YASARA Molecular Dynamics Trajectory Analysis for YourStructure

1. About the simulation

The trajectory **C:\Users\funaki\Desktop\tuto_md\5ns_normal\YourStructure** has been analyzed with YASARA version 26.3.5.W.64 over a period of 5.00 nanoseconds with 51 snapshots and the AMBER14 force field. Note that the MD simulation may have been run with a different force field, but AMBER14 was used to calculate the energies in this report. To change this, edit the ForceField setting at the start of this macro.

All plots and pictures in this report [like the simulated system below] are 1024 pixels wide, you can change the **figurewidth** variable in this macro as needed.

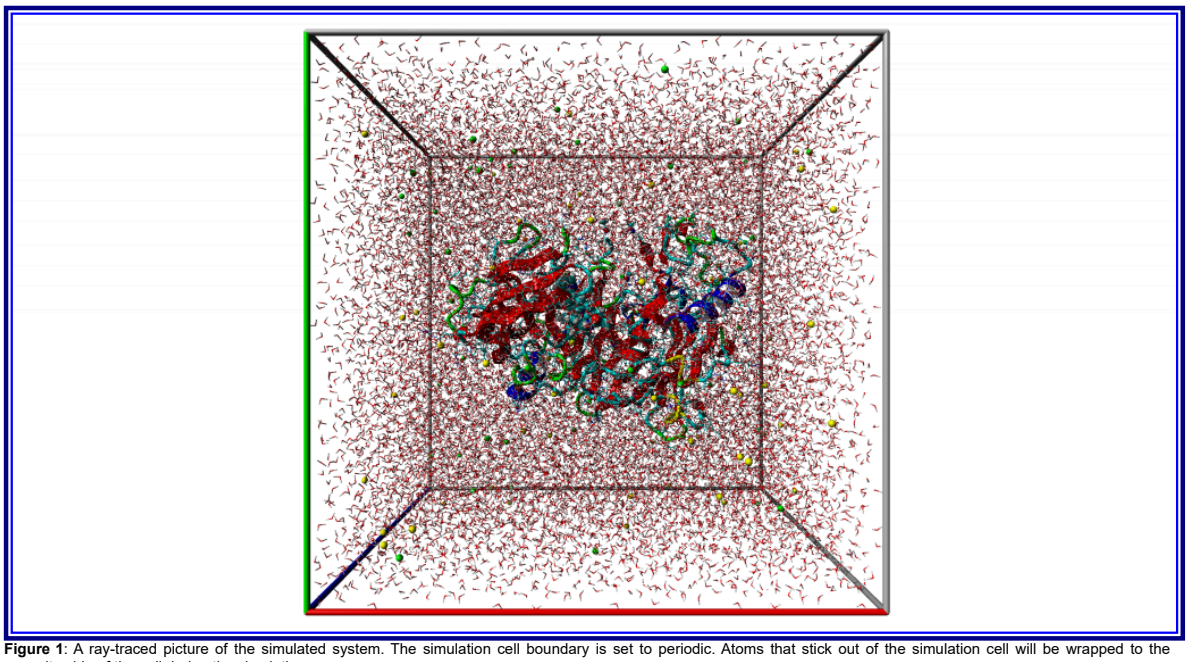

Figure 1: A ray-traced picture of the simulated system. The simulation cell boundary is set to periodic. Atoms that stick out of the simulation cell will be wrapped to the opposite side of the cell during the simulation.

1.1. Composition of the system

The components of the system are shown in the table below.

Type	Number
Protein molecules	3
Protein residues	370
Protein atoms	5766
Nucleic acid molecules	0
Nucleic acid residues	0
Nucleic acid atoms	0
Residue NME with 6 atoms	2
Residue ACE with 6 atoms	2
Residue "68M" with 63 atoms	1
Residue CIP with 1 atom of element Na	63
Residue CIM with 1 atom of element Cl	57
Water residues	21292
Total number of atoms	69849

Table 1: Composition of the simulated system

Object 1 with name **YourStructur** has been identified as the solute and is shown below. If this is not the intended solute, please change the **soluteobj** variable in this macro.

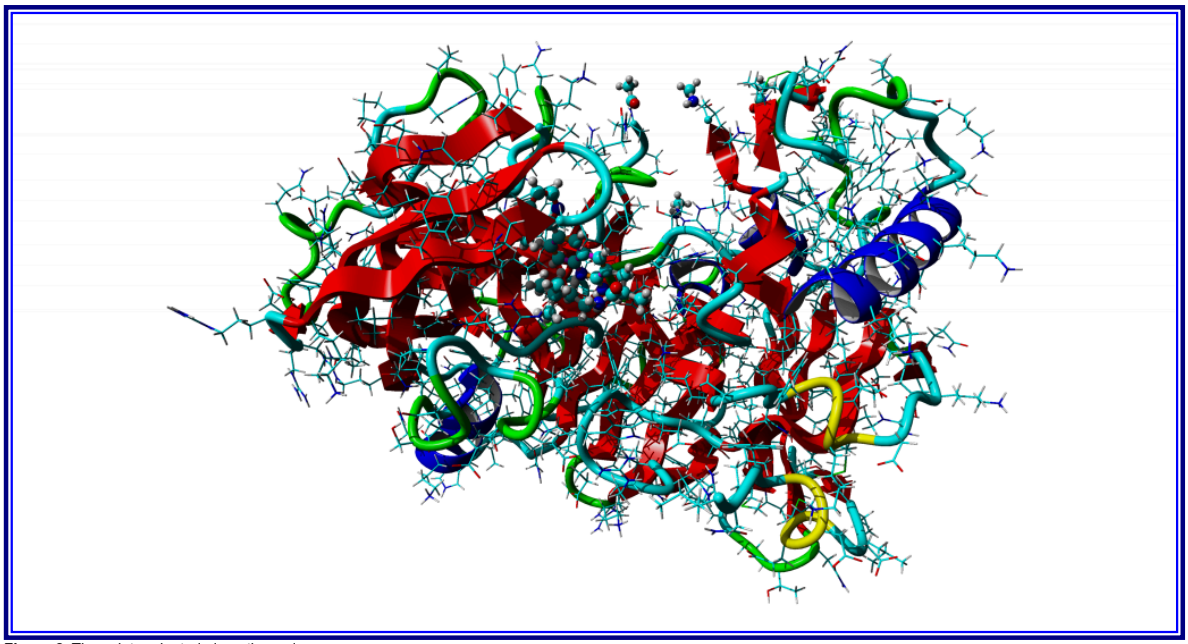

Figure 2: The solute oriented along the major axes.

1.2. The ligand

A special analysis has been performed for the ligand, chosen automatically by YASARA with the selection **Res "68M" Obj Solute**. The number of residues matching the ligand selection is 1, with 63 atoms. To change the ligand selection, edit the **ligandsel** variable at the beginning of this macro.

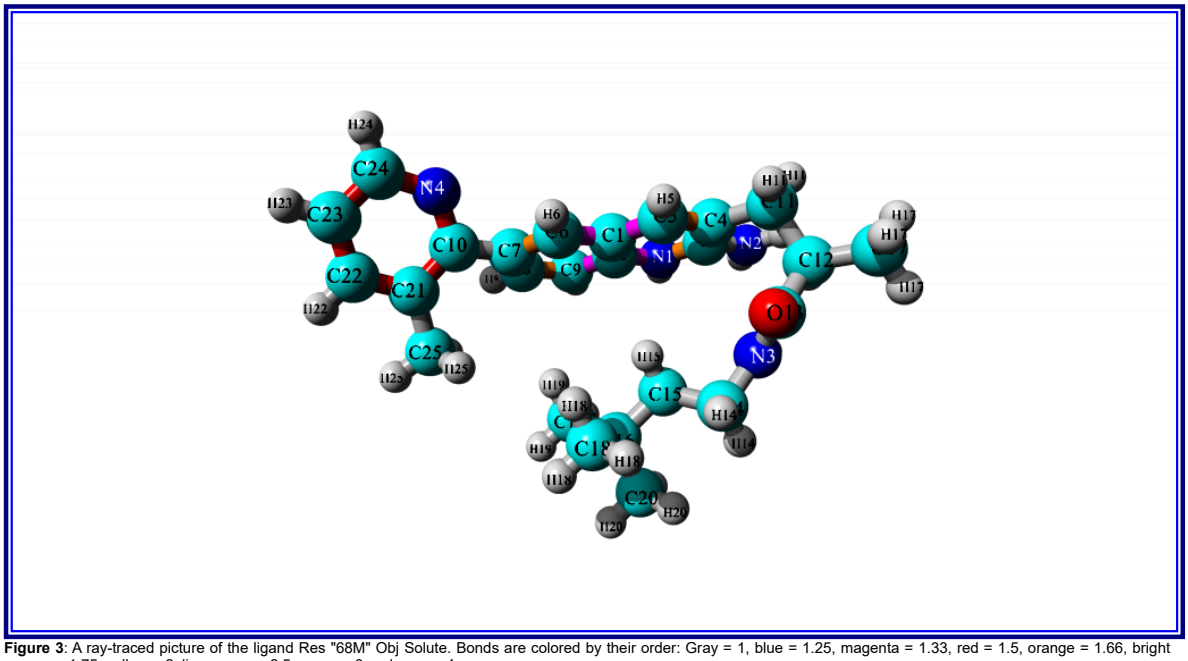

Figure 3: A ray-traced picture of the ligand Res "68M" Obj Solute. Bonds are colored by their order: Gray = 1, blue = 1.25, magenta = 1.33, red = 1.5, orange = 1.66, bright orange = 1.75, yellow = 2, lime green = 2.5, green = 3 and cyan = 4.

2. Analyses inside the simulation cell

This section shows all analyses that have been performed inside the simulation cell, when all atoms share the common coordinate system of the simulation cell.

Periodic boundaries are active and considered for distance measurements. Calculations that involve groups of atoms [center of mass, regression lines, enclosing spheres..] are ambiguous and should be placed in the next section, unless it is known that the atom group does not drift through a periodic boundary.

2.1. Simulation cell lengths

Conformational changes of the simulated solute molecules lead to fluctuations in density. If the simulation box has a constant size, changes in density lead to changes in pressure. This is not realistic, because molecules normally "live" in a constant pressure environment. During the simulation the cell is therefore rescaled to maintain a constant cell pressure. Depending on the chosen pressure control mode, the three cell axes are either rescaled together [Manometer1D], partly together [X- and Z-axes, Manometer2D, used for membrane simulations], independently [Manometer3D], or not at all [Off]. You can deduce the pressure control mode from the plot below.

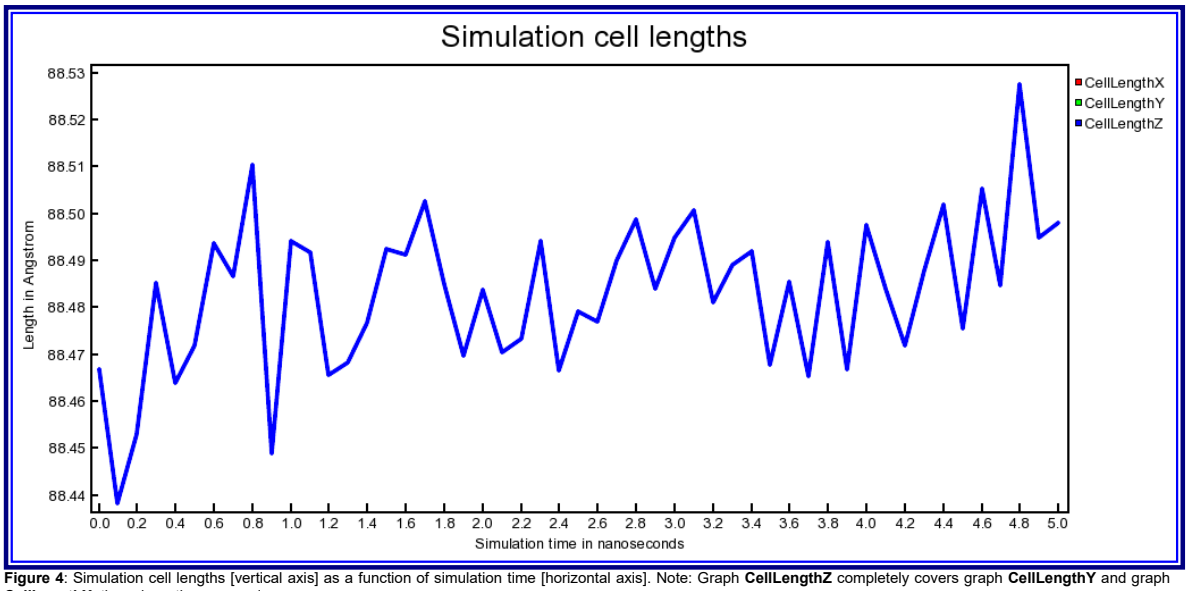

Figure 4: Simulation cell lengths [vertical axis] as a function of simulation time [horizontal axis]. Note: Graph **CellLengthZ** completely covers graph **CellLengthY** and graph **CellLengthX**, they share the same values.

2.2. Total potential energy of the system

The total potential energy of the system is plotted, according to the AMBER14 force field. If you ran the simulation with a different force field, you need to adapt the **ForceField** command at the top of this macro accordingly.

When the simulation is started from an energy-minimized "frozen" conformation, there is usually a sharp increase in energy during the first picoseconds, since the added kinetic energy is partly stored as potential energy. Also on a larger time-scale, the potential energy will often not decrease. A common reason are counter ions. These are initially placed at the positions with the lowest potential energy, usually close to charged solute groups, from where they detach to gain entropy, but also potential energy.

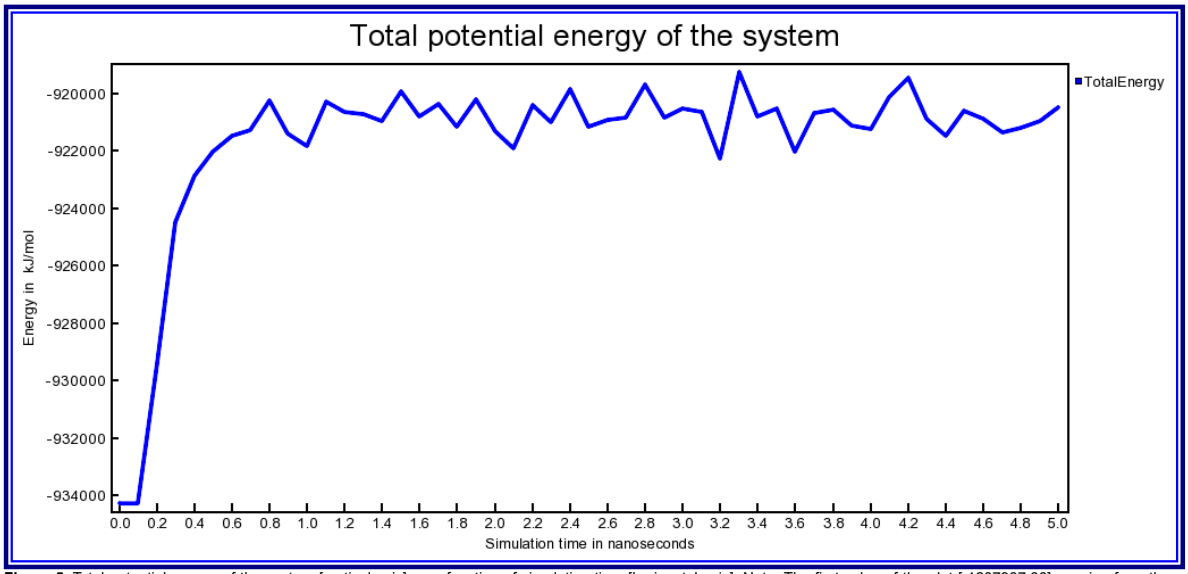

Figure 5: Total potential energy of the system [vertical axis] as a function of simulation time [horizontal axis]. Note: The first value of the plot [-1207967.66], coming from the energy minimized starting structure, has been replaced with the second value of the plot [-934244.89] to show this plot with a smaller energy range and thus a higher resolution.

