Molecular dynamics characterization of the conformational landscape of small peptides: A series of hands-on collaborative practical sessions for undergraduate students

Article *in* Biochemistry and Molecular Biology Education · January 2016

DOI: 10.1002/bmb.20941

3 authors:

João P G L M Rodrigues
Stanford University
30 PUBLICATIONS **311** CITATIONS
SEE PROFILE

Adrien Melquiond
32 PUBLICATIONS **489** CITATIONS
SEE PROFILE

Alexandre M J J Bonvin
Utrecht University
297 PUBLICATIONS **10,070** CITATIONS
SEE PROFILE

Molecular Dynamics Characterization of the Conformational Landscape of Small Peptides: A series of hands-on collaborative practical sessions for undergraduate students.

João P.G.L.M. Rodrigues[1*], Adrien S.J. Melquiond[2*], Alexandre M.J.J. Bonvin[1*]

1. Computational Structural Biology Group, Bijvoet Center for Biomolecular Research, Faculty of Science-Chemistry, Utrecht University, Utrecht, Netherlands.
2. Biomedical genomics, Hubrecht Institute - KNAW & University Medical Center Utrecht, Utrecht, The Netherlands
* Corresponding Authors
Contact Information: j.p.g.l.m.rodrigues@gmail.com, a.melquiond@hubrecht.eu, a.m.j.j.bonvin@uu.nl

Keywords: *protein modelling, molecular dynamics, GROMACS, virtualization, jigsaw teaching*

Abstract

Molecular modelling and simulations are nowadays an integral part of research in areas ranging from physics to chemistry to structural biology, as well as pharmaceutical drug design. This popularity is due to the development of high-performance hardware and of accurate and efficient molecular mechanics algorithms by the scientific community. These improvements are also benefitting scientific education. Molecular simulations, their underlying theory, and their applications are particularly difficult to grasp for undergraduate students. Having hands-on experience with the methods contributes to a better understanding and solidification of the concepts taught during the lectures. To this end, we have created a computer practical class, which has been running for the past five years, composed of several sessions where students characterize the conformational landscape of small peptides using molecular dynamics simulations in order to gain insights on their binding to protein receptors. In this report, we detail the ingredients and recipe necessary to establish and carry out this practical, as well as some of the questions posed to the students and their expected results. Further, we cite some examples of the students' written reports, provide statistics, and share their feedbacks on the structure and execution of the sessions. These sessions were implemented alongside a theoretical molecular modelling course but have also been used successfully as a standalone tutorial during specialized workshops. The availability of the material on our web page also facilitates this integration and dissemination and lends strength to the thesis of open-source science and education.

Introduction

The last four decades of developments in computational biology and chemistry democratized molecular modelling and its applications [1,2]. Given the widespread usage of these computational methods in chemistry, structural biology and the pharmaceutical industry, there is an increasing demand for their inclusion in educational curricula [3,4]. The challenge lies, though, in attracting students towards computational experiments, since the core of molecular modelling and simulation is a combination of chemistry, physics, mathematics, and computer science, while its most noteworthy applications are in biology. While chemistry students usually have the necessary mathematical know-how to understand the equations behind the algorithms, they might lack the biology curriculum to understand the impact of the results and thus opt-out of such courses. Those coming from a biology background also tend to opt-out, although the reasons are usually tied with the mathematical concepts and difficulty of the course content and methodology. Further, in most cases, students in those undergraduate programs have little exposure to programming, scripting and working under a Linux environment. Not surprisingly, the educational community is divided about the best approach to teach computational biology at an undergraduate level [5]. In contrast to other authors, we believe that a molecular modelling course should offer sufficient in-depth knowledge while maintaining strong ties to biological applications. This is supported by the attendance numbers in our classes, offered to second-year undergraduate chemistry students, and by their performance at the final exam. Nevertheless, Dutch second year bachelor students have already a strong background in physics, mathematics, and (quantum) chemistry. Therefore, it is

easier to introduce rather advanced (practical) concepts of molecular modelling and molecular dynamics, such as those described in this article. In other settings, this set of practical lessons could perfectly fit at more advanced stages, such as a Master or early Ph. D. educational programmes.

Overview of the Practical Sessions

The sessions here detailed are integrated in a larger molecular modelling course that includes theoretical lectures and exercise sessions. Although not strictly required, these lectures and exercise sessions are advantageous as they expose the students to the mathematical concepts behind molecular modelling algorithms: the energy functions used to describe a molecular system at the classical level; and optimization and search techniques, such as energy minimization and molecular dynamics simulations. The practical sessions themselves are scheduled once a week, lasting four hours each. Given the workload of the students in our curriculum, the sessions stretch over six weeks (Figure 1). The length of the course can, however, be easily modified. In our case, we take advantage of a two-week Christmas break to run the simulations and therefore require the students to finish the preparation and setup by the fourth session. However, it is also possible to pre-compute all the simulations and therefore avoid these constraints, enabling for example, a week-long intensive course. The students are evaluated for their performance and interest during the sessions, as well as by a written report due after the course exam.

