

# **Molecular simulations of a helix-turn-helix motif for use as a biosensor**

Mart Ernits

Master's thesis

Nanobiotechnology

Spring 2017



**Department of  
Physics and Nanotechnology** Nanotechnology  
Skjernvej 4A  
9220 Aalborg Ø

Abstract

**Title:**

Molecular simulations of a helix-turn-helix motif for use as a biosensor

**Theme:**

Master's thesis

**Project period:**

Spring 2017

**Student:**

Mart Ernits

**Supervisors:**

Peter Fojan  
Eva Maria Petersen

**Number of copies: 3**

**Number of Pages: 77**

**Finished: 11 June 2017**

Biosensors are used in a variety of applications in science and medicine. Computational methods are used in this project to study the ion binding of P3W, an EF-hand like protein.

A homology model is created and analyzed using molecular dynamics. Simulations are done for  $\text{Eu}^{3+}$ ,  $\text{Fe}^{3+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  ions. A protocol termed MDAnalysis is used and improved during the project.

The N to C terminus distances of the resulting configurations are measured. FRET efficiency is calculated from these distances and ions are ranked based on their ability to quench FRET.

The homology model has a number of structural and quality problems and the MDAnalysis algorithm appears to suffer from an inadequate energy minimization technique.

Further research is necessary to validate the results and improve the quality of the predictions.

*The contents of this project are freely accessible but publication (with reference) may only happen upon agreement with the student.*

## **Preface**

This project is a master's thesis done during one semester in the Nanobiotechnology master's program at Aalborg University. The author of this thesis is Mart Ernits, the supervisors are Peter Fojan and Eva Maria Petersen.

In text, references consist of the name or names of the authors of the referenced work, followed by the year of publication in parentheses. References where the author's name and year are both surrounded by parentheses reference the entire preceding block of text if they come after a period, and just the sentence if they are placed before a period. If a figure was not created by the author, then the associated reference is placed in the caption of the figure. The bibliography section contains full information about the references used.

Mart Ernits

---

# Contents

Introduction	6
Biosensors	6
EF-Hand	6
P3W	7
FRET	8
Structure modeling	8
Knowledge-based potentials	9
QMEAN	10
Materials and methods	12
Homology modeling	12
Structural alignment	12
Simulations	12
MDAnalysis	13
Restraining potential	13
Distance measurements	14
FRET efficiency	14
Binding energy	14
Restraint energy	14
Results and discussion	15
Homology modeling	15
Simulations	21
Simulations to check the stability of the reached configurations	51
FRET efficiency calculations	67
General Discussion	71
Conclusion	73
Outlook	74
Bibliography	75

# Introduction

The aim of this project is to do molecular simulations to support investigations of a FRET-based biosensor that could be used to investigate conformational changes of P3W, which could then be used, for example, to create a contrast agent for the imaging of cancer cells. Another potential use for a biosensor like this could be to detect the presence of ions it binds to in solutions. In addition, learning about molecular dynamics and modeling can be named as a secondary objective of this project.

## Biosensors

Biosensors are a relatively widely used set of tools in research and medicine. They are molecules or sets of molecules acting together to produce a signal when they encounter a specific target molecule, or another type of trigger, such as change of temperature. The output signal can be some sort of a change in electrical conductivity, a light or color change, a change in temperature or even a mass change. (Bohunicky and Mousa 2010)

A biosensor can be based on a variety of molecule types, such as antibodies, enzymes, other kinds of proteins and nucleic acids. Antibodies are some of the most convenient types, because they are highly specific and do not require purification. (Bohunicky and Mousa 2010)

However, depending on the scenario, different approaches are sensible. For example aptamers, which are oligonucleotides that are discovered via screening and can be produced in vitro. Some aptamers are based on proteins as well. They have numerous advantages over antibodies, such as greater reusability and the fact that they can be used on a wider array of targets. (Kuchelmeister and Schmuck 2013)

DNA can be detected using a variety of methods, one of these is a helix-turn-helix (HTH) structure. This motif even has a level of sequence specificity, making it a possible base for a DNA detecting biosensor. (Brennan and Matthews 1989)

## EF-Hand

EF-hands are a relatively widespread and well-studied type of calcium binding proteins that were discovered in 1973. The majority of EF-hands are comprised of two or more HTH motifs, where the turn is responsible for the binding of  $\text{Ca}^{2+}$  and other similar ions. The motifs are usually comprised of about 30 amino acids while the central loop or turn is around 9 amino acid residues long. (Kretsinger, Uversky et al. 2013)