2.3. Potential energy components

The following individual components of the total potential energy are plotted: bond energies [Bond], bond angle energies [Angle], dihedral angle energies [Dihedral], planarity or improper dihedral energies [Planarity], Van der Waals energies [VdW] and electrostatic energies [Coulomb]. Force field energies help to judge the structural quality of a protein: distortions of local covalent geometry can be found by looking at the bond, angle and planarity energies. Unrealistically close contacts [bumps] lead to a high Van der Waals energy, just like a large number of hydrogen bonds [since they pull the atoms closer than their normal Van der Waals contact distance]. The Coulomb energy is the least informative, because it strongly depends on the amino acid composition [e.g. proteins with a net charge have a higher Coulomb energy].

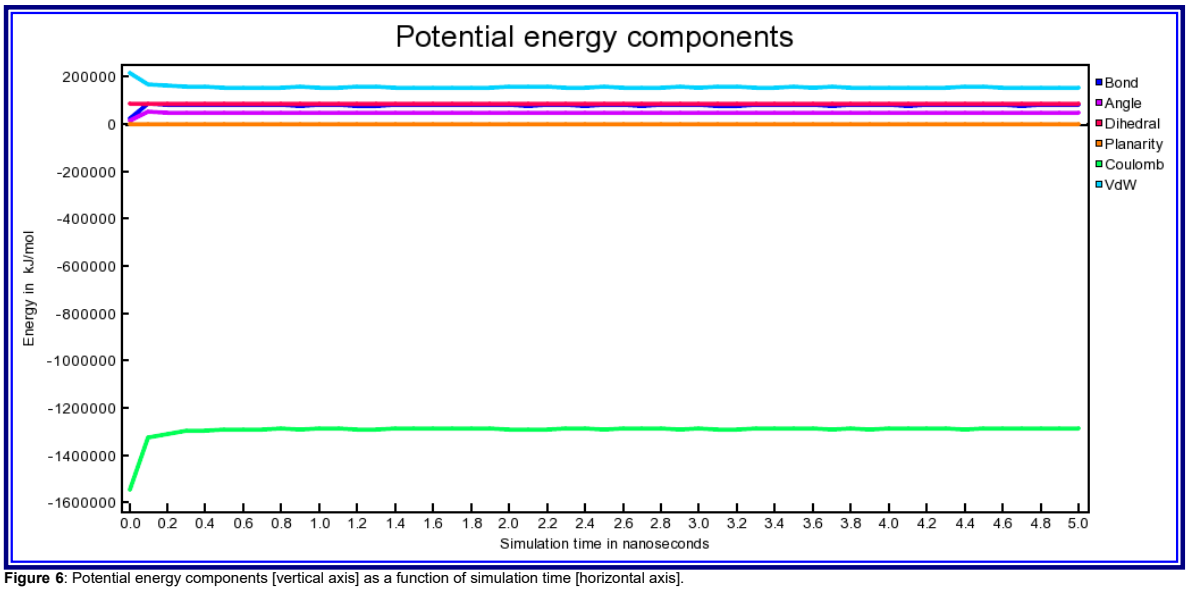

Figure 6: Potential energy components [vertical axis] as a function of simulation time [horizontal axis].

2.4. Surface areas of the solute

The Van der Waals [SurfVdW], molecular [SurfMol] and solvent accessible [SurfAcc] surface areas of the solute in A^2 are plotted. The difference between these surface types can be summarized as follows:

Van der Waals surface: if you think of atoms as spheres with a given Van der Waals radius, then the Van der Waals surface consists of all the points on these spheres that are not inside another sphere. In practice, the Van der Waals surface is of limited use, because it can be found throughout a protein and does not tell much about the interaction with the solvent.

Molecular surface: this is the Van der Waals surface from the viewpoint of a solvent molecule, which is a much more useful concept. The water is assumed to be a sphere of a given radius [also called the water probe], that rolls over the solute. Those parts of the Van der Waals surface that the water probe can touch are simply copied to the molecular surface [and called the contact surface]. Clefts in the Van der Waals surface that are too narrow for the water probe to enter are replaced by the Van der Waals surface of the water probe itself [and called the reentrant surface]. So the molecular surface is a smooth composition of two Van der Waals surfaces: the one of the solute and the one of the solvent molecule while it traces the contours of the solute. Other common names for the molecular surface are solvent excluded surface or Connolly surface.

Solvent accessible surface: this surface consists of all the points that the center of the water probe [i.e. the nucleus of the oxygen atom in the water molecule] can reach while rolling over the solute. The shortest possible distance between the water oxygen nucleus and a solute atom is simply the sum of the Van der Waals radii of the solute atom and the water probe.

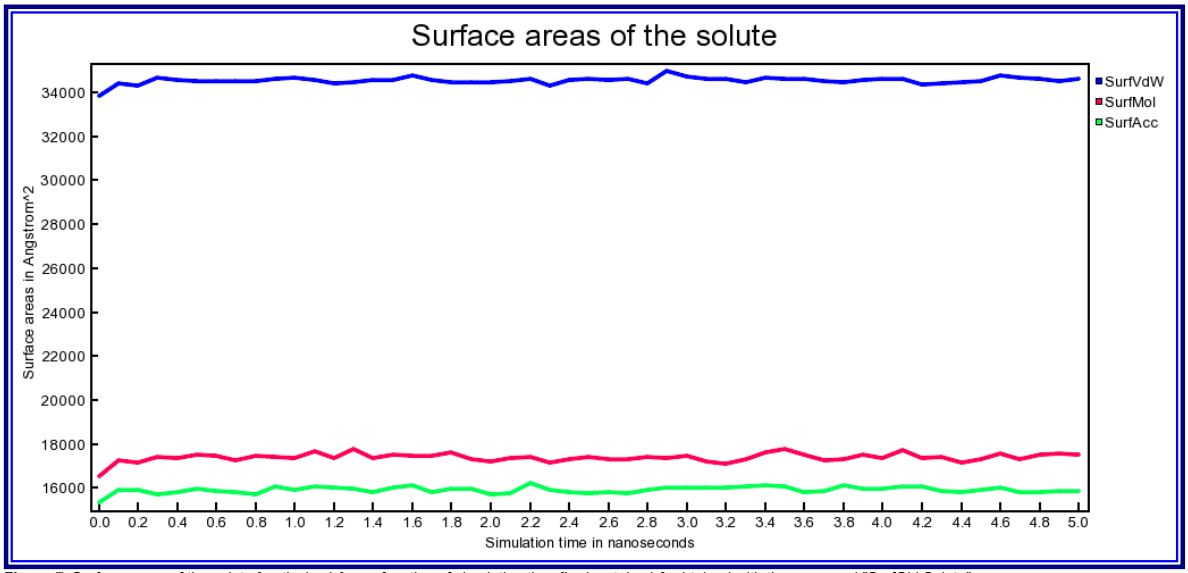

Figure 7: Surface areas of the solute [vertical axis] as a function of simulation time [horizontal axis], obtained with the command "SurfObj Solute".

2.5. Number of hydrogen bonds in the solute

The number of hydrogen bonds inside the solute is plotted below. One hydrogen bond per hydrogen atom is assigned at most, picking the better one if two acceptors are available.The following formula yields the bond energy in [kJ/mol] as a function of the Hydrogen-Acceptor distance in [A] and two scaling factors:

$$Energy_{HBo} = 25 * \frac{2.6 - max(Dis_{H-A}, 2.1)}{0.5} * Scale_{D-H-A} * Scale_{H-A-X}$$

The first scaling factor depends on the angle formed by Donor-Hydrogen-Acceptor:

$$Scale_{D-H\ A} = \begin{cases} 0 & \text{in range } 0..100 \text{ degrees} \\ 0..1 & \text{in range } 100..165 \text{ degrees} \\ 1 & \text{in range } 165..180 \text{ degrees} \end{cases}$$

The second scaling factor is derived from the angle formed by Hydrogen-Acceptor-X, where X is the atom covalently bound to the acceptor. If X is a heavy atom:

$$Scale_{H\ A-X} = \begin{cases} 0 & \text{in range } 0..85 \text{ degrees} \\ 0..1 & \text{in range } 85..95 \text{ degrees} \\ 1 & \text{in range } 95..180 \text{ degrees} \end{cases}$$

If X is a hydrogen, slightly smaller angles are allowed:

$$Scale_{H\ A-H} = \begin{cases} 0 & \text{in range } 0..75 \text{ degrees} \\ 0..1 & \text{in range } 75..85 \text{ degrees} \\ 1 & \text{in range } 85..180 \text{ degrees} \end{cases}$$

A hydrogen bond is counted if the hydrogen bond energy obtained with this formula is better than 6.25 kJ/mol [or 1.5 kcal/mol], which is 25% of the optimum value 25 kJ/mol.

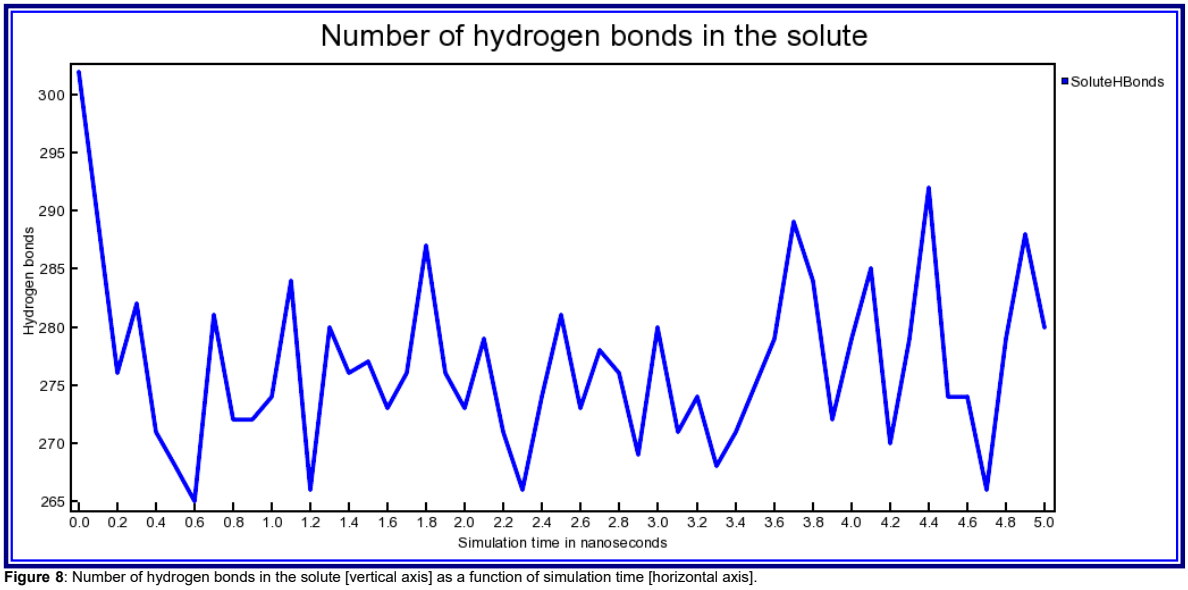

Figure 8: Number of hydrogen bonds in the solute [vertical axis] as a function of simulation time [horizontal axis].

2.6. Number of hydrogen bonds between solute and solvent

The plot shows the number of hydrogen bonds between solute and solvent. Together with the plot above, it is a good indicator for successful protein folding, indicated by a decreasing number of bonds with the solvent and a growing number of bonds within the solute.

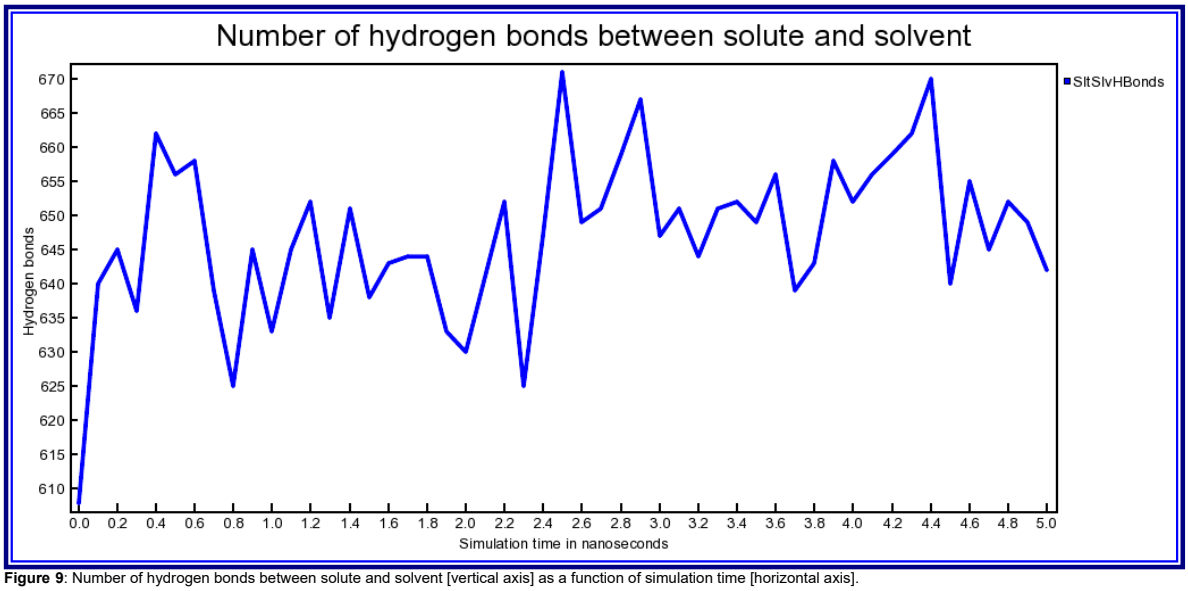

Figure 9: Number of hydrogen bonds between solute and solvent [vertical axis] as a function of simulation time [horizontal axis].

2.7. Number of hydrogen bonds made by Res "68M" Obj Solute

The plot shows three kinds of hydrogen bonds made by atoms selected with **hbosel**: Internal hydrogen bonds [blue], hydrogen bonds with the rest of the solute [green], those with the water [red] including membrane molecules if any, and the sum of all three [gray]. Internal bonds will be counted twice, once in each direction, so that the total number of bonds stays the same if alternative hydrogen bonds are formed. The green graph is an indicator for the completeness of a dissociation or docking event, assuming that there are hydrogen bonds involved at all. It is noteworthy that hydrogen bonds do not contribute significantly to the free energy of binding, as long as receptor and ligand can form alternative hydrogen bonds internally or with the solvent. Only if unsatisfied hydrogen bond donors or acceptors get buried, i.e. the sum of all hydrogen bonds shown in gray gets smaller, there is an adverse impact on the binding energy.