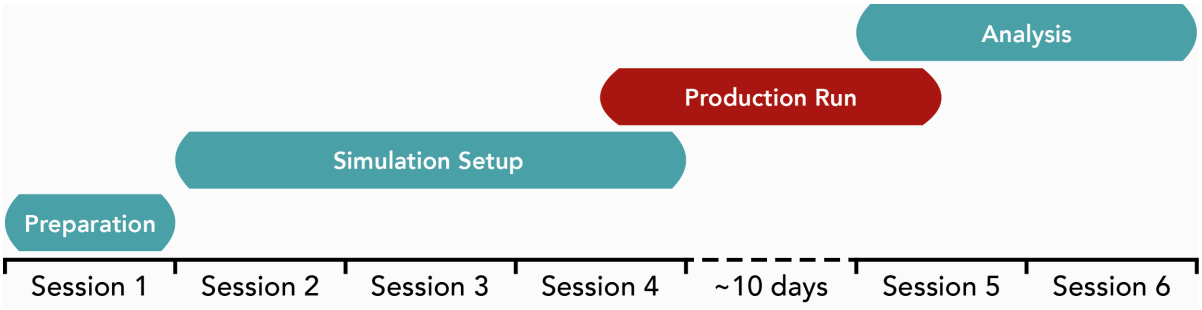


Figure 1. Schedule of the course: the 6 sessions are divided in 6 weeks in our curriculum, with a break of roughly 10 days to allow the simulations to complete. This break can be shortened given more computational resources, or skipped if the simulations are pre-computed and made available to the students.

In the first practical session, the students are posed the research question: does the binding event of a particular protein-peptide complex obey to the theory of conformational selection or does it experience induced fit? A conformational selection mechanism would imply that the peptide samples a number of conformations in its free state, including the one it adopts when bound to the protein receptor - the bound state. Induced fit, in contrast, implies that the peptide only adopts this bound state upon binding to its target protein and, therefore, this conformation will not be sampled in any simulation of the peptide alone. In order to find out, the students are asked to carry out extensive molecular dynamics simulations of the isolated peptide. If the peptide visits the bound conformation during the simulations, the answer to the research question is conformational selection; otherwise it leans towards induced fit.

To stimulate collaboration, the students are organized in groups of four, each group simulating a peptide, and each peptide being simulated by at least two groups. This setup, vaguely based on the ‘jigsaw technique’ as opposed to direct instruction, ensures not only the reproducibility of the results, but encourages the students to collaborate with other groups in order to complement their final reports. In a group, each student simulates the peptide using a different starting conformation in order to maximize the sampling of the conformational landscape. Besides the technical advantage, this also instils a sense of responsibility in the students: if their simulation is not properly set up, the work of the entire group is jeopardized. From our experience, this occurred very rarely as errors were caught early on by the students themselves, reflecting an internal peer-review process.

The material of the course is available online (http://nmr.chem.uu.nl/~adrien/course/molmod) and encourages the students to progress at their own pace. Given the time needed to run the simulations, the students are only given the deadline of having their peptide system ready for production simulation by the fourth practical - before the Christmas break. This means that students can opt to follow the material step-by-step while reading and understanding the concepts, or just prepare the system as quickly as possible and then spend time discussing with the instructors on specific sections or concepts that are of interest to them. This flexibility also releases the instructors from having to lecture. Instead, as we experienced, they focus more on troubleshooting and engaging the students in discussions, often individually or with a single group. Not only teaching becomes less monotonous, and so the instructors are more eager and motivated for the sessions, but the students also build stronger bonds with the instructors, which make communication easier and further motivate the students toward the material. All this is reflected in a very high satisfaction rate (93%) from the students' feedback at the end of the course, as well as personal mentions and praise to the instructors in the open questions.

Technical Requirements

The molecular dynamics simulations use the GROMACS [6] software and the AMBER99SB-ILDN force field [7]. This setup is in theory mutable, in particular the force field. An interesting experiment that could easily be added to the course would be to have two groups simulate the same peptide using different force fields and then compare the sampling. To facilitate teaching in the absence of dedicated computer laboratories, we prepared a virtual image compatible with the VirtualBox software (http://www.virtualbox.org). This image contains a Linux operating system and has GROMACS pre-installed, as well as a number of other programs including the molecular visualization package PyMOL (http://www.pymol.org) and the Xmgrace plotting tool (http://plasma-gate.weizmann.ac.il/Grace/). This setup allows for the deployment of a uniform system across all machines, usually the students' laptops, and simplifies installation and troubleshooting, at a cost of some performance.