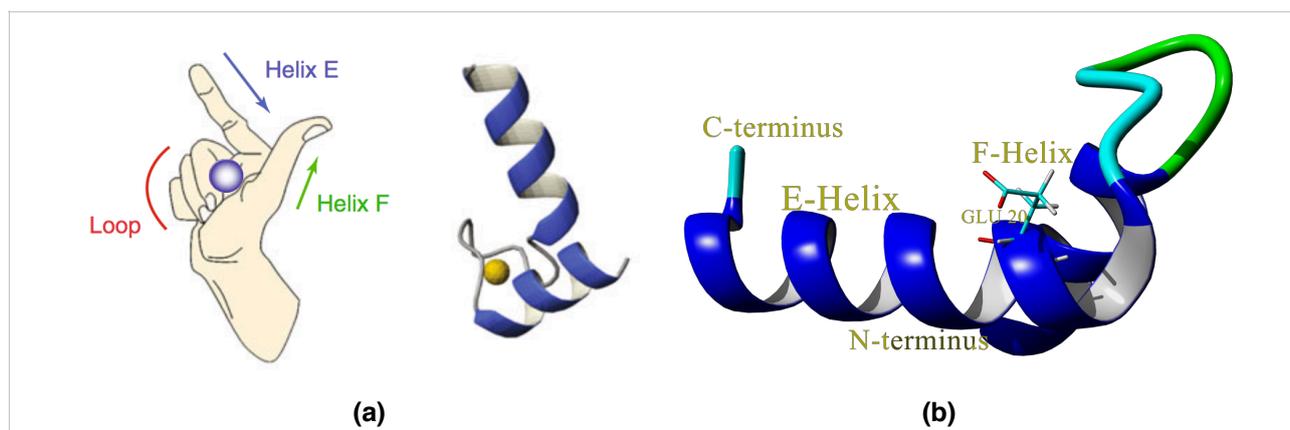
These motifs are usually found in pairs, but proteins containing four and even five instances of the HTH pattern have been described. In the latter case, the fifth motif is often found to form a pair with the lone HTH of another molecule, forming a dimer of 10 HTH motifs. (Kretsinger, Uversky et al. 2013)

## Introduction

EF-hands can be used as calcium sensors, to gauge the amount of calcium ions in a solution. They are naturally DNA-binding and have some sequence selectivity, therefore making them useable as DNA detectors and cleaving devices. (Brennan and Matthews 1989) (Kovacic, Welch et al. 2003)

Their ability to bind certain sequences of DNA could also be used to deliver payloads to some targets. In nature, they can be found in transcription factors. (Ikura, Osawa et al. 2002)

The structure of a single HTH motif of an EF-hand is described in figure 1. The longer helix is called the E-helix and the shorter one is called the F-helix. The figure also demonstrates the way most simulation and modeling results are presented in this report, the E-helix is pointing left and the residue Glu 20 is positioned above the helix axis.



**Figure 1:** EF hand helix-turn-helix motif. (a) Comparison of the canonical form with a right hand. Calcium ions shown as balls. (b) The angle used to show most results in this report. The E-helix is pointed left and rotated so that the backbone part of Glu 20 is approximately above or between the camera and the axis of the helix. (Kretsinger, Uversky et al. 2013)