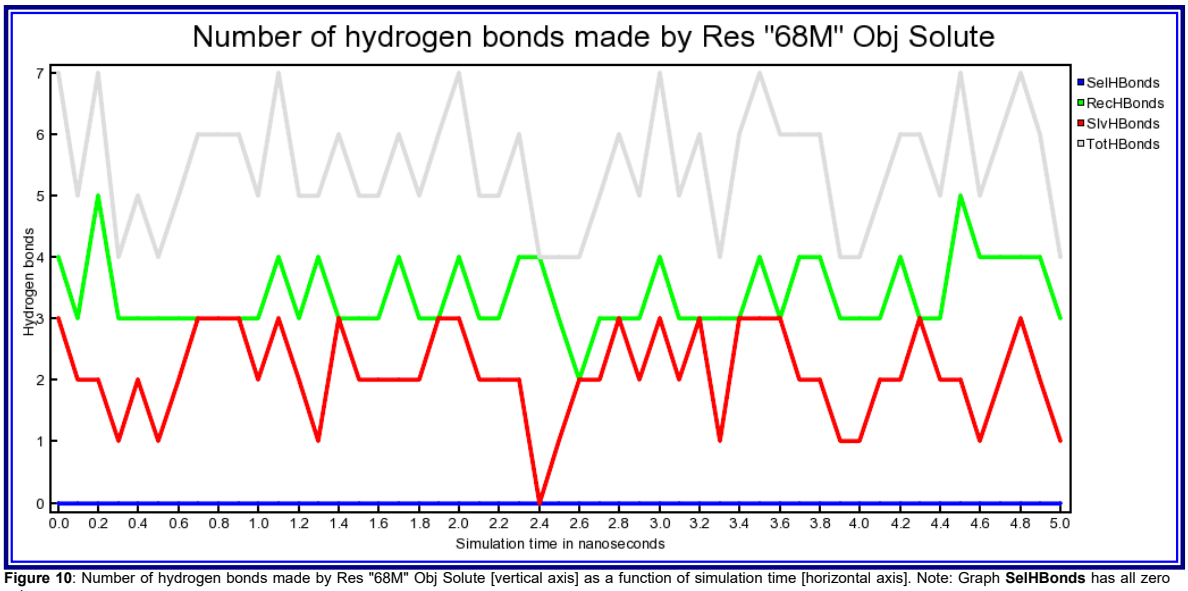

Figure 10: Number of hydrogen bonds made by Res "68M" Obj Solute [vertical axis] as a function of simulation time [horizontal axis]. Note: Graph **SelHBonds** has all zero values.

2.8. Hydrogen bonds made by Res "68M" Obj Solute

The following table shows all hydrogen bonds made by Res "68M" Obj Solute, excluding water molecules. With 5 acceptors and 4 donors, a total number of 14 hydrogen bonds are possible. 14 hydrogen bonds are listed - labeled HB1 to HB14. The donor of the bonding pair is labeled as Atm1 and the acceptor as Atm2, respectively. Internal bonds of Res "68M" Obj Solute will be listed twice, once in each direction. The atom ID separates atom name, residue ID and molecule name with dots. A lower-case "h" indicates hetgroups. E and D are short for the hydrogen bonding energy in [kJ/mol] and the distance between the bonding partners in [A]. To list other hydrogen bonds, edit the **hbosel** variable at the beginning of this macro. To list more or fewer hydrogen bonds, edit the **hbondsmax** variable at the beginning of this macro.

Time [ns]	HB1Atm1	HB1Atm2	HB1E	HB1D	HB2Atm1	HB2Atm2	HB2E	HB2D	HB3Atm1	HB3Atm2	HB3E	HB3D	HB4Atm1	HB4Atm2	HB4E	HB4D	HB5Atr
0.00	N3.h404.L	O.G34.R	13.80	2.13	N1.h404.L	OD2.D32.R	25.00	1.75	N2.h404.L	OD1.D32.R	25.00	1.78	N2.h404.L	OD2.D228.R	25.00	1.81	-
0.10	N1.h404.L	OD2.D32.R	17.00	1.75	N2.h404.L	OD1.D32.R	25.00	1.63	N2.h404.L	OD2.D228.R	21.50	1.98	-	-	-	-	-
0.20	N3.h404.L	O.G34.R	11.70	2.23	OH.Y71.R	O1.h404.L	20.32	1.83	N1.h404.L	OD1.D32.R	24.43	2.01	N2.h404.L	OD2.D32.R	21.50	1.72	N2.h40
0.30	N3.h404.L	O.G34.R	7.80	2.31	N1.h404.L	OD1.D32.R	23.45	1.88	N2.h404.L	OD2.D228.R	19.55	1.86	-	-	-	-	-
0.40	N3.h404.L	O.G34.R	8.45	2.22	OH.Y71.R	O1.h404.L	25.00	1.97	N1.h404.L	OD1.D32.R	22.88	1.90	-	-	-	-	-
0.50	OH.Y71.R	O1.h404.L	25.00	1.75	N1.h404.L	OD1.D32.R	16.63	1.93	N2.h404.L	OD2.D228.R	25.00	1.79	-	-	-	-	-
0.60	N3.h404.L	O.G34.R	12.88	2.12	N1.h404.L	OD1.D32.R	21.50	1.95	N2.h404.L	OD2.D228.R	11.93	1.98	-	-	-	-	-
0.70	N3.h404.L	O.G34.R	20.80	2.12	N1.h404.L	OD2.D32.R	20.32	1.81	N2.h404.L	OD1.D32.R	21.88	1.93	-	-	-	-	-
0.80	N3.h404.L	O.G34.R	14.65	1.95	N1.h404.L	OD2.D32.R	13.00	2.07	N2.h404.L	OD1.D32.R	23.85	1.70	-	-	-	-	-
0.90	N1.h404.L	OD2.D32.R	21.88	1.77	N2.h404.L	OD1.D32.R	25.00	1.91	N2.h404.L	OD2.D228.R	8.03	1.89	-	-	-	-	-
1.00	N3.h404.L	O.G34.R	21.88	2.00	N1.h404.L	OD2.D32.R	25.00	1.77	N2.h404.L	OD1.D32.R	23.85	1.75	-	-	-	-	-
1.10	N3.h404.L	O.G34.R	6.70	2.27	N1.h404.L	OD2.D32.R	25.00	1.77	N2.h404.L	OD1.D32.R	23.85	1.86	N2.h404.L	OD2.D228.R	17.68	2.11	-
1.20	N1.h404.L	OD2.D32.R	25.00	1.86	N2.h404.L	OD1.D32.R	22.88	1.81	N2.h404.L	OD2.D228.R	18.18	1.88	-	-	-	-	-
1.30	N3.h404.L	O.G34.R	21.10	2.06	OH.Y71.R	O1.h404.L	20.73	1.79	N1.h404.L	OD2.D32.R	25.00	1.70	N2.h404.L	OD1.D32.R	18.75	2.02	-
1.40	N1.h404.L	OD2.D32.R	25.00	1.75	N2.h404.L	OD1.D32.R	25.00	1.65	N2.h404.L	OD2.D228.R	21.10	1.89	-	-	-	-	-
1.50	N1.h404.L	OD2.D32.R	25.00	1.80	N2.h404.L	OD1.D32.R	25.00	1.66	N2.h404.L	OD2.D228.R	18.57	1.90	-	-	-	-	-
1.60	N1.h404.L	OD2.D32.R	25.00	2.00	N2.h404.L	OD1.D32.R	23.85	1.83	N2.h404.L	OD2.D228.R	14.48	1.87	-	-	-	-	-
1.70	N3.h404.L	O.G34.R	10.55	2.32	N1.h404.L	OD2.D32.R	22.88	1.87	N2.h404.L	OD1.D32.R	16.63	1.83	N2.h404.L	OD1.D228.R	19.93	1.78	-
1.80	N1.h404.L	OD2.D32.R	25.00	1.82	N2.h404.L	OD1.D32.R	25.00	1.68	N2.h404.L	OD1.D228.R	14.65	1.85	-	-	-	-	-
1.90	N3.h404.L	O.G34.R	12.80	2.23	N1.h404.L	OD2.D32.R	25.00	1.87	N2.h404.L	OD1.D32.R	18.75	1.80	-	-	-	-	-

Table 2: Hydrogen bonds made by Res "68M" Obj Solute as a function of simulation time [first column]. Note: At most 20 table rows are shown. Change the **tabrowsmax** variable in the macro t

2.9. Protein secondary structure content

The total percentages of alpha helices, beta sheets, turns, coils, 3-10 helices and pi helices are calculated and plotted. For clarification, a turn is simply a stretch of four residues that are not part of other secondary structure elements and form a hydrogen bond between the O of the first and the NH of the last residue. A coil is anything that does not fit into the other categories. Note that pi-helices [helices with hydrogen bonds between residues N and N+5] are rather unstable and thus do not normally occur in proteins, except for short bulges in alpha helices [which are often the result of single residue insertions and prolines].

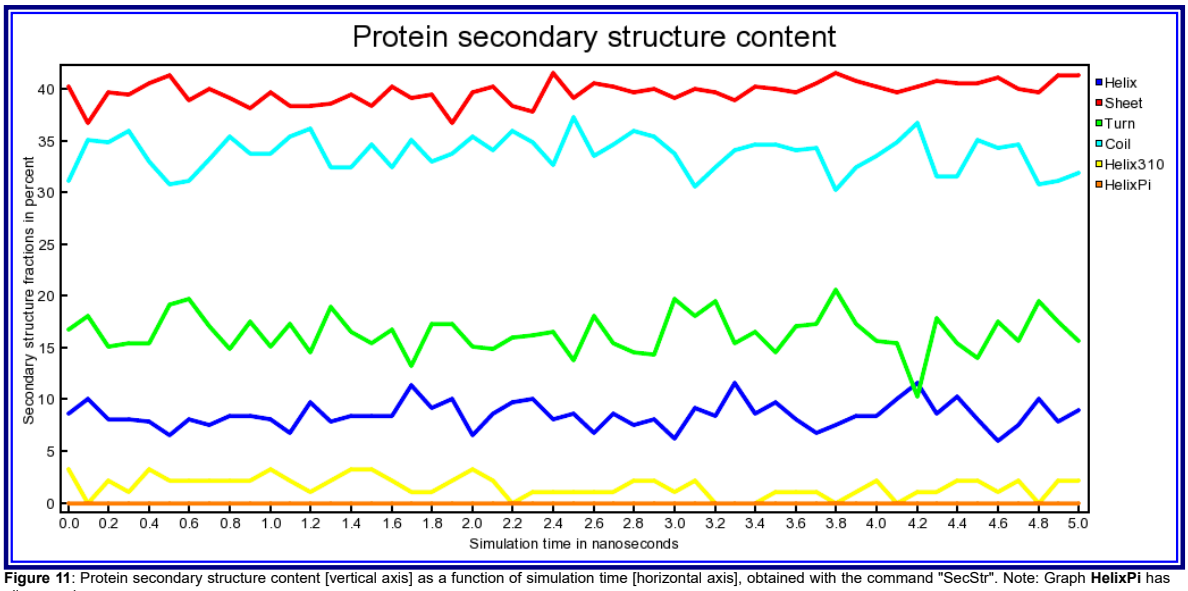

Figure 11: Protein secondary structure content [vertical axis] as a function of simulation time [horizontal axis], obtained with the command "SecStr". Note: Graph **HelixPi** has all zero values.

2.10. Per-residue protein secondary structure

The following plots show the protein secondary structure per residue as a function of simulation time. They are helpful to monitor protein folding and all other kinds of structural changes. The default secondary structure colors are used, you can change them at View > Color > Parameters > Secondary structure colors. One plot per protein molecule is shown.

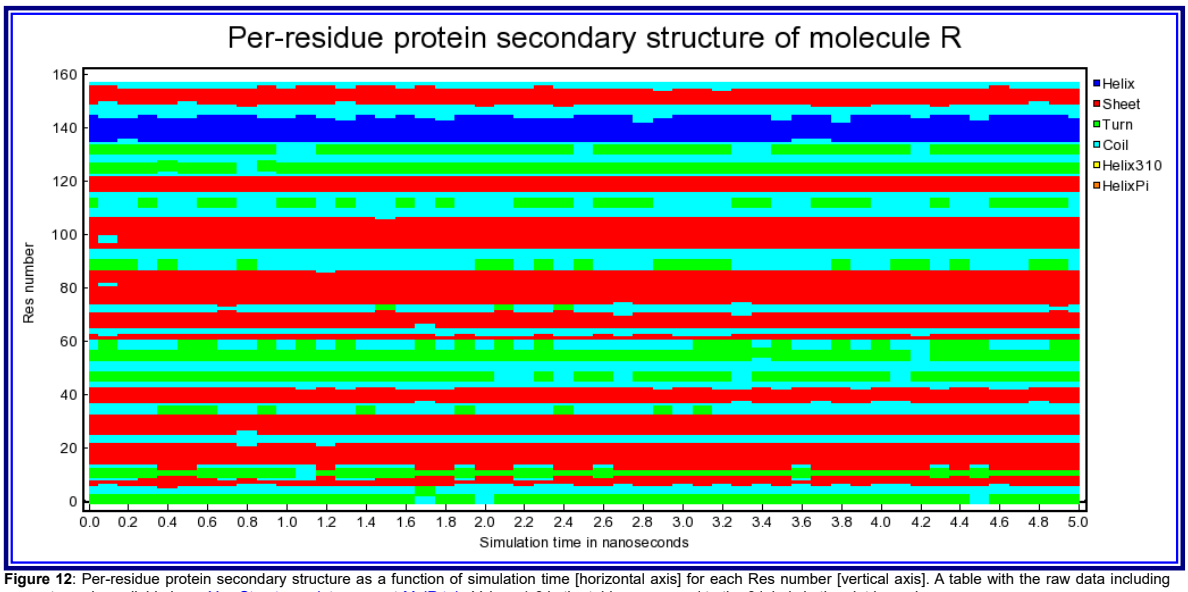

Figure 12: Per-residue protein secondary structure as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_secstrMolR.tab. Values 1-6 in the table correspond to the 6 labels in the plot legend.

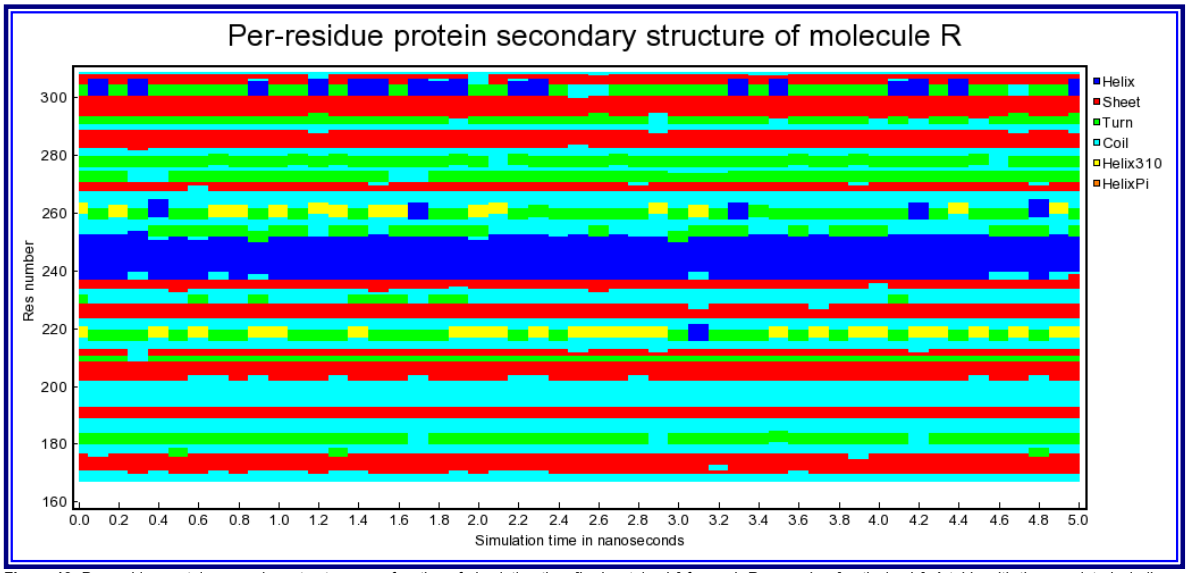

Figure 13: Per-residue protein secondary structure as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_secstrMolR.tab. Values 1-6 in the table correspond to the 6 labels in the plot legend.