Since the students' laptops vary in terms of hardware and computing power, we chose to run the production simulations somewhere else. During the several iterations of the course, we experimented with two different alternatives: running on our on-site cluster and running on a virtual cloud. The cloud service was provided by our national software infrastructure at SURFSara, under a NWO computing grant "Chemical Education in the Cloud" [8]. In this setup, the students accessed the cloud themselves and carried out all the production commands and monitoring of the simulation on their own. Given the availability and relatively low-cost of these services (e.g. Amazon EC2 provides 750 hours of computer time for free), this might be an option for those that do not have access to a local cluster. In the cloud setup, we used two cores per simulation, which translated in roughly 10 days of simulation time. When running on-site using 16 cores, the simulations complete within 24-36 hours, while on 8 cores and one NVIDIA GTX280 GPU card, they complete in less than a day. We expect these numbers to be consistent across the different cloud providers and even on high-end laptops, assuming of course reasonably modern processors.

The Sessions

As mentioned before, there is no well-defined order or checkpoints throughout the sessions. The students are only asked to complete the setup of the simulation after 4 sessions, to comply with the course schedule and the time needed to run the simulations. The following sections will use data from simulations on a nuclear receptor coactivator peptide (PDB ID: 2A3I) and offer an overview of what the sessions comprise, illustrating the steps of the protocol and highlighting those the students found more challenging. This system, one of the many available to the students, is part of a published protein-peptide docking benchmark [9][10]. Although not necessarily important for the course, we ask the students to write a small description that defines the biological significance of their system. For this particular protein-peptide complex, 2A3I, a group of students wrote: *"we are investigating a fragment of a larger peptide, a mineralocorticoid receptor. This protein controls sodium homeostasis and blood pressure through hormone binding and co-activator recruitment"*.

Setting up the simulation

Any molecular dynamics simulation requires an initial structure. All peptides used in the practical are available in the RCSB PDB [11] in the bound conformation together with the partner protein receptor. However,

structural "features" such as gaps in numbering, duplicate atoms (from partial occupancies in crystal structures), and cofactors, to list a few examples, are not trivially spotted by the untrained eye of a second-year bachelor student that never opened a PDB-formatted file before, and cause problems for the setup of the simulation. In particular, the 2A3I structure includes solvent atoms from the crystal, which must be culled before preparing the simulation. The naming of the water molecules might differ in the crystal and in the chosen force field (e.g. WAT or HOH vs. TIP) and cause problems during topology building. Another common issue in PDB files are missing atoms (in flexible regions), which prevent the topology from being built and must be modelled prior to setting up the simulations. As such, we provide processed structures of these complexes. Nevertheless, we walk the students through their own PDB file(s) in a text editor, paying special attention to the fixed-width format and the meaning of each field (http://deposit.rcsb.org/adit/docs/pdb_atom_format.html). In addition, we show and review some examples of the original PDB files and their problems.

Using PyMOL, the students visualize and explore the three dimensional structures of the protein-peptide complex, of the peptide itself, and its binding pocket. Guided by the online instructions, they extract the peptide in its bound conformation to a separate PDB file. Then, they manipulate the backbone dihedral angles of the peptide and generate three more structures: alpha-helix, beta sheet, and polyproline II. These ideal structures comprise the vast majority of peptide conformations observed in crystal structures of protein-peptide complexes [10] and consequently provide perfect initial states to explore the conformational landscape of peptides using molecular dynamics simulations. Afterwards, since all these manipulations are carried out in PyMOL, the students can easily compare and contrast the 4 conformations after structural alignment.

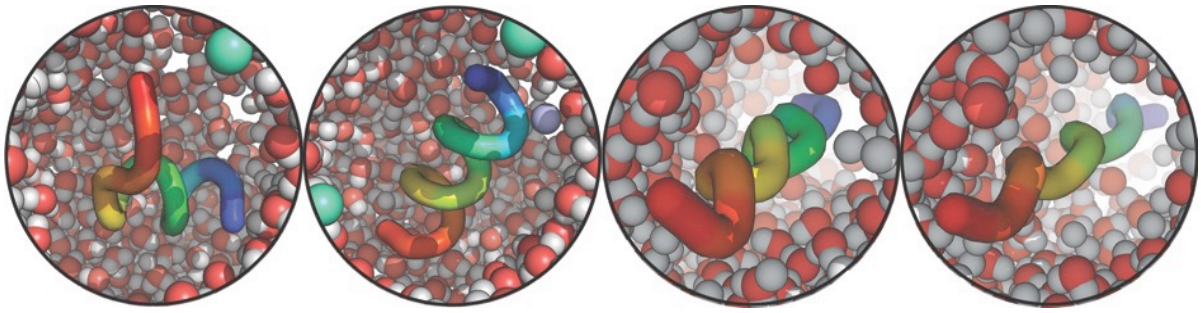


Figure 2. Initial conformations of the 2A3I peptide in a solvated simulation box. From left to right: bound, alpha-helix, polyproline II, beta sheet. The peptide residues are colored sequentially from red to blue. Water, sodium and chlorine molecules are represented by spheres, colored red/white, pale blue, and pale green respectively.

