SNARE complex. It is necessary to perform several types of simulations, each addressing a specific aspect of the whole system. Thus we are also pursuing simulations on the cytosolic soluble domain of the fusion complex and coarse-grained studies of the transmembrane domains (Figure 1B)<sup>2a,b</sup>. In the present paper we will discuss a particularly ambitious simulation of a single SNARE complex embedded in a double lipid bilayer. This challenging simulation was enabled via the DEISA Extreme Computing Initiative. Biological systems - and membranes in particular - are 'soft matter' with a delicate balance of forces and interactions. The size of the model is important and long simulation times are necessary. Memory and disk space requirements further add to the complexity and can impose changes to existing software for efficient processing of the data.

A brief overview of the overall project organisation is given in Sec. 2. The setup of this simulation is outlined in Sec. 3. Computational performance and benchmark results are discussed in Sec. 4. We conclude with an outlook on possible improvements and extensions in Sec. 5.

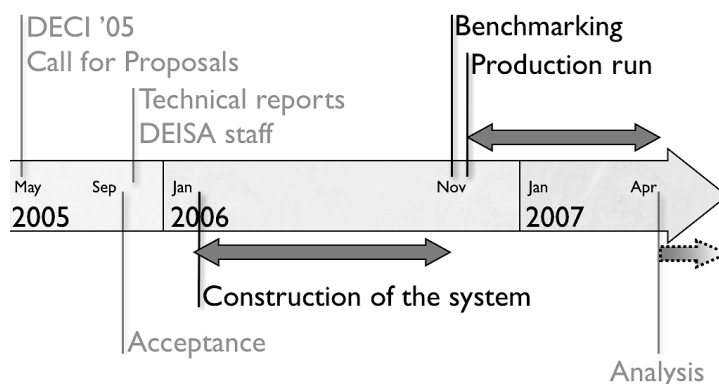


Figure 2. Timeline and project overview.

## 2 Timeline / Project overview

The project started 24 months ago with a substantial grant of computing time (see Figure 2). In collaboration with the DEISA staff, we first assessed the technical requirements of the simulation by analyzing the potential runtime characteristics. Next, we started with the construction of the atomistic model. This part, which required a little more than 12 months, will be described in more detail in the next section. Benchmarking was then achieved within 1-2 weeks, after which the production runs started for a duration of four months. Analysis of the simulation is currently under way and we expect it to become the longest part of the project. It should be pointed out that the initial lead time needed for setting up the simulation system cannot be neglected. It took longer than the actual production run and is an integral part of the challenges that one has to face when modeling complex 'soft matter' systems.

## 3 Simulation Setup

As mentioned in Sec. 2, the initial construction of the simulation system was one of the longest tasks of the project. In this section we describe the overall procedure that was adopted, then we provide details on one important sub-step.

We set up a full atomic model in explicit solvent, inserted into two fully hydrated mixed lipid bilayers consisting of 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) and 1-palmitoyl-2-oleoyl phosphatidylserine (POPS). The construction process is depicted in Figure 3. From left to right: 1/ We started from pre-existing all-atom simulations for the soluble cytosolic domain and the two transmembrane domains Synaptobrevin (Sb) and Syntaxin (Sx)<sup>2a,c</sup>. The system construction was guided by data from AFM experiments<sup>4</sup>, aiming at an approximate 50 Å separation of the two membranes, a rather short distance compared to the size of vesicles and cells. The fragments were brought into a suitable arrangement. 2/ In order to connect the cytosolic and transmembrane fragments, it was necessary to stretch the core complex. We thus carried out adaptive biasing force runs<sup>7</sup> combined with subsequent interactive molecular dynamics<sup>8</sup>. This will be described in more

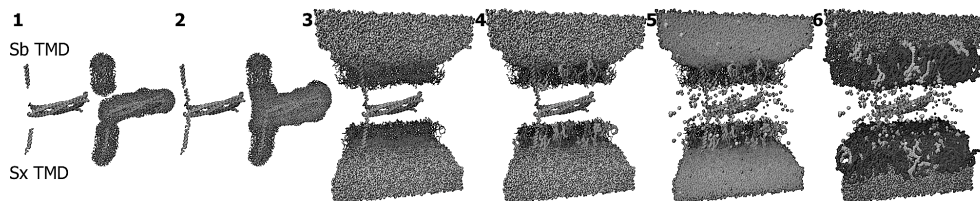


Figure 3. Stepwise construction of the starting system.

detail in the next subsection. 3/ We then created two larger hydrated POPC lipid bilayers spanning the whole simulation system. 4/ Subsequently these bilayers were mutated into mixed charged bilayers by replacing one in eight POPC lipids with POPS. 5/ Neutralizing counter ions were introduced by mutating water molecules into ions. 6/ The final structure with a mixed bilayer of 11% POPS and a 0.1 mol/l NaCl solution was equilibrated during several nanoseconds of molecular dynamics. The GROMOS-87 forcefield<sup>5</sup> with additional POPS parameters<sup>6</sup> was used. Simulations were carried out during 40 ns using full electrostatics with the Particle Mesh Ewald (PME) method. Such a timescale is short with respect to biological processes, but currently limited by the necessary computing time. Compared to other simulations of biological systems of this size, it is one of the longest simulations reported so far.