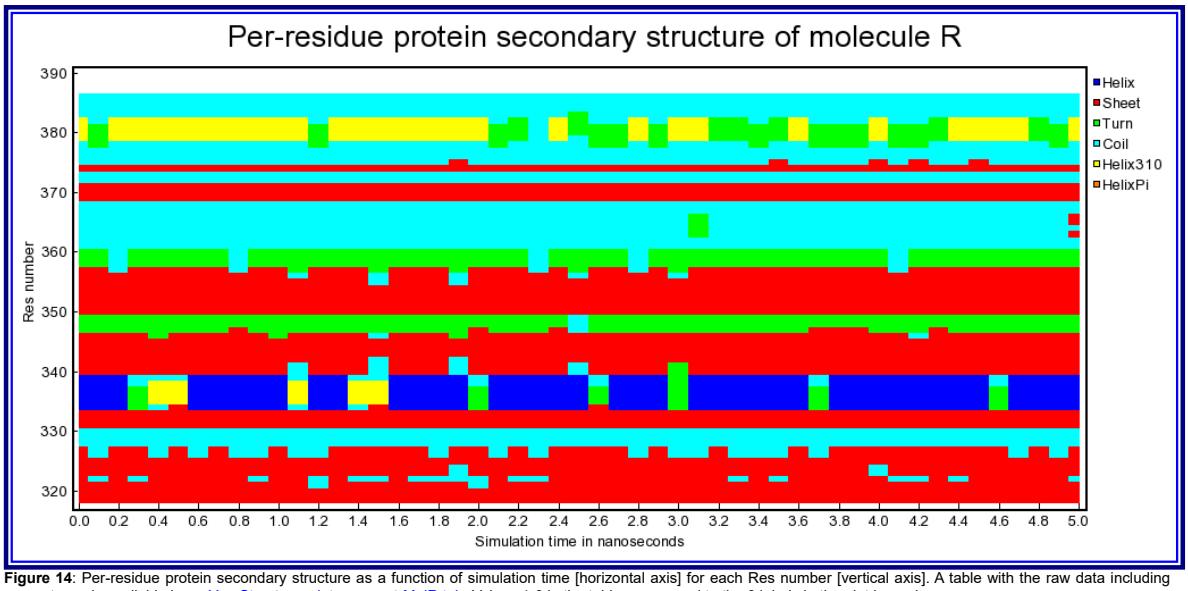

Figure 14: Per-residue protein secondary structure as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_secstrMolR.tab. Values 1-6 in the table correspond to the 6 labels in the plot legend.

2.11. Per-residue number of contacts

The number of contacts per residue as a function of simulation time is shown in the following plots. There is one plot for each protein or nucleic acid molecule. Even though contacts between atoms separated by up to four chemical bonds are excluded, neighboring residues in the molecule usually have enough close atoms to be counted as a contact. Consequently residues with zero contacts are very rare and often glycines. The number of contacts tells you how densely a certain residue range is packed and allows to identify structurally very important residues, e.g. a phenylalanine in the hydrophobic core can contact 15 or more other residues.

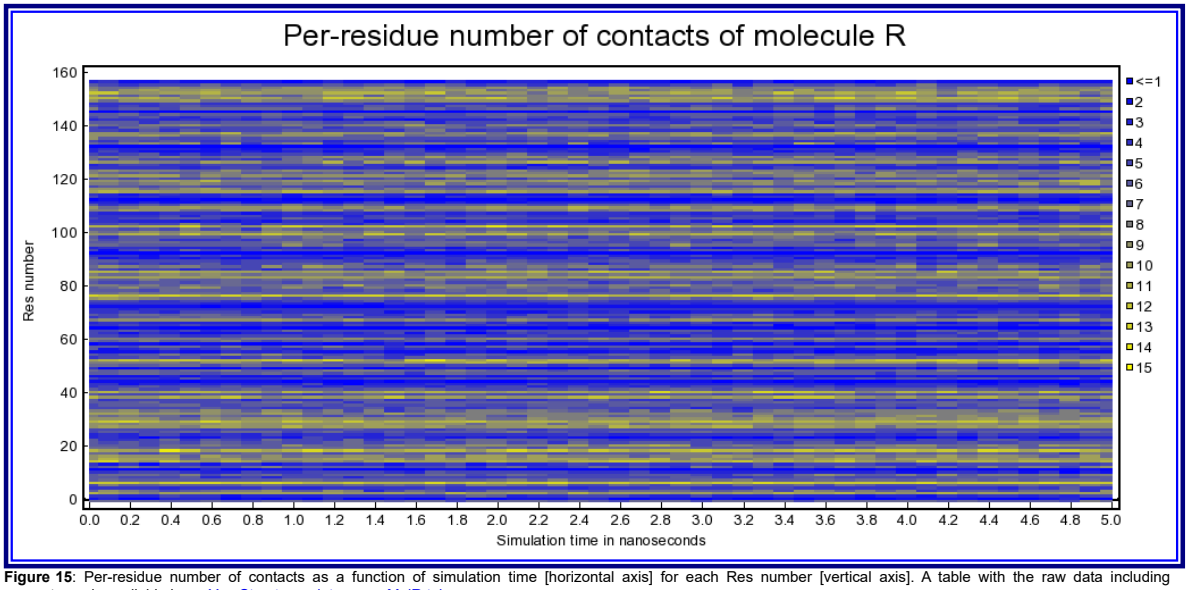

Figure 15: Per-residue number of contacts as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_conMolR.tab

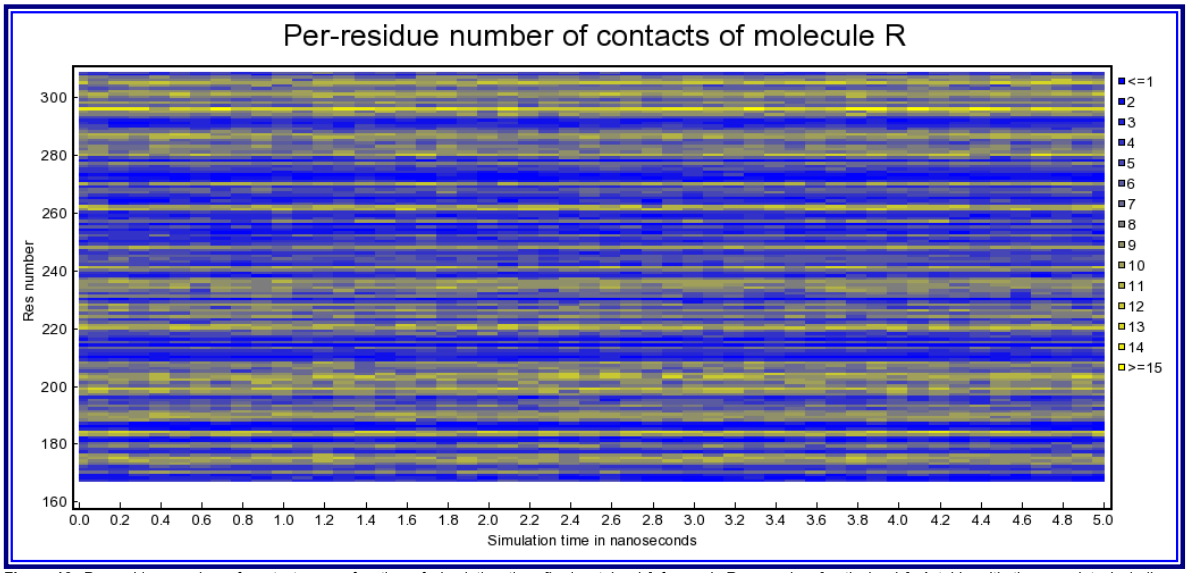

Figure 16: Per-residue number of contacts as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_conMolR.tab

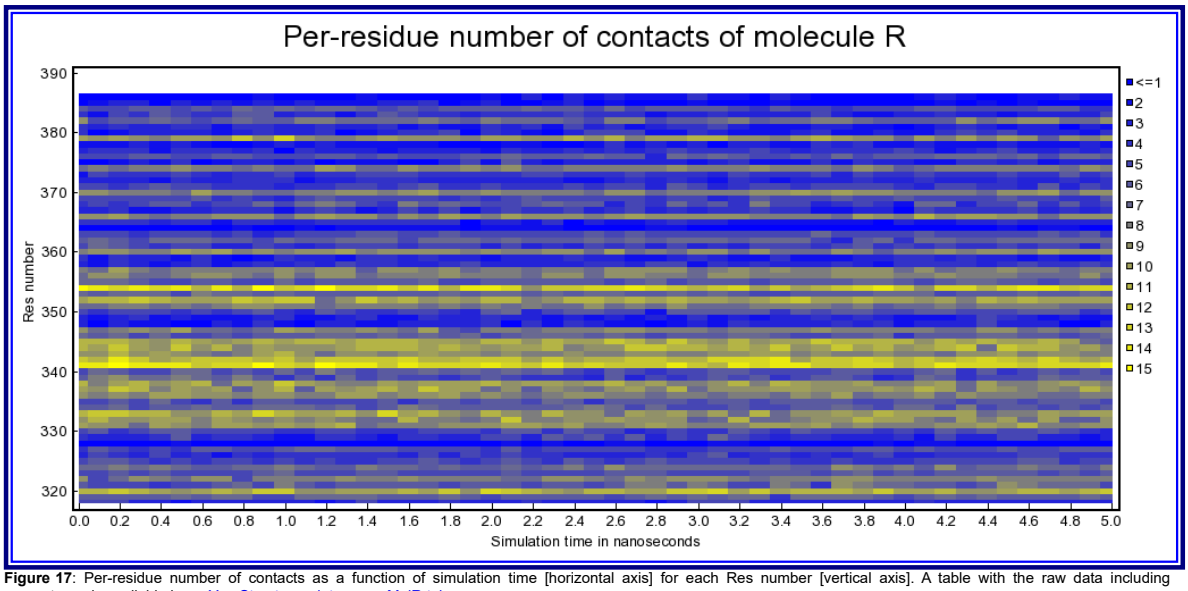

Figure 17: Per-residue number of contacts as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_conMolR.tab

2.12. Per-residue ligand interactions of the receptor

The following plots show the types of interactions by receptor residues with the ligand as a function of simulation time. There is one plot for each protein or nucleic acid molecule. Three types of interactions are shown: Hydrogen bonds [red] ,hydrophobic [green] and ionic interactions [blue]. Also mixtures of these three colors can show up if a certain residue is involved in more than one type of interaction with the ligand [see plot legend].

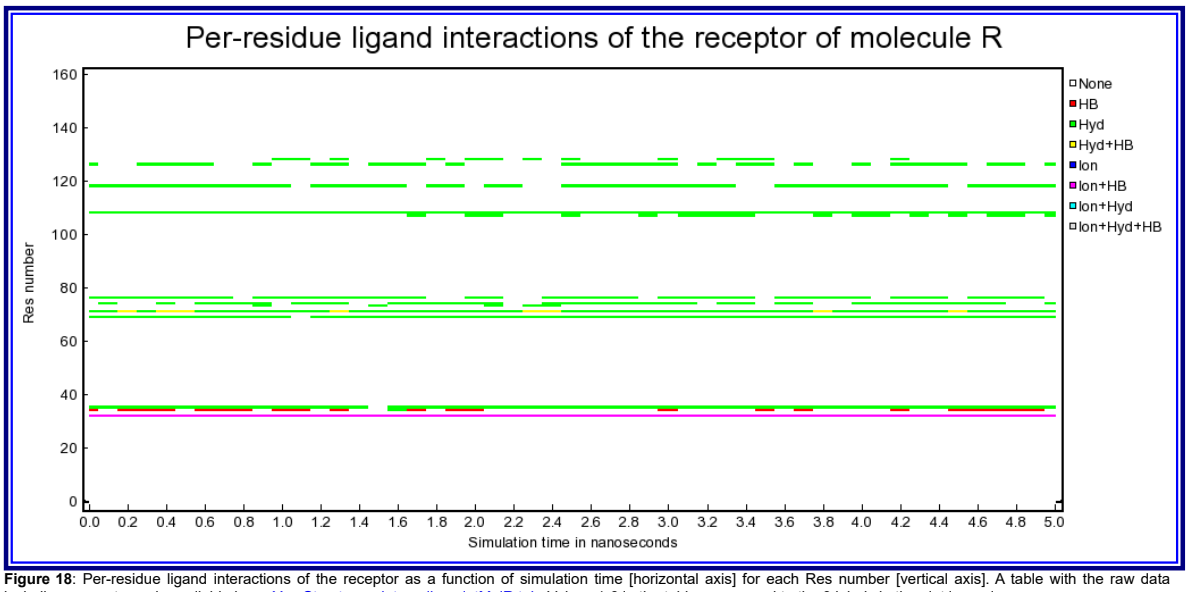

Figure 18: Per-residue ligand interactions of the receptor as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_ligresintMolR.tab. Values 1-8 in the table correspond to the 8 labels in the plot legend.

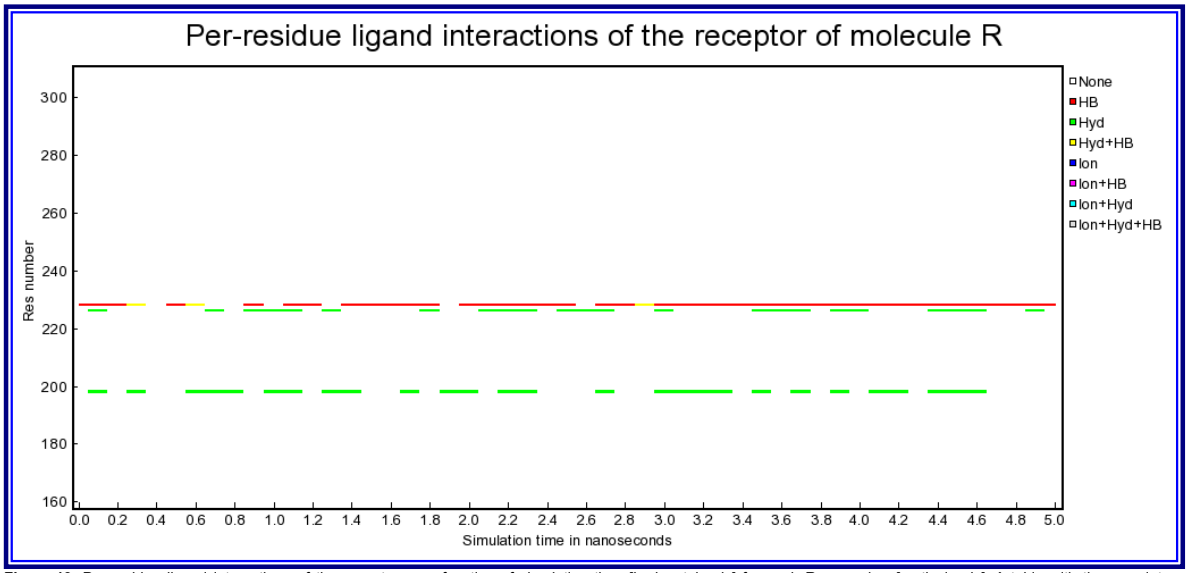

Figure 19: Per-residue ligand interactions of the receptor as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_ligresintMolR.tab. Values 1-8 in the table correspond to the 8 labels in the plot legend.