The four conformations - bound, alpha helix, polyproline II, and beta sheet - will seed four different molecular dynamics simulations (Figure 2). Making use of the virtual machine and preset image, each student creates a folder where they will store all the simulation setup data, and begins preparing the simulation. As mentioned before, all the instructions are available online, as well as an in-depth explanation of what entails each preparation step and its importance. Alongside these instructions, the online material also contains questions that the students *should* try to answer on their own, as they will help complete the written report due at the end of the course.

The preparation starts with creating a force field-compliant structure and corresponding topology. The students are asked to select the AMBER99sb-ILDN all-atom force field, which has been shown not to particularly favour certain secondary structures [7], but in principle any other force field would suit the course. Since the simulation will be carried out in water, the students also have to select the appropriate water model for the force field - TIP3P in this case. Besides these parameters, the students have to make the decision of capping (or not) the termini of their peptide. If the peptide is a fragment of a larger molecule, the termini should be capped in order to avoid artificial charge-charge interactions. Otherwise, the termini can be left charged. This information, whether the peptide is a fragment or not, is up to the students to find in the RCSB PDB entry for the complex

and represents their first challenge. A wrong choice will very likely lead to very different simulation results. For the 2A3I system, the termini must be capped, as the peptide is a fragment of a longer protein. The next step involves setting up periodic boundary conditions and choosing an appropriate box type and size. During the lectures that precede the practical, the students were exposed to the concept of periodic boundaries to simulate a large (infinite) system, as well as the importance of not letting the different periodic images interact with each other. As such, the students are instructed to create a rhombic dodecahedron box with a particular radius (1.2 nm) and are asked to check the unit cell volume. This is often a first point of discussion between the students and the instructors, in particular the importance of the radius and the differences between the different unit cells for the different conformations. After a short energy minimization in vacuum, the students can compare the minimized and initial structures and observe the differences, then proceed with solvating the unit cell as well as adding counter ions to neutralize the overall charge of the system. The salt concentration (Na+ and Cl-) is set to 0.15M, regardless of the system, to mirror physiological conditions. While most of the procedure is automated in GROMACS, there is a small correction that must be made manually to the topology file - balancing the number of solvent molecules and adding ions - that requires parsing the output files and understanding the format of the topology file. Although this step could in principle be automated, since this is likely the students' first contact with a command-line environment, we find it important they pay attention to each command they type and learn to interpret its output. Moreover, like this, the students correlate the peptide conformation, to the box size and the number of solvent and counter ions added. An extended conformation will create a larger simulation box, with more water molecules. This will also have an impact on the simulation performance, which the students realize later.

Conformation	Bound	Helix	Sheet	Polyproline II
No. of amino acids	12			
No. of protein atoms	209			
Charged Termini	No			
No. of solvent molecules	1856	1881	6972	5281
No. of counter ions	10	10	38	30
Total no. of atoms	5787	5862	16082	21163

Table 1. General information on the different conformations of the simulated 2A3I peptide

For the remainder of the preparation, the students perform molecular dynamics simulations on the system under a variety of conditions. At first, the system is heated to the desired temperature (310K) and it is simulated for a few thousand steps under NVT conditions (constant number of particles, volume and temperature) and position restraints on the heavy atoms of the peptide. This allows the solvent, ions, and hydrogen atoms of the peptide to relax further and adapt to the injection of velocities and consequently, kinetic energy. Being the first real dynamics step, the students are asked several questions and posed several challenges. For example: Why 310K degrees? How are the position restraints defined and on which atoms? What is the length of this particular simulation in picoseconds? What is the importance of setting a unique random seed value for velocity generation? What happens to the temperature during the simulation? These and other questions require the students to draw on their biology, chemistry, and mathematics knowledge, and apply a mix of those. In addition, the students gain familiarity with more computational file formats, such as those describing the topology and the position restraints, as well as plotting tools to help them analyze the thermodynamical properties of the simulation. To increase the realism of the simulation, pressure coupling is then added to the system, effectively changing the thermodynamical ensemble to isothermal-isobaric (NPT). In addition, the position restraints are progressively released until the system is completely free. This stepwise relaxation, involving several iterations

of simulations, exposes students to the care molecular systems need to avoid 'explosions' during the production runs. Further, the change from NVT to NPT introduces the concept of macroscopic properties and how these are controlled in a simulation environment. By plotting the pressure and density of the system throughout the equilibration simulations, and comparing these with the temperature variation, the students realize that while temperature is connected to the kinetic energy and therefore easily adjusted, pressure is coupled to the volume of the simulation box, and that its value fluctuates wildly around the preset (1 bar) (Figure 3).

Finally, before launching the simulation, the students are responsible for calculating of the number of steps necessary to perform 50 ns of simulation time. They then have to edit an existing *mdp* parameter file to insert this value and define others, such as how frequently to write positions and energies to disk, to optimize their simulation. Since the simulations run on the cloud, or on a local cluster, the instructors check the final simulation parameters before allowing the students to launch the production run. This is effectively the only external quality control step.