### 3.1 SNARE separation via Adaptive Biasing Force simulations

The Adaptive Biasing Force approach<sup>7</sup> was used to determine the potential of mean force (PMF) for the separation of the four-helical SNARE bundle and to drive the C-termini apart so that they connect to the transmembrane domains. We chose the distance between the C-termini of Sb and Sx as reaction coordinate. The PMF was refined in three successive runs with a total sampling time of three nanoseconds. Given the elongated shape of the complex, the fully solvated system requires a large solvation box and amounts to 140 000 atoms. It was simulated with the NAMD software<sup>9</sup> using the CHARMM forcefield<sup>10</sup>. The high observed pulling speed of 0.1 Å/ps in this exploratory simulation contributes to the observed loss of secondary structure (Fig. 4B) at the termini. Nevertheless the overall structure of the complex is little perturbed and only the connecting parts moved. Visual analysis shows an unzipping of the C-termini for about 1/4 of the length of the complex with the destruction of the knobs-into-holes packing of the helices. As illustrated by the PMF, the complex shows elastic behavior under the simulation conditions. The contact interface between Sb and Sx shown on Fig. 4C is remarkably unaffected by the C-terminal perturbation.

## 4 Computational Performance

### 4.1 Benchmarks

Benchmarks were carried out on an IBM SP Power 4 system at the national DEISA site IDRIS. A short 0.2 ps test run was performed for up to 128 processors. The latest stable

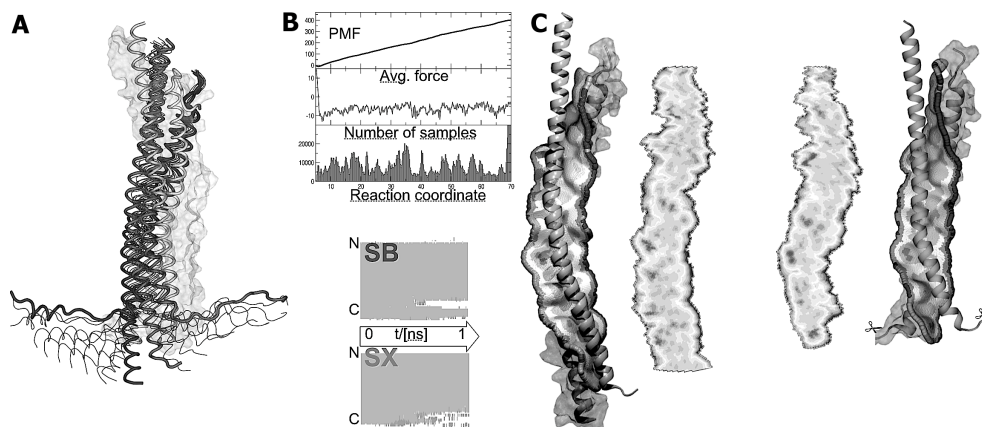


Figure 4. A: Cumulated snapshots during SNARE separation with initial and final structures highlighted. B (top): ABF estimated PMF, average biasing force and number of samples as a function of the reaction coordinate. B (bottom): secondary structure of Sb and Sx as function of time. C: Sb/Sx contact interface at the beginning and end of the ABF simulation. On the right the N-termini are omitted for clarity.

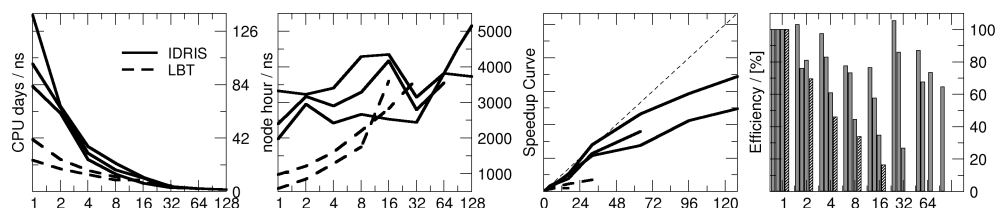


Figure 5. Benchmarks carried out at the IDRIS supercomputer center and locally at the LBT laboratory. From left to right: CPU days per nanosecond (ns), node hours per ns, speedup curve and computational efficiency.

and development versions of the Gromacs software<sup>3</sup> were used. Figure 5 shows that the CPU days per nanosecond ratio decreased rapidly up to 16 processors, then the gain leveled off. A comparison of the calculation cost in node hours per nanosecond showed that up to eight processors, small in-house clusters are significantly cheaper than running on a supercomputer. Above 32 processors, it is preferable to run at a dedicated computing infrastructure. The speedup curve revealed a degradation in scaling beyond 64 processors. The efficiency for a 96-processor run was a little above 60 percent. With such a setup, a 50 ns simulation would take 98.4 days and consume 227 000 CPU hours.