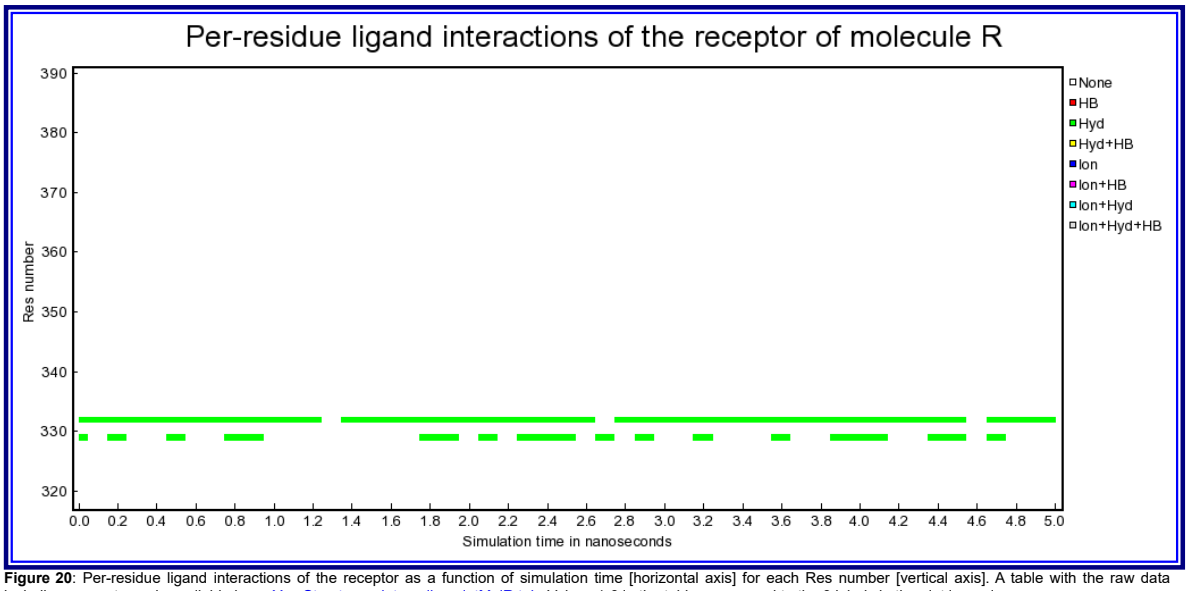

Figure 20: Per-residue ligand interactions of the receptor as a function of simulation time [horizontal axis] for each Res number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotres_ligresintMolR.tab. Values 1-8 in the table correspond to the 8 labels in the plot legend.

2.13. Per-atom receptor interactions of the ligand

The following plots show the types of interactions by atoms of the ligand with the receptor as a function of simulation time. There is one plot for each ligand molecule. Three types of interactions are shown: Hydrogen bonds [red] hydrophobic [green] and ionic interactions [blue]. Also mixtures of these three colors can show up if a certain ligand atom is involved in more than one type of interaction [see plot legend]. Atom numbers in the plots match the atom numbers shown in the screenshot of the ligand in the figure below.

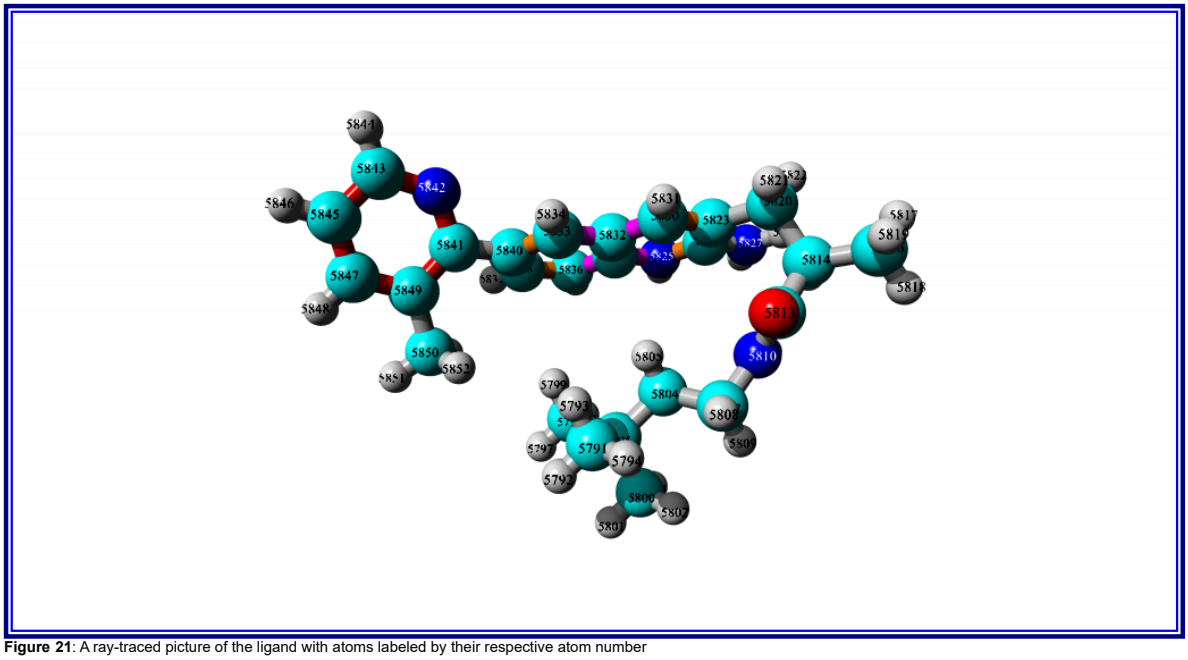

Figure 21: A ray-traced picture of the ligand with atoms labeled by their respective atom number

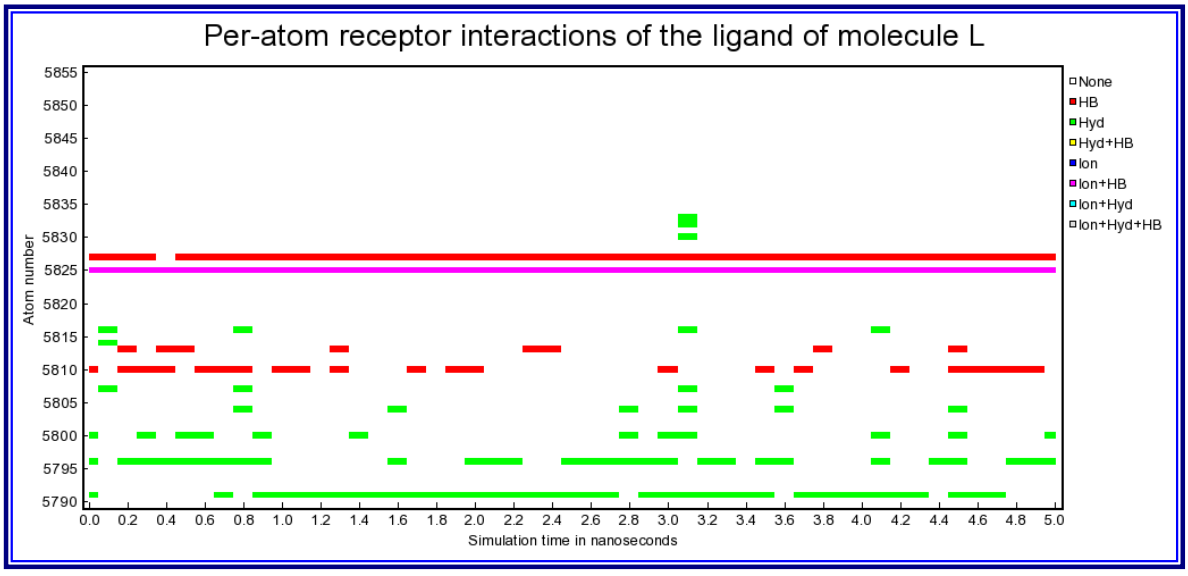

Figure 22: Per-atom receptor interactions of the ligand as a function of simulation time [horizontal axis] for each Atom number [vertical axis]. A table with the raw data including percentages is available here: YourStructure_plotatom_ligatomintMolL.tab. Values 1-8 in the table correspond to the 8 labels in the plot legend.

2.14. Interaction table

The following table lists all interactions [Hydrophobic, PiPi, CationPi, Ionic] with their average strength and all hydrogen bonds with their average energy between the ligand and the receptor. The occupancy is calculated by counting the occurrences and dividing them by the total number of analyzed snapshots. The interactions with the highest occupancy are sorted to the top of the table.

Interaction	Atom1	Atom2	Average	Occupancy
Hydrophobic	C15.h404.L	CB.S35.R	0.81	1.00
Hydrophobic	C5.h404.L	CE2.Y71.R	0.82	1.00
Hydrophobic	C6.h404.L	CG.Y71.R	0.70	1.00
Hydrophobic	C6.h404.L	CD1.Y71.R	0.75	1.00
Hydrophobic	C6.h404.L	CD2.Y71.R	0.70	1.00
Hydrophobic	C6.h404.L	CE1.Y71.R	0.85	1.00
Hydrophobic	C9.h404.L	CD1.I118.R	0.83	1.00
Hydrophobic	C7.h404.L	CG.Y71.R	0.89	1.00
Hydrophobic	C7.h404.L	CD2.Y71.R	0.86	1.00
Hydrophobic	C24.h404.L	CD2.Y71.R	0.74	1.00
Hydrophobic	C23.h404.L	CB.F108.R	0.68	1.00
Hydrophobic	C22.h404.L	CB.Y71.R	0.80	1.00
Hydrophobic	C22.h404.L	CB.F108.R	0.68	1.00
Hydrophobic	C22.h404.L	CG.F108.R	0.88	1.00
Hydrophobic	C21.h404.L	CD1.F108.R	0.76	1.00
Hydrophobic	C25.h404.L	CG1.V69.R	0.64	1.00
PiPi	C6.h404.L	CD1.Y71.R	0.93	1.00
PiPi	C6.h404.L	CE1.Y71.R	0.92	1.00
PiPi	C22.h404.L	CD1.F108.R	0.98	1.00
Ionic	N1.h404.L	CG.D32.R	0.94	1.00

Table 3: Interaction table. The last two columns show the average value and the ratio of occurrence in the analyzed snapshots. The table is sorted by decreasing occurrence ratio, with the most frequently occurring entries on top. Note: At most 20 table rows are shown. Change the **tabrowsmax** variable in the macro to adjust the number of shown table rows. The full table can be found in YourStructure_analysis_Pair.tab.

3. Analyses outside the simulation cell

The following section presents data gathered outside the simulation cell, where each object has its own local coordinate system and no periodic boundaries are present. Calculations that involve the interaction between objects [common surface areas, contacts between objects..] must be placed in the previous section.

3.1. Radius of gyration of the solute

After determining the center of mass of the solute, the radius of gyration is calculated and plotted according to this formula:

$$Radius_{gyr,Mass} = \sqrt{\frac{\sum_{i=1}^{N} Mass_i(\vec{R}_i - \vec{C})^2}{\sum_{i=1}^{N} Mass_i}}$$

In this formula, **C** is the center of mass, and **Ri** is the position of atom **i** of **N**.

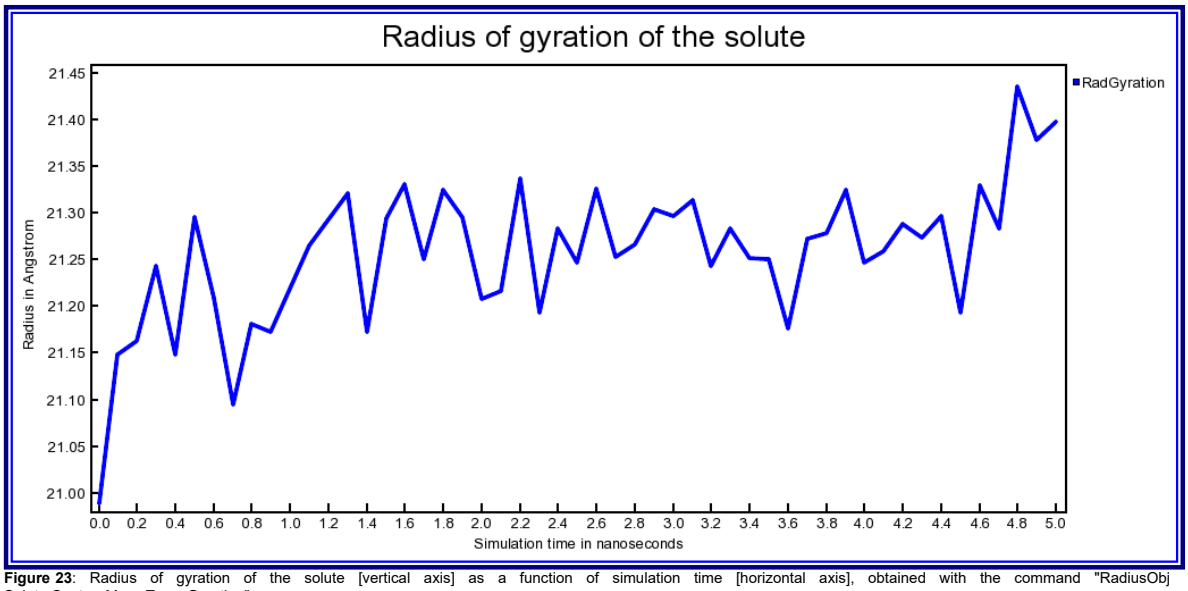

Figure 23: Radius of gyration of the solute [vertical axis] as a function of simulation time [horizontal axis], obtained with the command "RadiusObj Solute,Center=Mass,Type=Gyration".

4. Analyses performed with respect to the starting structure

Analyses performed with respect to the starting structure are shown in this section. These are also done outside the simulation cell, where each object has its own local coordinate systems and no periodic boundaries are present. To choose another reference snapshot than 0, edit the **refsnapshot** variable at the beginning of this macro.

4.1. Solute RMSD from the reference structure

The plot shows Calpha [RMSDCa], backbone [RMSDBb] and all-heavy atom [RMSDAll] RMSDs calculated according to this formula, where **Ri** is the vector linking the positions of atom **i** [of **N** atoms] in the reference snapshot and the current snapshot after optimal superposition:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} R_i * R_i}{n}}$$

The selection for the Calpha RMSD calculation is **CA Protein or C1* NucAcid and Obj Solute**, this matched 370 atoms. The Calpha selection thus includes the main backbone carbon C1* of nucleic acids, so the plot also shows a Calpha RMSD if you simulate just nucleic acids. In simulations of protein-DNA complexes, the Calpha RMSD therefore considers the DNA too. To change the Calpha selection, edit the **casel** variable at the beginning of this macro.

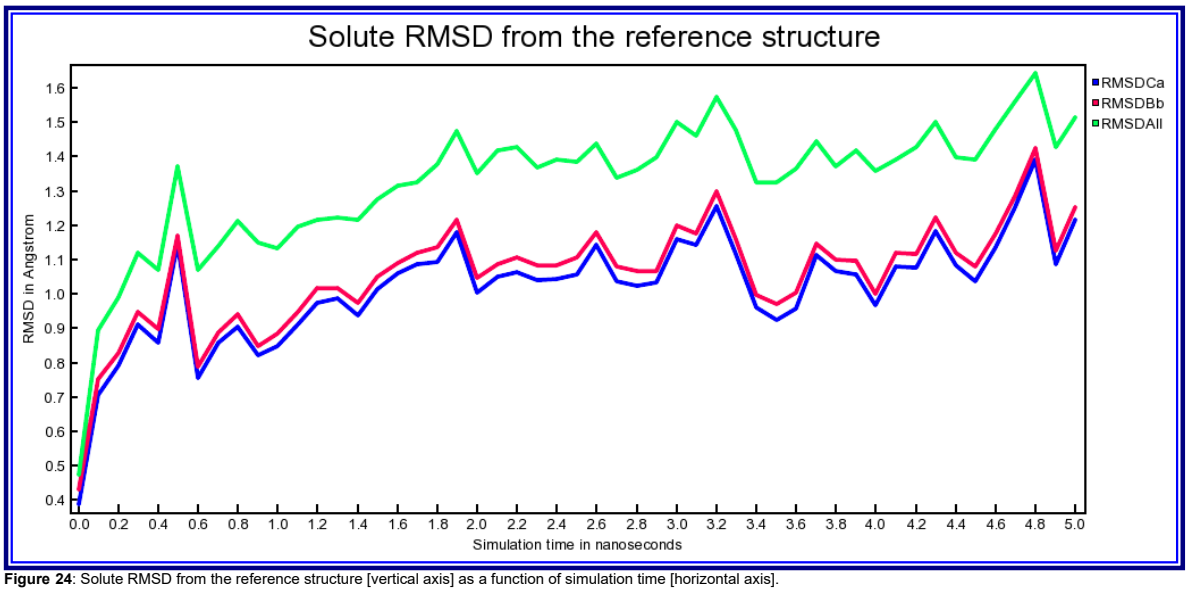

Figure 24: Solute RMSD from the reference structure [vertical axis] as a function of simulation time [horizontal axis].