Simulation

The 50ns-simulation of a single peptide conformation takes roughly 10 days using two CPU cores. Depending on the timeframe and available resources, it is possible also to simulate the systems in advance and have the students move directly from the simulation setup to the analysis steps. However, this might decrease the sense of responsibility of the students, since regardless of their actions during the preparation there will always be a correct trajectory to analyze. Over the years, we were positively surprised by the minuscule amount of problems that eluded both ours and the students' checks. This low rate of problems indicates that both the material is robust and the students are paying attention to their actions and those of their colleagues. A particular case worthy of note happened when a student mistyped the sequence of peptide at the very first step - generation of an ideal conformation in PyMOL - and thus worked with a shorter peptide throughout the preparation sessions. This went completely unnoticed by either the rest of the group and the instructors, causing a problem during the analysis when the trajectories had to be merged. In this case, it helped that another group was running the exact same peptide.

Analysis

After the simulations complete, the students spend 2 extra sessions performing several quality assurance checks and extracting data from the simulations to answer the research question: induced fit or conformational selection? As a bonus, the students can extract representatives from their combined simulations and perform protein-peptide docking predictions to obtain models of the complex. These models constitute *per se* another readout for the students when they mimic the starting structure of the complex they used for this tutorial.

First, we ask the students to verify if their simulation completed, i.e. if the number of steps corresponds to what they set in the parameter file, and extract a few trivia from the log file, such as raw performance (ns/day), duration (days, hours), and what part of the calculation took the longest. From these values they realize, for example, that the more extended conformations - sheet and polyproline II - require a larger box with more water molecules and that this translates to a worse performance since the most time is spent in calculating solvent-solvent interactions (65-80%). A more in-depth quality assurance analysis follows, checking the minimum distance between periodic images, the values of temperature and pressure, as well as structural parameters such as radius of gyration and root mean square deviation of atomic coordinates to the average and starting structures. In addition, the students calculate a frame-by-frame comparison of each simulation trajectory and build an RMSD matrix to gauge conformational variability (Figure 4, upper-left panel). After this battery of tests, the students conclude if their simulation is valid (and if not, where exactly their preparation failed) and can gauge the conformational variability of the peptide during the simulation.

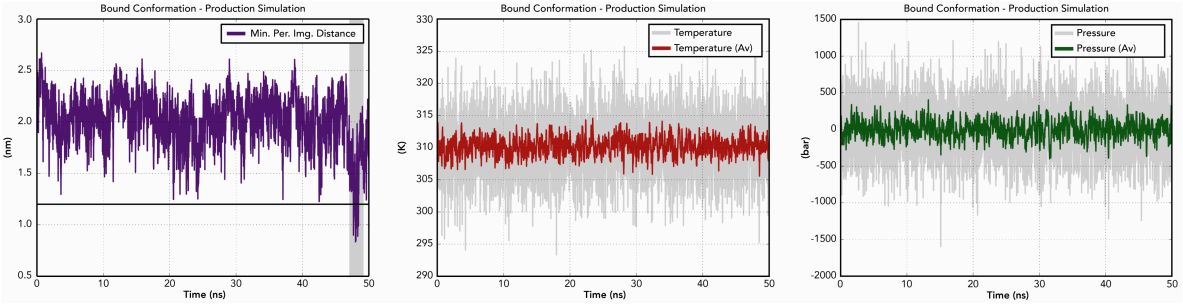


Figure 3. Minimum distance between periodic images (left), temperature (middle), and pressure (right) during the 50 ns production simulation of the bound conformation. Note the highlighted region (gray) towards the end of the simulation where the minimum distance between periodic images is smaller than the non-bonded cutoff (black line). This might invalidate this section of the simulation and indicate problems with the simulation setup. Note also the scale of the variations of temperature and pressure, substantially larger for the latter, reflecting the harder control of a macroscopic quantity.

The second and last part of the analysis consists on the calculation of several structural parameters, namely the number and type of secondary structure elements sampled during the simulation, Ramachandran plots for each residue, and hydrogen bond patterns between specific pairs of residues. Additionally, the 4 simulations of the group are merged and analyzed as a whole, after discarding the first 10 ns of each to account for an equilibration period. From this combined trajectory, the students are asked to evaluate the sampling of the different conformations and (if any) the degree of overlap between the difference trajectories, using an all-vs-all RMSD matrix (Figure 4, upper-right panel). In this simple visual representation of the conformational landscape, dark off-diagonal blocks indicate different frames with a similar conformation. The root mean square deviation to the native bound structure is also calculated from this trajectory, which allows finally to answer the research question (Figure 4, lower-right panel). The final piece in the analysis is a representation of the free energy landscape as a function of the radius of gyration and the root mean deviation to the average structure of the combined trajectory. We ask the students to build this plot using the software *Mathematica* [12], which allows them to visualize the "free energy landscape" in two (Figure 4, lower-left panel) and three-dimensions, and interactively rotate and zoom in particular regions. This representation complements the combined RMSD matrix, as regions with lower "free energy" indicate more popular conformations throughout the several trajectories. Further, using this landscape and an in-house Python script, the students can pick representatives of the simulation for an optional protein-peptide docking calculation.