## 4.2 Production runs

The very first production runs were carried out locally at the LBT laboratory. The main CPU time resources for this DECI project were allocated at the Rechenzentrum Garching, where subsequent runs were submitted. There we achieved a stable simulation throughput as shown in Figure 6 over the whole four months period. An improved version of the Gromacs software<sup>3</sup> was used, which allowed acceptable scaling up to 96 processors. The

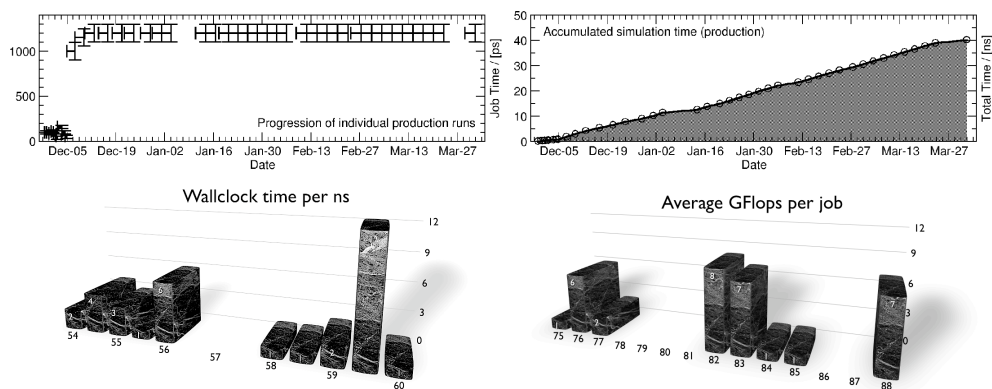


Figure 6. Top: History of job execution (left) and status of advancement (right) are shown in these graphs. The length of the line on the left plot corresponds to the execution time of a given job. The error bar on the ordinate is proportional to the number of processors used. Bottom: observed wall clock times for the production runs (left) and floating point performance (right).

simulated timescale was 40 ns corresponding to 20 million iterations. This equals a real timescale of 99 days and a total of 218 000 CPU hours. The overall performance was 30% slower than on the machine used for benchmarking. The average rate was 2.4 days (57 h) per ns, generating 75 GigaByte of compressed trajectory data.

Although we had access to a dedicated part of the supercomputer, the performance in a production environment was not completely stable, but varied by about 10%. A similar spread was observed for the floating point performance, with an average of 81.5 GigaFlops (min.: 74.9%; max.: 87.6%). Although we did not carry out detailed analysis to identify the source of the variability, it seems likely that network related issues are at its origin. The jobs were run on three P690 32-CPU nodes, with MPI communications between the three nodes sharing the same network as other jobs running on other nodes. I/O access through the Global File System also used the same network and influenced job performance. Both factors can lead to variations in performance depending on all other jobs running on the whole infrastructure at the same time.

## 5 Outlook

### 5.1 Calibrating coarse-grained simulations

Atomistic simulations may serve as reference for calibrating lighter, faster and more approximate coarse-grained simulations<sup>15</sup>. The computational gain is huge! The current 340 000 atom simulation of 40 ns duration remains a *tour de force* given the size of the system and the long simulation time. It consumed five months of computing time on 96 processors, whereas a corresponding coarse-grained calculation with 37 000 particles would run for one week on one processor and produce about 200 ns of trajectory.



## 5.2 Anchoring the SNARE complex deeper in the membrane

A molecular dynamics simulation of a double membrane with periodic boundaries is difficult, because water molecules cannot travel between the two resulting compartments. The distance between the two membranes is thus more or less fixed and needs to be chosen carefully. We started the simulation presented in this paper with a membrane distance of 100 Å (center to center) based on an estimation using Ref.<sup>4</sup>. After 40 nanoseconds, we found that one of the anchor helices had been partly pulled out of the membrane, indicating that the initial distance might have been slightly too large. We used YASARA<sup>11</sup> to move the membranes closer together in small steps of 0.5 Å, each one followed by a short steepest descent energy minimization and 200 steps of simulated annealing with the Yamber force field<sup>12</sup>. Force field parameters for the phospholipid molecules were derived automatically using AutoSMILES<sup>13</sup> in the framework of the GAFF force field<sup>14</sup>. The hydrogen bonds within alpha-helices were constrained to keep the SNARE protein fully intact. At a membrane distance of 86 Å, the procedure was stopped because the critical linker residues Trp 89 and Trp 90 had been buried again. This choice proved sensible, since so far the anchors stay firmly attached to the membrane (new production run; work in progress<sup>2b</sup>).