4.2. Ligand movement RMSD after superposing on the receptor

The following plot shows the RMSD of the ligand heavy atoms over time, measured after superposing the receptor on its reference structure. This procedure delivers information about the movement of the ligand in its binding pocket.

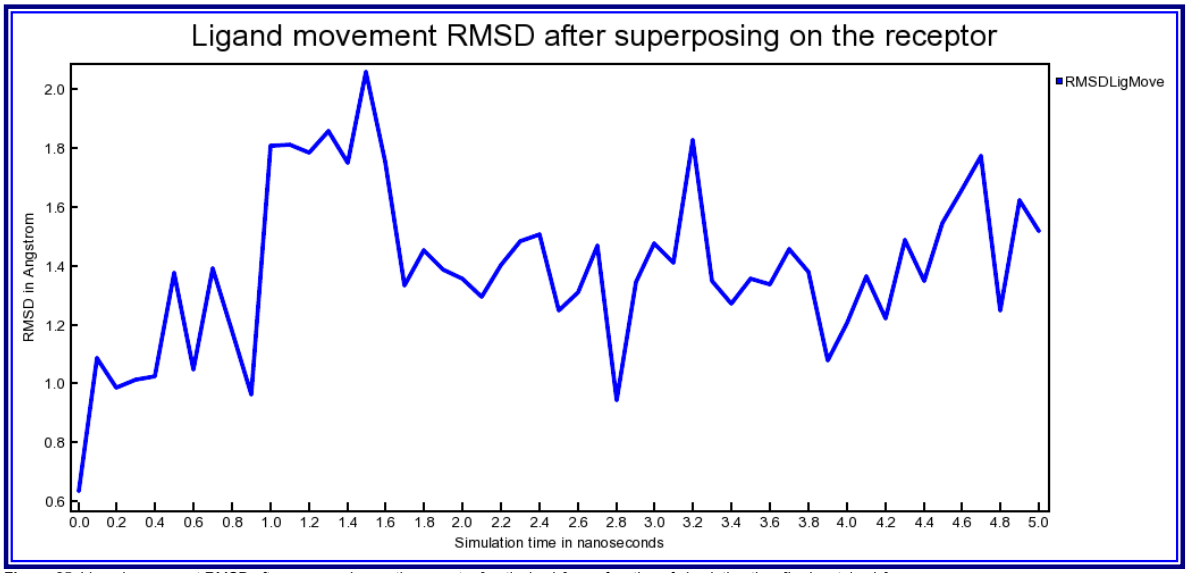

Figure 25: Ligand movement RMSD after superposing on the receptor [vertical axis] as a function of simulation time [horizontal axis].

4.3. Ligand conformation RMSD after superposing on the ligand

This plot displays the RMSD of the ligand atoms over time, measured after superposing on the reference structure of the ligand. The gained data summarize the conformational changes of the ligand.

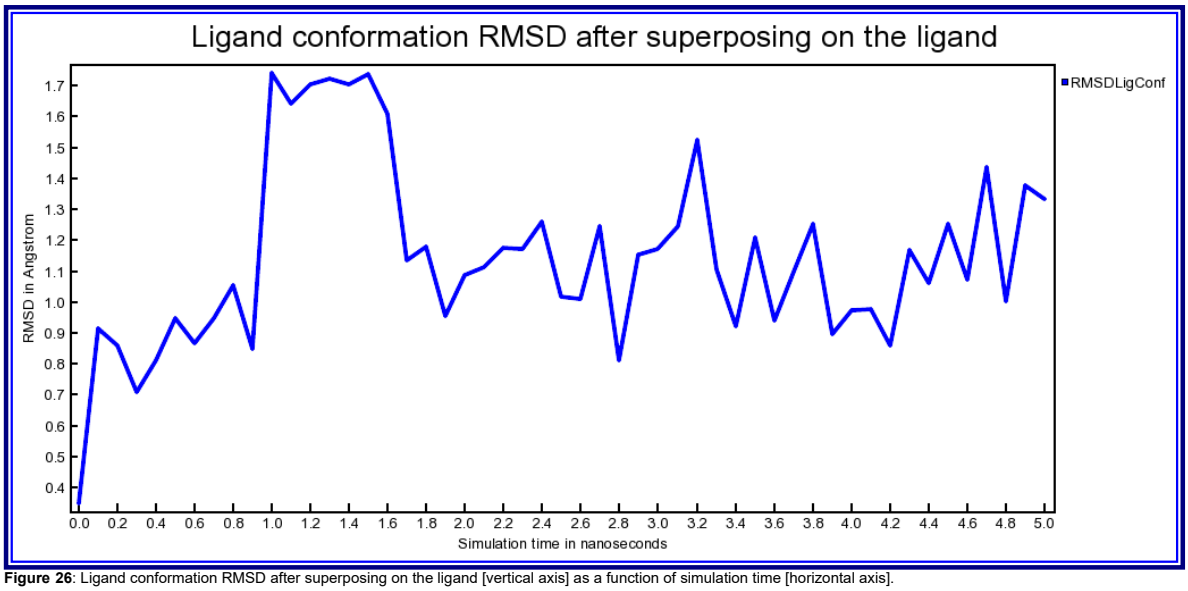

Figure 26: Ligand conformation RMSD after superposing on the ligand [vertical axis] as a function of simulation time [horizontal axis].

5. Solute residue RMSF

The Root Mean Square Fluctuation [RMSF] per solute residue is calculated from the average RMSF of its constituting atoms. The RMSF of atom i with j runing from 1 to 3 for the x, y, and z coordinate of the atom position vector P and k runing over the set of N evaluated snapshots is given by following formula:

$$RMSF_i = \sqrt{\sum_{j=1}^{3}(\frac{1}{N}\sum_{k=1}^{N}P_{ikj}^2 - \overline{P_{ij}}^2)}$$

Each graph in the following plot represents one molecule, so that you can easily see differences between molecules. Note: Residue numbers are not unique, so graphs can overlap.

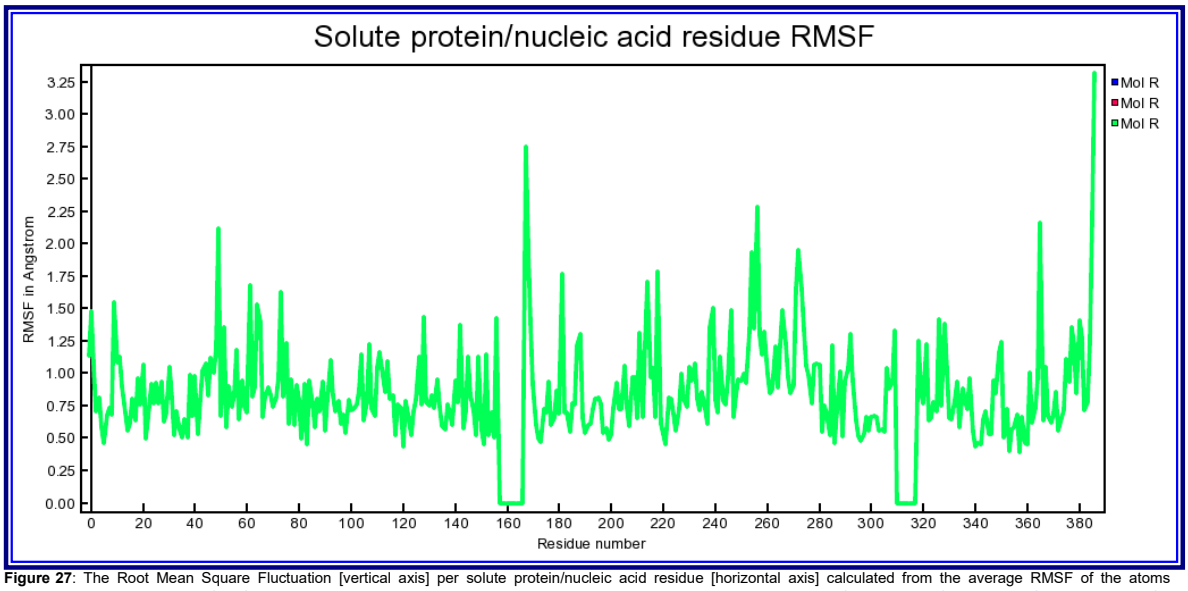

Figure 27: The Root Mean Square Fluctuation [vertical axis] per solute protein/nucleic acid residue [horizontal axis] calculated from the average RMSF of the atoms constituting the residue. A RMSF of exactly zero means that that residue number is not present in the molecule. Atom RMSF table: YourStructure_rmsf.tab, residue RMSF table: YourStructure_rmsfres.tab

In case the plot above is too crowded, the per-residue RMSF values are shown separately for all 3 molecules in the following plots:

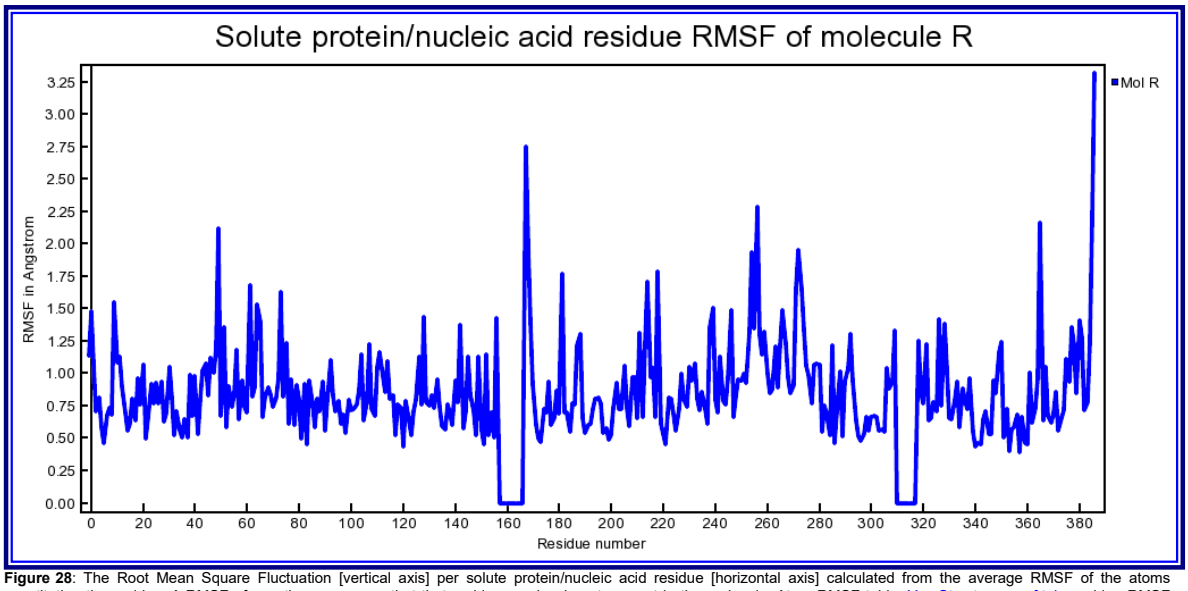

Figure 28: The Root Mean Square Fluctuation [vertical axis] per solute protein/nucleic acid residue [horizontal axis] calculated from the average RMSF of the atoms constituting the residue. A RMSF of exactly zero means that that residue number is not present in the molecule. Atom RMSF table: YourStructure_rmsf.tab, residue RMSF table: YourStructure_rmsfres.tab

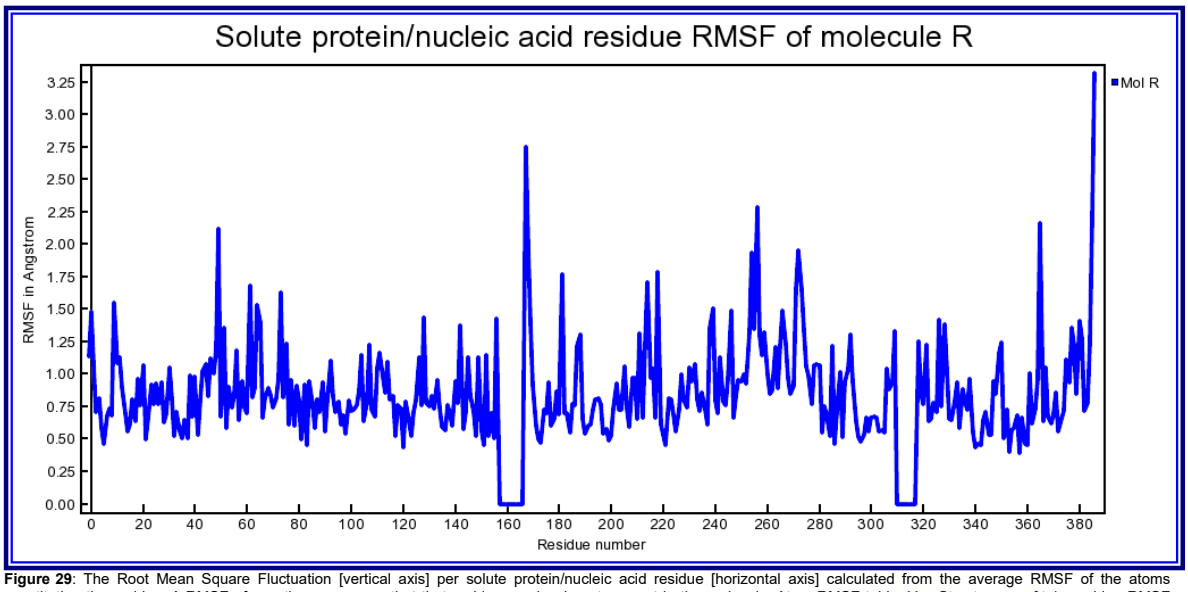

Figure 29: The Root Mean Square Fluctuation [vertical axis] per solute protein/nucleic acid residue [horizontal axis] calculated from the average RMSF of the atoms constituting the residue. A RMSF of exactly zero means that that residue number is not present in the molecule. Atom RMSF table: YourStructure_rmsf.tab, residue RMSF table: YourStructure_rmsfres.tab

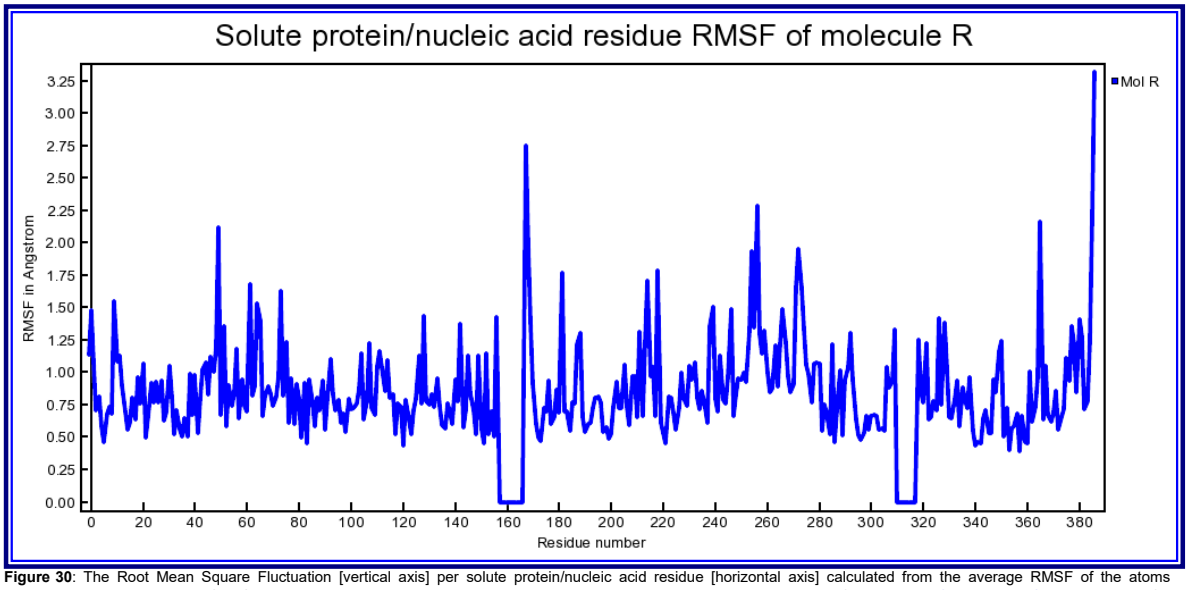

Figure 30: The Root Mean Square Fluctuation [vertical axis] per solute protein/nucleic acid residue [horizontal axis] calculated from the average RMSF of the atoms constituting the residue. A RMSF of exactly zero means that that residue number is not present in the molecule. Atom RMSF table: YourStructure_rmsf.tab, residue RMSF table: YourStructure_rmsfres.tab

Mol	Residue	First atom	RMSF[A]
R	Nme 157	2435	2.76
R	Ace 166	2441	3.17
R	Nme 310	4740	1.85
R	Ace 317	4746	1.77

L	68m 404	5791	0.96

Table 4: RMSF in Angstrom for non-protein/nucleic acid residues in the solute.