As observed in the plots in figure 4, for this particular system, the native trajectory deviates slightly but rapidly from the starting structure, probably due to the instability of the bound conformation in the absence of the protein receptor. The helical trajectory, surprisingly, is extremely stable and similar to the crystallized bound structure and its trajectory. From the "free energy" landscape, it follows that this particular peptide seems to prefer compact conformations (lower radius of gyration). The secondary structure analysis yields further insights on the conformational dynamics of the peptides (Supplementary Figure 1). The bound structure is helical only in particular region, which remains mostly stable. The differences in RMSD stem therefore from the unstructured termini. The helix peptide remains helical throughout the trajectory, except the termini that show some flexibility. Interestingly, the extended peptide trajectory shows several residues with a "bend" and "turn" character, which might indicate the tendency of this region to adopt a helical structure. All in all, the peptide structures do not deviate substantially from the native, in particular the bound and helical trajectories. Answering the question of conformational selection vs. induced fit would, however, require that all 4 trajectories sample similar regions of the conformational space; due to the limited simulation time, this is not (yet) the case here. It is therefore not possible, from this analysis alone, to rule out induced fit. To some students, however, this is enough to make a stronger conclusion: *"the original peptide structure bound to the protein is formed without the protein being present, so it can be concluded that the conformational selection theory is supported by the simulations and that the induced fit can be dismissed."*. These different conclusions can have multiple explanations: different random seeds to generate initial velocities, which caused the

trajectories to explore different regions of the conformational space; the usage of different force fields (the course used GROMOS in the past); different (wrong) capping. Regardless, we grade the students based on their ability to critically argue in favor of their ideas, as long as they are supported by the simulation data. The students also can, and have done so in the past, opt to perform additional analyses to rule out particular scenarios.