### P3W

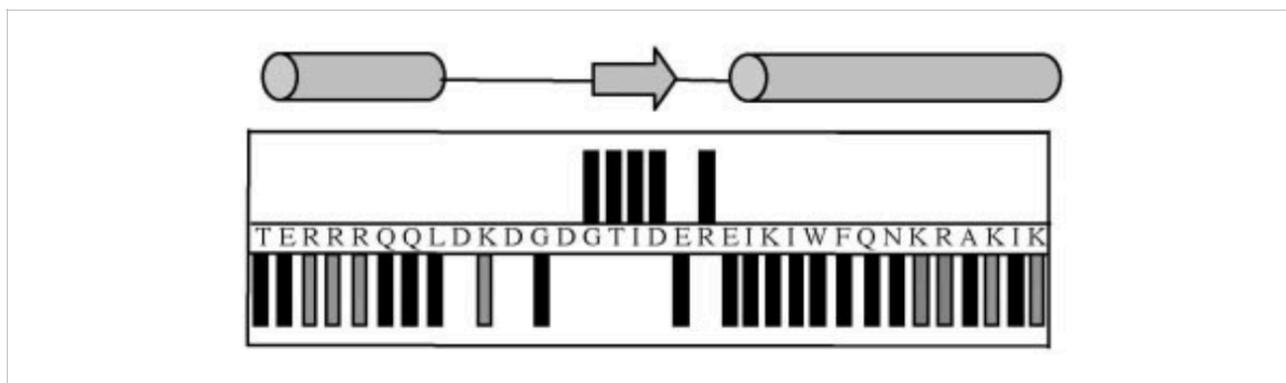
The protein of interest in this study is a single EF-hand like HTH motif that has been engineered by combining parts of proteins named *calmodulin* and *engrailed* in order to be able to bind ions such as  $\text{Eu}^{3+}$  and  $\text{La}^{3+}$ . The goal of the designers in this case was to create a very small protein that could cleave DNA with at least some sequence selectivity.

The solution structure of the molecule with  $\text{Eu}^{3+}$  bound to it has been studied using nuclear magnetic resonance (NMR) and it provides a convenient starting point for the current study. The distances from the ion to its ligand oxygens determined by Joel T. Welch (2003) are used in simulations in this study.

A secondary structure determined using chemical shift analysis is shown in figure 2. The secondary structure looks very much like an EF-hand, there are two helices with a loop and a  $\beta$ -strand in between. (Joel T. Welch 2003)

Unfortunately there is no 3D structure file for this protein available in the protein databank.

## Introduction



**Figure 2:** Secondary structure prediction of P3W via  $\alpha$ -proton chemical shift analysis. (Joel T. Welch 2003)

It has been experimentally determined that the protein is relatively disordered when no ion is bound. (Kovacic, Welch et al. 2003)

It is known that Trp 24 has an important effect in the stabilization of the secondary structure of P3W. This might be due to hydrophobic interactions with neighboring side chains. NMR data has shown that Trp 24 and Phe 25 become a part of a hydrophobic pocket upon the binding of a suitable ion. (Joel T. Welch 2003)

The structure is still flexible when an ion is bound.  $\text{Eu}^{3+}$  is located within 10 Å of residues Asp 9, Asp 11, Asp 13 and Glu 20.  $\text{Eu}^{3+}$  appears to have two water molecules interacting with it directly. Unlike traditional EF-hand proteins, P3W is a monomer in solution. (Joel T. Welch 2003)

## FRET

FRET is a method of energy transfer between molecules without converting to light in between. This process is highly dependent on distance. The efficiency of the process diminishes rapidly with increasing distance between the chromophore molecules. (Hussain 2012)

This distance dependence makes it a useful phenomenon for building biosensors, allowing the conversion of a conformational change to a visual signal.

## Structure modeling

Homology Modeling is used to create 3D models of proteins that do not have experimentally determined structures. The structure is then predicted based on homology with other structures that have had their 3D structures determined experimentally. Various tools exist to do this, such as YASARA, SWISS-MODEL, and RaptorX. These tools use different approaches to try to achieve the same goal.

SWISS-MODEL and YASARA both use a traditional homology modeling approach where the backbone is assembled from rigid pieces of template molecules. Insertion and deletion areas are modeled using constraint space programming and using loop libraries in the case of SWISS-MODEL. YASARA just uses the PDB as a loop library.

## Introduction

In SWISS-MODEL, side chains are placed on the backbone using a rotamer library, the different configurations are scored to select the most favorable, taking into account interactions such as hydrogen bonding, disulfide bridging, and close contacts. (Schwede 2003)

In YASARA, side chains are first simply placed on the backbone and optimized using dead end elimination, which is based on repulsive forces. After that, loop regions are optimized by trying out a large number of possible configurations of the backbone as well as the side chains. Next comes fine tuning of side chain rotamers using knowledge-based potentials, followed by optimization of the hydrogen bond network. (Krieger 2017)

After these steps, the energy of the system is minimized using a steepest descent energy minimization. (Schwede 2003) (Krieger 2017)

This process may be repeated many times for different template molecules, the results are analyzed and scored using various checking techniques.