6. Principal Component Analysis

Principal component analysis [PCA] has been performed on the atoms selected with **N CA C Protein Backbone or C1* C3* C5* NucAcid** and data colletced for the first **3** of 3330 principal components. Edit the **pcasel** or **pcacomps** variable at the beginning of this macro to change the atom selection or the number of analyzed components respectively.

Each principal component corresponds to a pair of eigenvalues and eigenvectors of the covariance matrix:

$$Cov_{ij} = \frac{1}{N}\sum_{k=1}^{N}(X_{i,k} - \overline{X}_i)(X_{j,k} - \overline{X}_j)$$

with **N** denoting the number of analyzed snapshots, **X1** to **X3** the Cartesian coordinates of the first atom, **X4** to **X6** the Cartesian coordinates of the second atom and so forth. An eigenvector of the covariance matrix represents a linear combination of highly correlated atom coordinates and its corresponding eigenvalue gives a measure of how much variance in the position data is explanied by this given component. The collective motion of analyzed atoms is hence well approximated by the first few components with the largest eigenvalues, drastically reducing the high dimensionality of the position data to a handful of PCA components.

The movement of the solute along each analyzed PCA component is reconstructed by loading the average strucure and then varying each component value from its minimum to its maximum while holding all other component values constant.

Figure 31: Visualization of the movement along PCA component 1. The minimum and maximum value for each analyzed component can be found in the last rows of the PCA components data table: YourStructure_pca.tab

Figure 32: Visualization of the movement along PCA component 2. The minimum and maximum value for each analyzed component can be found in the last rows of the PCA components data table: YourStructure_pca.tab

Figure 33: Visualization of the movement along PCA component 3. The minimum and maximum value for each analyzed component can be found in the last rows of the PCA components data table: YourStructure_pca.tab

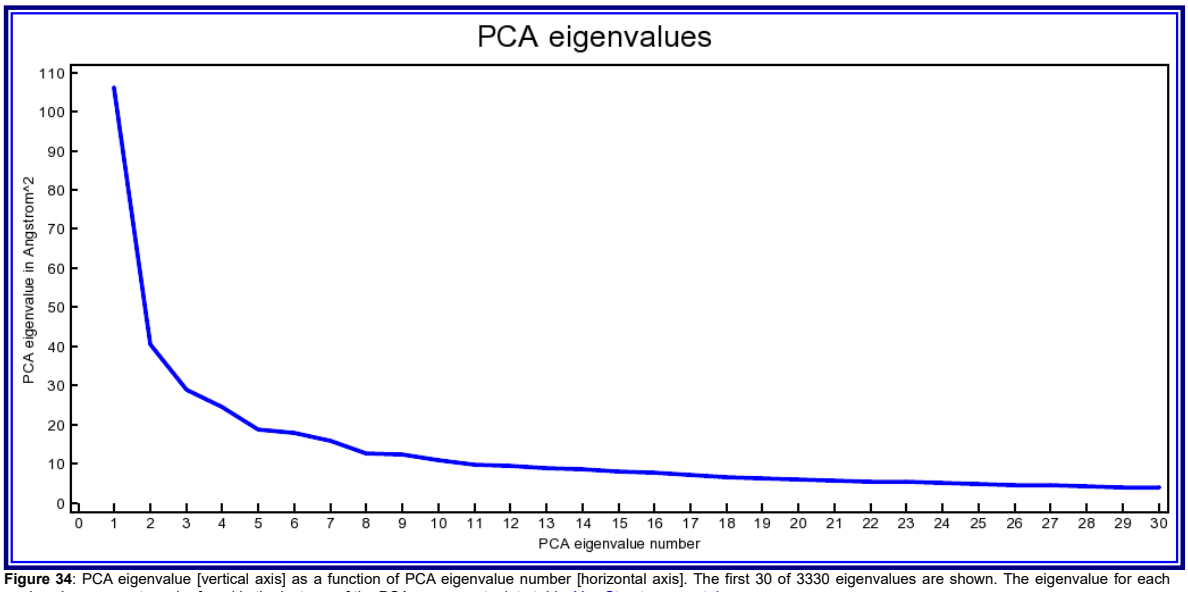

Figure 34: PCA eigenvalue [vertical axis] as a function of PCA eigenvalue number [horizontal axis]. The first 30 of 3330 eigenvalues are shown. The eigenvalue for each analyzed component can be found in the last row of the PCA components data table: YourStructure_pca.tab

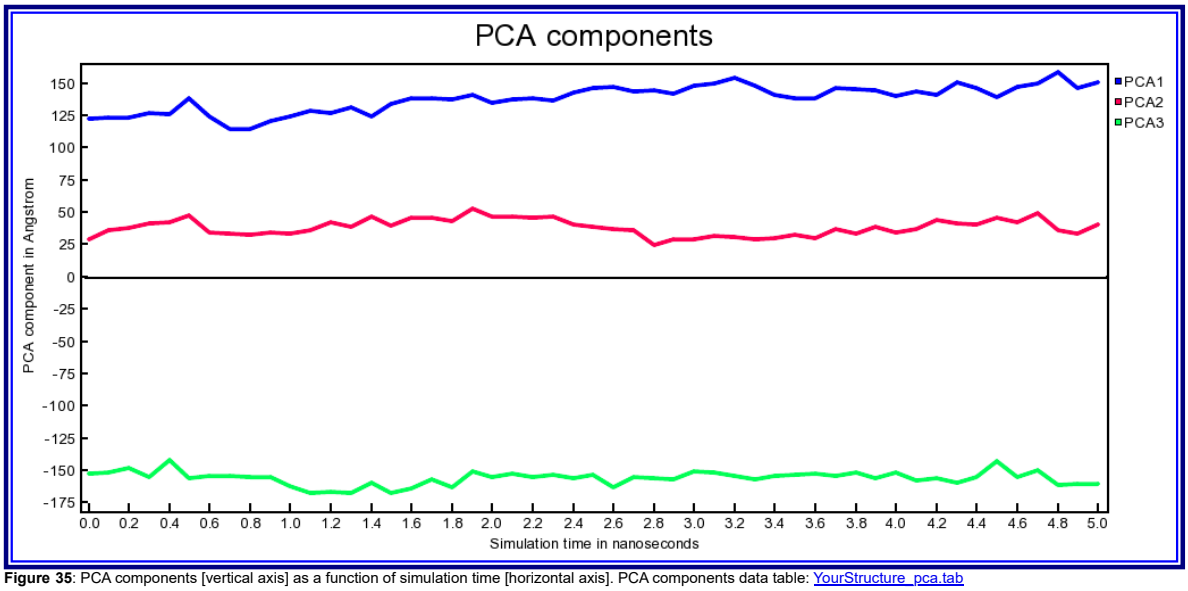

Figure 35: PCA components [vertical axis] as a function of simulation time [horizontal axis]. PCA components data table: YourStructure_pca.tab

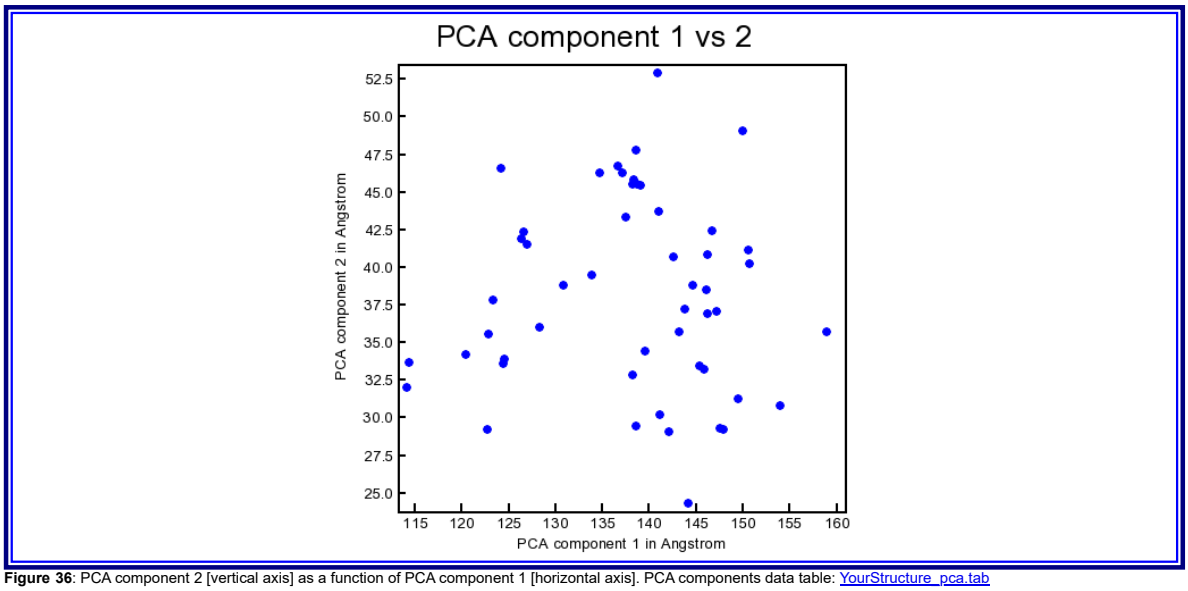

Figure 36: PCA component 2 [vertical axis] as a function of PCA component 1 [horizontal axis]. PCA components data table: YourStructure_pca.tab

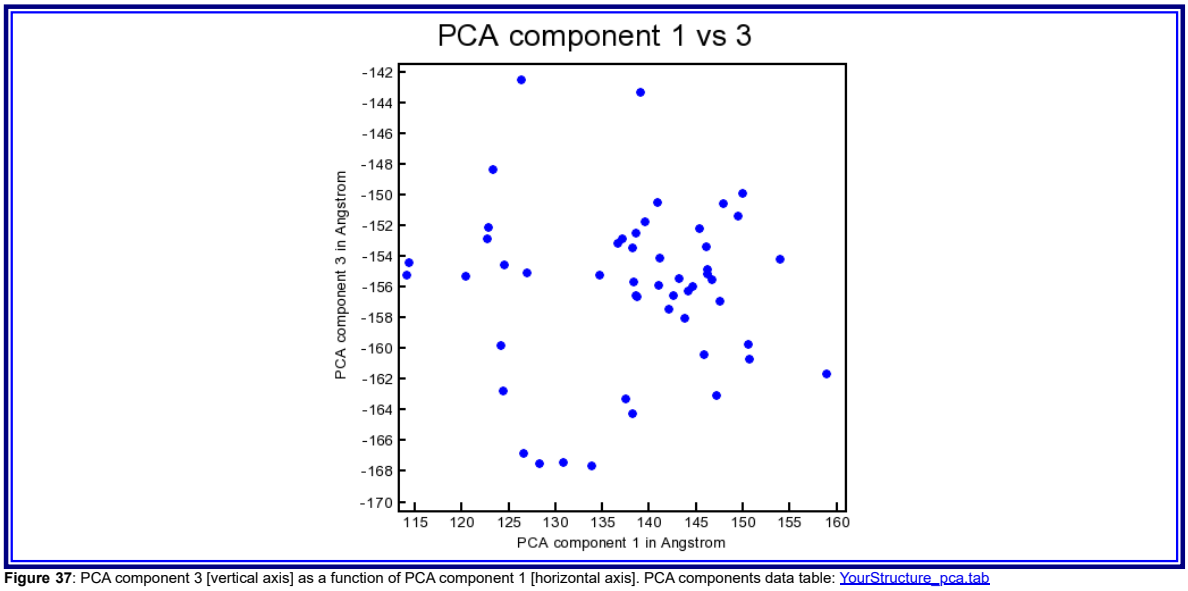

Figure 37: PCA component 3 [vertical axis] as a function of PCA component 1 [horizontal axis]. PCA components data table: YourStructure_pca.tab

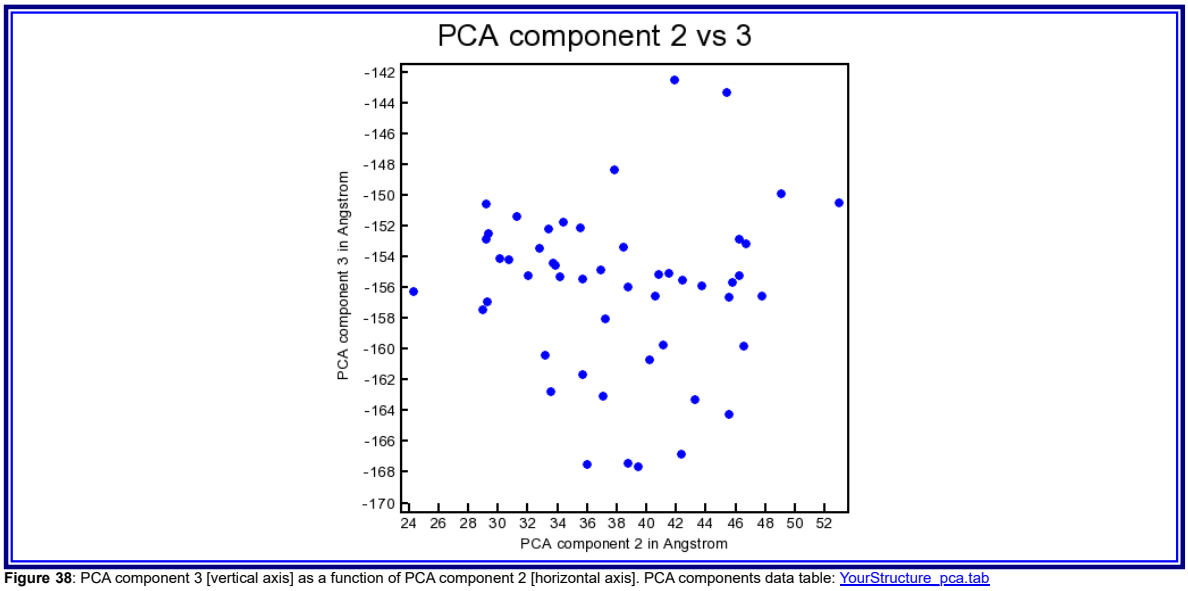

Figure 38: PCA component 3 [vertical axis] as a function of PCA component 2 [horizontal axis]. PCA components data table: YourStructure_pca.tab

7. Dynamic Cross-Correlation Matrix

The dynamic cross-correlation matrix [DCCM] is a square matrix, whose rows and columns match the selected units **Atom CA Protein or C1* NucAcid**. To change this selection, edit the **dccmsel** variable at the beginning of this macro. The DCCM shows how the movements of all selected pairs correlate. The values in the DCCM range from -1 [perfectly anti-correlated] to +1 [perfectly correlated]. The values along the diagonal are always +1 [because the motion of an atom is perfectly correlated to itself]. The DCCM element for units i and j is obtained with the following formula:

$$DCCM_{i,j} = \frac{\langle \vec{d}_i \cdot \vec{d}_j \rangle}{\sqrt{\langle d_i^2 \rangle \langle d_j^2 \rangle}}$$

Here **d** is the displacement between the current position and the average position of the selected unit, and the angle brackets indicate the average over all samples. The highest correlations off the diagonal can often be found for bridged cysteines.