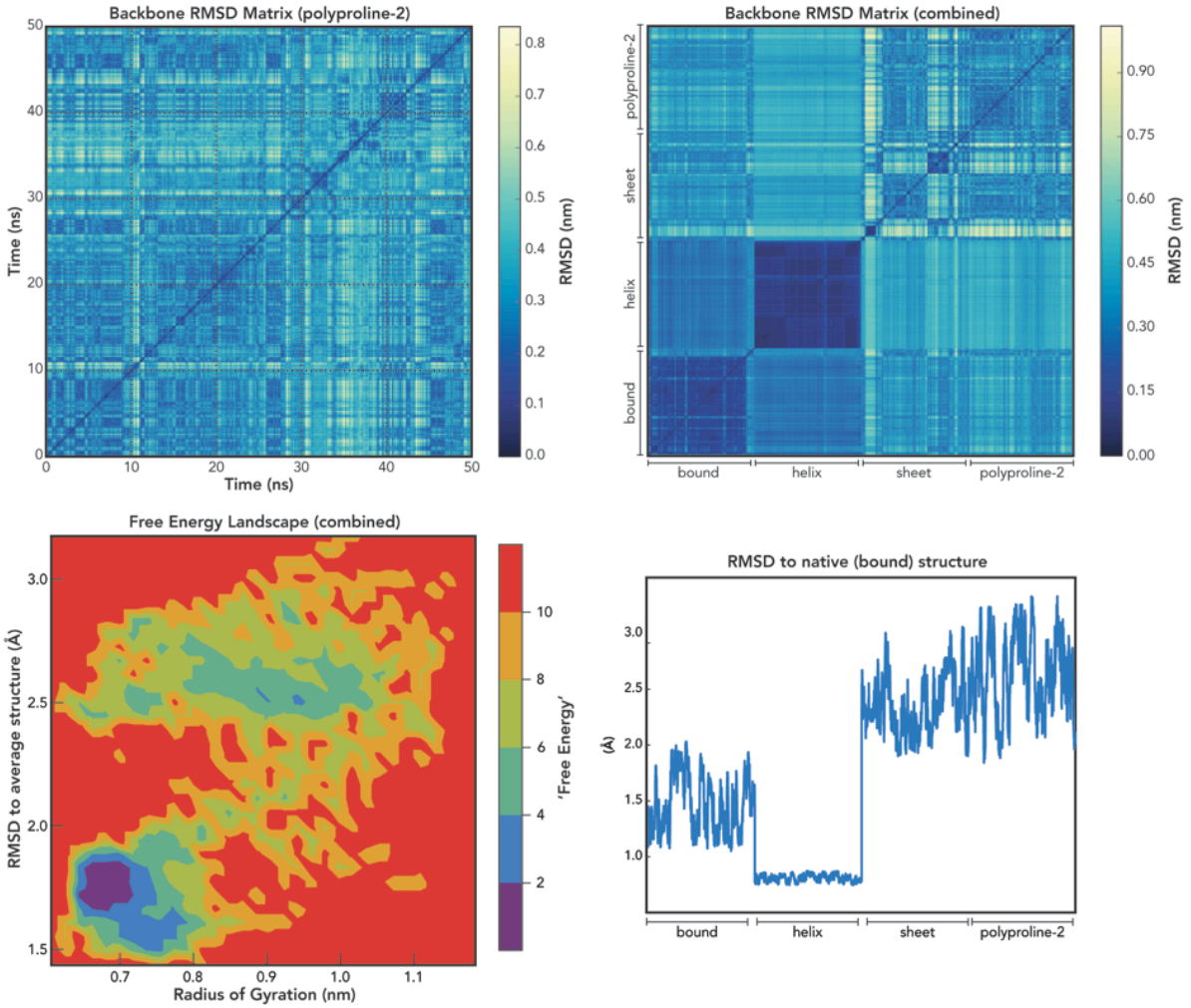


Figure 4. Conformational variability during the production simulations. Upper-left panel: All-vs-all comparison of trajectory frames in the polyproline II simulation; upper-right panel: all-vs-all comparison of the combined trajectories; lower-left panel: free energy landscape built using radius of gyration and RMSD to average structure. Lower free energy corresponds to more populated regions of the landscape; lower-right panel: RMSD calculation of each frame of the combined trajectory to the native bound structure.

Feedback

After each lecture, the students were asked to write a short paragraph reflecting on what they learned and what aspects they found most challenging. These ‘one-minute papers’ are a source of continuous evaluation, for both the lectures and the practical sessions, and help pinpoint problems with particular parts of the curriculum that can be reviewed the week after. The students also took this opportunity to convey some thoughts and opinions on the practical sessions, and on the link between them and the lectures, for example,

"I also wondered about the last slide [of the lecture], for which the practical session (...) was really useful. I see now the outline of energy minimization: why first it is important for the simulation, in vacuum and then in solvent." [translated].

Using an anonymous online survey, after taking the final exam, the students are encouraged to evaluate the course regarding its content, the lectures, and the practical sessions. Unfortunately, this course is taught together with a mathematics component and the evaluations cannot be entirely separated. Nevertheless, of the 15 students (~40% of the total enrolled in the course) that turned in the evaluations in the last iteration of the course, the majority felt challenged (86.7%) by the course content and 60% considered the study load adequate (33.4% considered it somewhat heavy). In the questions specifically targeting the practical sessions, nearly all the students (93.3%) reported them to be well led, with 60% awarding them the highest possible grade (5/5). In addition, in a set of open questions available at the end of the survey, several students highlighted the practical sessions and this setup as an 'outstanding' aspect of the course. The availability of the practical instructors to interact more with the students, due to the material being hosted online, was also noted by the students in several direct mentions as 'outstanding' aspects.

"I found the practical to complement the lectures very well. You get a better picture of what actually happens during modelling, which from the lectures themselves is sometimes hard to imagine." [translated].

"The practical gave a good idea of what the [modelling] profession entails." [translated].

"[The instructors] explained well (...) and were there frequently (...) helping anyone as soon as possible." [translated].

As for negative aspects and criticisms, the students often noted that a better introduction to the practical would help them understand the motivation behind the simulations. Also, both during the practical and in the surveys, several students indicated that the command-line environment was a major challenge. Both these criticisms are natural in the context of the students' background (Chemistry), but suggest improvements to take into account for the next iteration of the course.

Summary

This series of practical sessions learns students the details and important aspects of setting up, running, and analyzing a molecular dynamics simulation. The goal of the practical is clearly set at the beginning - conformational selection vs. induced fit - and is reminded often during the several steps of the protocol. Using a combination of GROMACS, PyMOL, and plotting tools, the students explore the conformational landscape of a peptide molecule and learn how to extract valuable biochemical information that provides further insight into the molecular system. Further, the setup of the sessions, combining online material, teamwork, and an active participation of the instructors, contributes to an enjoyable and positive learning experience. Finally, by making the material freely available online, we hope to encourage its application in other curricula and teaching environments.

Acknowledgments

The authors would like to thank all the practical instructors that helped in the course over the years for their enthusiasm and expertise, in particular drs. Ezgi Karaca and Tsjerk Wassenaar. In addition, the authors also thank the students for their patience and eagerness to learn, in particular those that completed the surveys and as such contributed to an increase in the quality of these sessions. Financial support from the Graduate School of Life Sciences of Utrecht University is acknowledged.

References

1. Schlick, T., Collepardo-Guevara, R., Halvorsen, L. A., Jung, S. & Xiao, X. Biomolecular modeling and simulation: a field coming of age. *Quarterly Reviews of Biophysics* **44,** 191–228 (2011).

2. Levitt, M. Birth and future of multiscale modeling for macromolecular systems (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **53,** 10006–10018 (2014).

3. Miller BT, Singh RP, Schalk V, Pevzner Y, Sun J, Miller CS, et al. Web-Based Computational Chemistry Education with CHARMMing I: Lessons and Tutorial. *PLoS Comput Biol* **10(7)**: e1003719 (2014).