In YASARA structure, these techniques are triggered using the “Check” command and include isomer, peptide bond, naming convention, water positioning, bond lengths and angles, dihedral (Phi-Psi) angles, non-bonded interactions, and packing checks. (Krieger 2017)

In SWISS-MODEL, the models are scored using the QMEAN tool, which considers a torsion angle potential over three consecutive amino acids, distance dependent interactions between C $\beta$  and all atoms, a solvation potential, and finally comparisons of predicted and calculated secondary structure, and the same for solvent accessibility. (Benkert, Künzli et al. 2009)

QMEANDisCo uses an additional distance term comparing distances between C $\alpha$  atoms with those of a model based on homologous structures. (Biozentrum 2017)

The final score returned by QMEAN is a linear combination of all the component scores, which are all normalized to lie between 0 and 1. The final score is also normalized to lie between 0 and 1, where 1 is the perfect score. The result is also converted to a Z-score. (Benkert, Künzli et al. 2009)

RaptorX takes a somewhat different approach. It was originally developed to model proteins with low template homology. It can combine several low quality templates to create hybrid models. It uses a nonlinear function that takes into account the number of available non-redundant homologs to score its templates. In this way, templates with large variance in their known structures are given less weight in the selection of best alignments. (Källberg, Wang et al. 2012)

In RaptorX, the actual modeling is done as the last step based on the highest scoring template, unlike SWISS-MODEL and YASARA, where models are created the best one is selected. (Källberg, Wang et al. 2012)

### **Knowledge-based potentials**

Knowledge-based potentials, unlike physics-based potentials, are artificial force fields that have been found to be helpful for the optimization of protein structures towards various properties. For

## Introduction

example, one potential could be really useful to find the native state and another might be good to check for the correctness of the fold of the protein. These potentials are based on the assumption that the more common a given structure is, the lower its energy is, in other words, the Boltzmann principle holds. (Sippl 1995)

Statistical potentials that are based on database statistics are generally of the form  $E(r) = -kT\ln[f(r)]$  where  $k$  is the Boltzmann constant,  $T$  is the absolute temperature and  $f(r)$  is the probability density for a given value of  $r$ , which can be distance or angle or any other measurable parameter. (Sippl 1995)

### QMEAN

QMEAN is a tool to estimate the quality of protein models, it is used internally by SWISS-MODEL to rank its candidate models. It considers six knowledge-based potentials, a torsion angle potential over three consecutive amino acids, distance dependent interactions between  $C\beta$  and all atoms, a solvation potential and finally, comparisons of predicted and calculated secondary structure and the same for solvent accessibility. (Benkert, Künzli et al. 2009)

Distance dependent interactions between  $C\beta$  and all atoms (2 potentials) are calculated using a knowledge-based potential that was obtained by observing a high resolution set of experimentally determined structures. Pairs of atoms that are less than four residues apart are excluded from this analysis. Further, secondary structure is taken into account so that an atom in an alpha helix is judged by a potential that was obtained from a dataset where at least one of the atoms belonged to an alpha helix. (Benkert, Tosatto et al. 2008)

The solvation potential is another knowledge-based potential that counts the number of  $C\beta$  atoms within 9 Å of an amino acid. This is meant to judge how accessible or inaccessible the residue is to the solvent. This potential is based on the idea that certain residues are more hydrophobic and are therefore more likely to be situated on the inside of structures. (Benkert, Tosatto et al. 2008)

The torsion/Psi-Phi/dihedral angle potential over three consecutive amino acids is calculated according to

$$E_{torsion}^a(a, \phi_{i-1}, \psi_{i-1}, \phi_i, \psi_i, \phi_{i+1}, \psi_{i+1}) = RT \ln(1 + M_a \sigma) - RT \ln \left( 1 + M_a \sigma \frac{f_{observed}(a)}{f_{reference}} \right)$$

where  $\sigma$  is a weight constant applied to each observation, commonly 0.01, and  $M_a$  is the number of times type  $a$  is observed.  $f_{observed}$  shows how frequently the local conformation occurs with the angles described by type  $a$  and  $f_{reference}$  shows how frequently the conformation at index  $i$  occurs at all. (Benkert, Tosatto et al. 2008)

Comparison of predicted and calculated secondary structure uses PSIPRED to predict the secondary structure of the protein, and then divides the number of residues that agree between the two models with the total number of residues, obtaining the score as a value between zero and one. (Benkert, Tosatto et al. 2008)