The image below shows the correlation directly in the solute object:

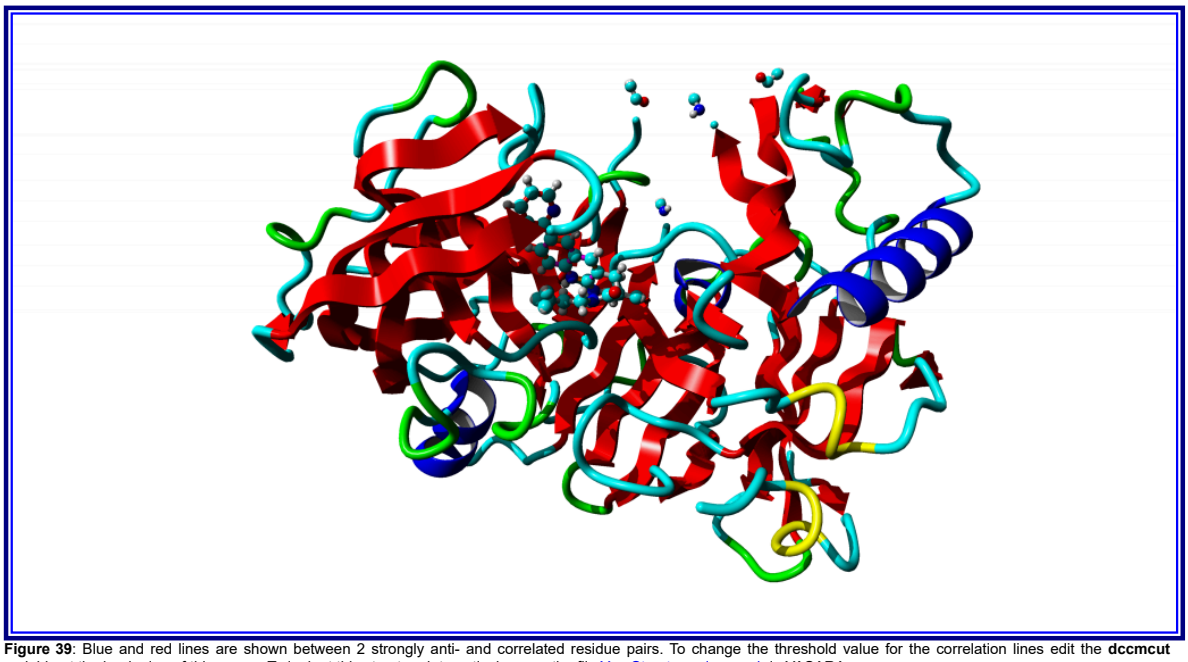

Figure 39: Blue and red lines are shown between 2 strongly anti- and correlated residue pairs. To change the threshold value for the correlation lines edit the **dccmcut** variable at the beginning of this macro. To look at this structure interactively, open the file YourStructure_dccm.yob in YASARA.

In the image below, the DCCM is visualized with colors ranging from blue [-1, fully anti-correlated] to yellow [+1, fully correlated].

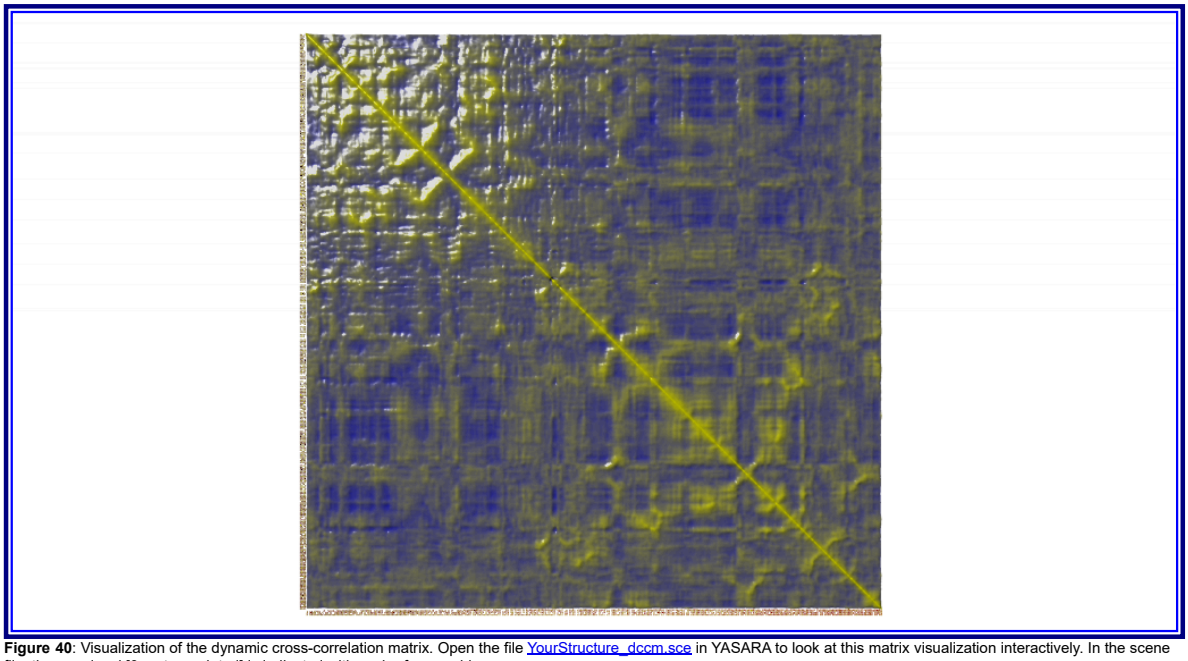

Figure 40: Visualization of the dynamic cross-correlation matrix. Open the file YourStructure_dccm.sce in YASARA to look at this matrix visualization interactively. In the scene file, the zero level [0, not correlated] is indicated with a wire-frame grid.

DCCM	R Phe -1	R Val 0	R Glu 1	R Met 2	R Val 3	R Asp 4	R Asn 5	R Leu 6	R Arg 7	R Gly 8	R Lys 9	R Ser 10	R Gly 11	R Gln 12	R Gly 13	R Tyr 14	R Tyr 15	R Val 16	R Glu 17	R Met 18	R Thr 19	R Val 20	R Gly 21	R Ser 22	R Pro 23	R Pro 24	R Gln 25
R Phe -1	1.00	0.77	0.45	0.39	0.26	0.19	0.21	0.14	0.14	0.08	-0.01	-0.06	-0.03	-0.02	-0.03	0.10	0.15	0.27	0.18	0.03	-0.00	-0.07	-0.11	-0.26	-0.21	-0.21	-0.0
R Val 0	0.77	1.00	0.60	0.41	0.32	0.21	0.23	0.12	0.10	0.05	-0.04	-0.07	-0.06	-0.02	-0.01	0.08	0.10	0.16	0.12	-0.01	-0.04	-0.07	-0.10	-0.21	-0.17	-0.19	-0.0
R Glu 1	0.45	0.60	1.00	0.71	0.42	0.32	0.35	0.24	0.16	0.06	-0.06	-0.06	0.01	-0.05	0.05	0.12	0.14	0.15	0.09	0.01	0.06	-0.03	-0.09	-0.07	-0.12	-0.16	-0.0
R Met 2	0.39	0.41	0.71	1.00	0.75	0.59	0.54	0.34	0.16	-0.02	-0.13	-0.16	-0.13	-0.13	-0.01	0.10	0.18	0.20	0.14	0.01	0.05	-0.01	-0.15	-0.22	-0.23	-0.23	-0.0
R Val 3	0.26	0.32	0.42	0.75	1.00	0.74	0.54	0.35	0.13	-0.12	-0.20	-0.20	-0.15	-0.21	-0.14	0.09	0.19	0.21	0.13	0.02	0.01	0.04	-0.18	-0.29	-0.22	-0.19	-0.0
R Asp 4	0.19	0.21	0.32	0.59	0.74	1.00	0.74	0.54	0.31	-0.01	-0.06	-0.02	0.04	-0.03	0.02	0.30	0.35	0.33	0.24	0.16	0.07	0.07	-0.08	-0.22	-0.16	-0.02	0.0
R Asn 5	0.21	0.23	0.35	0.54	0.54	0.74	1.00	0.69	0.48	0.23	0.16	0.12	0.14	0.20	0.28	0.45	0.53	0.51	0.43	0.34	0.18	0.01	-0.12	-0.15	-0.05	0.01	-0.0
R Leu 6	0.14	0.12	0.24	0.34	0.35	0.54	0.69	1.00	0.72	0.27	0.18	0.19	0.23	0.29	0.32	0.61	0.69	0.59	0.49	0.39	0.24	-0.05	-0.13	-0.16	-0.03	0.14	0.0

R Arg 7	0.14	0.10	0.16	0.16	0.13	0.31	0.48	0.72	1.00	0.67	0.53	0.43	0.31	0.37	0.35	0.52	0.59	0.46	0.31	0.19	0.06	-0.16	-0.19	-0.22	-0.16	-0.00	-0.1
R Gly 8	0.08	0.05	0.06	-0.02	-0.12	-0.01	0.23	0.27	0.67	1.00	0.77	0.60	0.39	0.51	0.46	0.29	0.28	0.19	0.11	0.02	-0.06	-0.19	-0.07	0.00	-0.02	0.01	-0.1
R Lys 9	-0.01	-0.04	-0.06	-0.13	-0.20	-0.06	0.16	0.18	0.53	0.77	1.00	0.85	0.58	0.66	0.48	0.29	0.27	0.16	0.04	-0.08	-0.12	-0.19	-0.07	0.07	0.08	0.06	-0.1
R Ser 10	-0.06	-0.07	-0.06	-0.16	-0.20	-0.02	0.12	0.19	0.43	0.60	0.85	1.00	0.80	0.72	0.51	0.38	0.24	0.10	-0.01	-0.13	-0.15	-0.17	0.02	0.12	0.14	0.16	-0.1
R Gly 11	-0.03	-0.06	0.01	-0.13	-0.15	0.04	0.14	0.23	0.31	0.39	0.58	0.80	1.00	0.76	0.50	0.50	0.32	0.20	0.12	0.02	-0.03	-0.01	0.13	0.21	0.23	0.27	0.0
R Gln 12	-0.02	-0.02	-0.05	-0.13	-0.21	-0.03	0.20	0.29	0.37	0.51	0.66	0.72	0.76	1.00	0.74	0.63	0.44	0.23	0.22	0.13	0.10	-0.08	0.00	0.18	0.27	0.24	0.0
R Gly 13	-0.03	-0.01	0.05	-0.01	-0.14	0.02	0.28	0.32	0.35	0.46	0.48	0.51	0.50	0.74	1.00	0.63	0.42	0.21	0.21	0.14	0.12	-0.05	-0.05	0.16	0.17	0.15	-0.0
R Tyr 14	0.10	0.08	0.12	0.10	0.09	0.30	0.45	0.61	0.52	0.29	0.29	0.38	0.50	0.63	0.63	1.00	0.80	0.55	0.52	0.40	0.30	0.03	-0.08	-0.04	0.11	0.20	0.1
R Tyr 15	0.15	0.10	0.14	0.18	0.19	0.35	0.53	0.69	0.59	0.28	0.27	0.24	0.32	0.44	0.42	0.80	1.00	0.80	0.66	0.51	0.36	0.05	-0.10	-0.05	0.11	0.22	0.1
R Val 16	0.27	0.16	0.15	0.20	0.21	0.33	0.51	0.59	0.46	0.19	0.16	0.10	0.20	0.23	0.21	0.55	0.80	1.00	0.81	0.60	0.37	0.13	-0.08	-0.06	0.07	0.17	0.1
R Glu 17	0.18	0.12	0.09	0.14	0.13	0.24	0.43	0.49	0.31	0.11	0.04	-0.01	0.12	0.22	0.21	0.52	0.66	0.81	1.00	0.81	0.55	0.21	-0.03	0.02	0.14	0.29	0.3

Table 5: Dynamic cross-correlation matrix. The full table is also available in text format, you need a proper text editor without line wrapping to look at this file: YourStructure_dccm.tab. Note: At m

8. Additional files

The following additional files have been created:

8.1. The main data table

The main table contains all collected data in a single file. The column names match the names used above for graphs in plots and columns in tables. You can find a more detailed explanation of this table in the user manual at Recipes > Run a molecular dynamics simulation > Analyzing a trajectory. If you parse this file automatically, keep in mind that the number of columns can change any time, so you have to use the names in the first table row to find the columns of interest: YourStructure_analysis.tab

8.2. Extra data tables

User defined extra data tables:

YourStructure_analysis_Pair.tab

8.3. Per-atom and per-residue data tables

Data of the per-atom and per-residue plots:

YourStructure_plotres_secstrMolR.tab

YourStructure_plotres_secstrMolR.tab

YourStructure_plotres_secstrMolR.tab

YourStructure_plotres_conMolR.tab

YourStructure_plotres_conMolR.tab

YourStructure_plotres_conMolR.tab

YourStructure_plotres_ligresintMolR.tab

YourStructure_plotres_ligresintMolR.tab

YourStructure_plotres_ligresintMolR.tab

YourStructure_plotatom_ligatomintMolL.tab

8.4. The structures

The **time averaged structure** in PDB format: YourStructure_average.pdb

The **snapshot with the minimum solute energy**. Either just the solute in PDB format YourStructure_energymin.pdb, or the complete system including solvent as a YASARA scene YourStructure_energymin.sce.

The **last snapshot** of the simulation. Either just the solute in PDB format YourStructure_last.pdb, or the complete system including solvent as a YASARA scene YourStructure_last.sce

8.5. The RMSF tables

A table that lists the Root Mean Square Fluctuations [RMSFs] of all atoms in [A] is available here: YourStructure_rmsf.tab. The RMSFs have also been converted to B-factors and stored in the B-factor field of the time-average structure above.

A table with average atom RMSFs per residue can be found here: YourStructure_rmsfres.tab.

8.6. High resolution plots

To facilitate publication, high resolution versions of the plots above have been created with a 4:3 aspect ratio suited for printing in a single column of a typical journal article. Just look at the figure number above to find the right file:

YourStructure_report_figure4_hires.png

YourStructure_report_figure5_hires.png

YourStructure_report_figure6_hires.png

YourStructure_report_figure7_hires.png

YourStructure_report_figure8_hires.png

YourStructure_report_figure9_hires.png

YourStructure_report_figure10_hires.png

YourStructure_report_figure11_hires.png

YourStructure_report_figure12_hires.png

YourStructure_report_figure13_hires.png

YourStructure_report_figure14_hires.png

YourStructure_report_figure15_hires.png

YourStructure_report_figure16_hires.png

YourStructure_report_figure17_hires.png

YourStructure_report_figure18_hires.png

YourStructure_report_figure19_hires.png

YourStructure_report_figure20_hires.png

YourStructure_report_figure22_hires.png

YourStructure_report_figure23_hires.png

YourStructure_report_figure24_hires.png

YourStructure_report_figure25_hires.png

YourStructure_report_figure26_hires.png

YourStructure_report_figure27_hires.png

YourStructure_report_figure28_hires.png

YourStructure_report_figure29_hires.png

YourStructure_report_figure30_hires.png

YourStructure_report_figure34_hires.png

YourStructure_report_figure35_hires.png

YourStructure_report_figure36_hires.png

YourStructure_report_figure37_hires.png

YourStructure_report_figure38_hires.png