4. Pevzner, P. A. Educating biologists in the 21st century: bioinformatics scientists versus bioinformatics technicians. *Bioinformatics* **20,** 2159–2161 (2004).

5. Cattley, S. A review of bioinformatics degrees in Australia. *Brief. Bioinformatics* **5,** 350–354 (2004).

6. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput* **4,** 435–447 (2008).

7. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **78,** 1950–1958 (2010).

8. http://www.nwo.nl/en/research-and-results/research-projects/65/2300179165.html

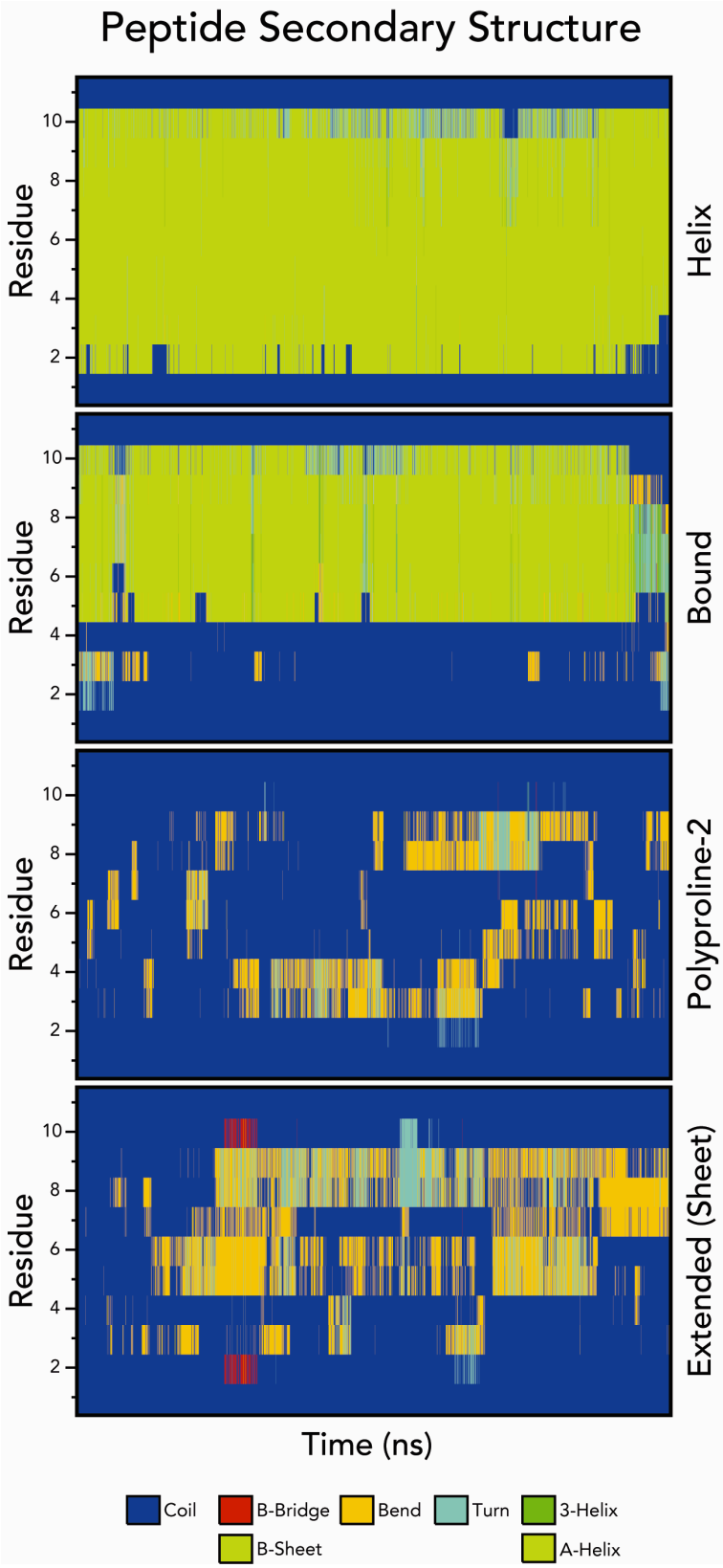
9. London N, Movshovitz-Attias D, Schueler-Furman O. The Structural Basis of Peptide-Protein Binding Strategies. *Structure* **18**:188–199 (2010).

10. Trellet, M., Melquiond, A. S. J. & Bonvin, A. M. J. J. A unified conformational selection and induced fit approach to protein-peptide docking. *PLoS ONE* **8,** e58769 (2013).

11. Rose, P. W. *et al.* The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res* **41,** D475–82 (2013).

12. Wolfram Research, Inc., Mathematica, Version 8.0, Champaign, IL (2010).

Supplementary Figure



Supplementary Figure 1. DSSP secondary structure analysis of the trajectories. The bound and helix trajectories show a central helical region that is stable throughout the simulation. The extended and polyproline-2 trajectories lack any helical residues but show some tendency for “bend” and “turn” in this central region.