## 5.3 Analysis

The analysis of our simulations will focus on several key aspects such as the structural integrity of the four-helical SNARE bundle under tension, the membrane insertion of Sb and Sx, and the perturbation of the membranes by the fusion complex.

## 6 Concluding Remarks

We have described the setup, benchmarking and production calculations for a model of the synaptic fusion complex inserted into two adjacent fully hydrated lipid bilayers. It was observed that the initial setup and system construction represents an unpredictably long and delicate step. Benchmarking and production characteristics indicated machine-dependent differences and important fluctuations. The most time-consuming part in such a project is often the post-production and analysis phase. DEISA as a vast HPC infrastructure provided a stable environment for production runs and proved valuable for enabling large, challenging simulations.

## Acknowledgements

We are indebted to the DEISA staff for support throughout the planning and production phase of this project. We thank the French supercomputer centers IDRIS and CINES for computing time (Projects No 041714, 051714 and LBT2411). The most CPU-intensive computations were performed with a grant of computer time provided by the DEISA Extreme Computing Initiative 2006 and carried out at the Rechenzentrum Garching. MPD acknowledges support from the French Ministry of Research.



## References

1. (a) R Jahn, T Lang, and TC Südhof. Membrane fusion. *Cell*, 112(4):519–33, Feb 2003. (b) YA Chen and RH Scheller. Snare-mediated membrane fusion. *Nat Rev Mol Cell Biol*, 2(2):98–106, Feb 2001.
2. Unpublished results. (a) MP Durrieu, R Lavery, and M Baaden. (b) MP Durrieu, P Bond, R Lavery, MSP Sansom, and M Baaden. (c) N Taib, and M Laguerre.
3. <http://www.gromacs.org>
4. A Yersin, H Hirling, P Steiner, S Magnin, R Regazzi, B Hüni, P Huguenot, P De los Rios, G Dietler, S Catsicas, and S Kasas. Interactions between synaptic vesicle fusion proteins explored by atomic force microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 100(15):8736–41, Jul 2003.
5. WF van Gunsteren, and HJC Berendsen *Gromos-87 Manual* (Biomos BV, Groningen) 1987
6. P Mukhopadhyay, L Monticelli, and DP Tieleman. Molecular dynamics simulation of a palmitoyl-oleoyl phosphatidylserine bilayer with Na<sup>+</sup> counterions and NaCl. *Biophys. J.*, 86(3):1601–9, Mar 2004.
7. C Chipot, and J Hénin. Exploring the free-energy landscape of a short peptide using an average force. *Chem. Phys.*, 123(24):244906, Dec 2005.
8. J Stone, J Gullingsrud, P Grayson, and K Schulten. A system for interactive molecular dynamics simulation. In *ACM Symposium on Interactive 3D Graphics* (John F. Hughes and Carlo H. Squin, editors, New York, ACM SIGGRAPH), 191–194, 2001.
9. JC Phillips, R Braun, W Wang, J Gumbart, E Tajkhorshid, E Villa, C Chipot, RD Skeel, L Kalé, and K Schulten. Scalable molecular dynamics with namd. *J Comput Chem*, 26(16):1781–802, Dec 2005.
10. AD Mackerell, D Bashford, M Bellott, RL Dunbrack, JD Evanseck, MJ Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-Mccarthy, L Kuchnir, K Kuczera, FTK Lau, C Mattos, S Michnick, T Ngo, DT Nguyen, B Prodhom, WE Reiher, B Roux, M Schlenkrich, JC Smith, R Stote, J Straub, M Watanabe, J Wiorkiewicz-Kuczera, D Yin, and M Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, April 1998.
11. E Krieger, G Koraimann, and G Vriend. Increasing the precision of comparative models with yasara nova—a self-parameterizing force field. *Proteins*, 47(3):393–402, May 2002.
12. E Krieger, T Darden, SB Nabuurs, A Finkelstein, and G Vriend. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins*, 57(4):678–83, Dec 2004.
13. <http://www.yasara.org/autosmiles>
14. J Wang, RM Wolf, JW Caldwell, PA Kollman, and DA Case. Development and testing of a general AMBER force field. *J Comput Chem*, 25(9):1157–74, Jul 2004.
15. M Baaden, and R Lavery. There’s plenty of room in the middle: multi-scale modelling of biological systems in *Recent Adv. In Structural Bioinformatics, 2007* (Research Signpost India; Editor A. G. de Brevem).
16. <http://www.baaden.ibpc.fr/projects/snaredeci/>