## Introduction

Comparison of predicted and calculated solvent accessibility uses ACCpro to predict, which residues are accessible to the solvent and which ones are not. The observed solvent accessibility is calculated using DSSP, and a cutoff value of 25% is used to convert the relative accessibility to binary, meaning that residues with over 25% surface accessible to the solvent are counted as accessible, and the ones below that threshold are counted as inaccessible. Finally, a fraction is obtained in a similar fashion as in the previous technique. (Benkert, Tosatto et al. 2008)

QMEANDisCo uses an additional distance term comparing distances between C $\alpha$  atoms with those of a model that is based on homologous structures. The tool does a search to find homologous models for the model being investigated, comparing the corresponding distances of those models with the one being targeted. This method is more accurate if there are many close homologues. The scores from each of the homologues are averaged with a weighted average where the weights are created using machine learning. (Biozentrum 2017)

The final score returned by QMEAN is a linear combination of all the component scores, which are all normalized to lie between 0 and 1 themselves. The final score is also normalized to lie between 0 and 1, where 1 is the perfect score. The result is also converted to a Z-score. (Benkert, Künzli et al. 2009)

# Materials and methods

## Homology modeling

As there is no published PDB structure for P3W, structure modeling was needed. This was done using three different tools — YASARA structure 15.9.6, SWISS-MODEL workspace, and RaptorX. (Krieger, Joo et al. 2009) (Källberg, Wang et al. 2012) (Jianzhu, Sheng et al. 2013) (Jian and Jinbo 2011) (Jian and Jinbo 2011)

SWISS-MODEL and RaptorX models were created by submitting the sequence of P3W for modeling through their respective web interfaces and waiting for the results. YASARA required slightly more input: a BLAST search was performed to find two homologues of P3W with known structures. These structures were provided to YASARA along with a sequence alignment of P3W with the two homologues in FASTA format. The alignment was created using the online tool called Clustal Omega.

## Structural alignment

Structural alignment was done using the *AlignObj* command in YASARA. This command uses the MUSTANG program to try and minimize the RMSD of C $\alpha$  atoms in the backbone of the objects. It returns a structural alignment over residues that are positioned within certain parameter values, and the RMSD over that alignment. (Krieger 2017) (Konagurthu, Whisstock et al. 2006)

The parameters used in this report are default values, DisMax=3.75 Å, AngleMax=75.00, LenMin=3, GapOpen=10, GapExtend=2, Overhang=1. These mean that two residues are considered structurally aligned if their distance is less than 3.75 Å, their C $\alpha$ -C $\beta$  bond angles are within 75 degrees of each other, the angle constraint is ignored if both neighboring residues are aligned based on their angles. (Krieger 2017)

LenMin sets the shortest allowed fragment in an alignment. GapOpen and GapExtend set the penalty for gaps in sequence alignments so that Penalty = GapOpen + GapLength \* GapExtend. Overhang sets the penalty for edge overhangs in sequence alignments. (Krieger 2017)

## Simulations

All molecular dynamics simulations, calculations and visualizations were performed using YASARA structure. Some of the results were obtained using version 15.9.6, most were obtained using 17.4.14. The results before the update have a special note. (Krieger and Vriend 2014) (Krieger and Vriend 2015)

The simulations where the older version of YASARA was used, were ran using the AMBER03 forcefield, with the newer version of YASARA, AMBER14 was used instead. (Duan, Wu et al. 2003) (Hornak, Abel et al. 2006)

Ray traced images of results are created inside YASARA using POVRay. (Krieger and Vriend 2014)

## Materials and methods

### MDAnalysis

The process used to find the conformation of the protein under different conditions, termed *MDAnalysis* in this report, is based on the default YASARA macros called *md\_analyze* and *md\_run*. (Krieger, Nielsen et al. 2006) (Essman, Perera et al. 1995) (Li, Roberts et al. 2013) (Li, Song et al. 2015) The process started by initializing the simulation, creating a cell around the protein, and filling the empty space with water. Salt concentration and pH were kept at physiological levels. In the first stage, a long MD simulation of 10 ns was performed with temperature  $T_0$ , taking 100 snapshots. This was supposed to probe the phase space. After the simulation, all snapshots were loaded and had their energies brought down to the local minima by simulated annealing. The five lowest energy snapshots after the initial equilibration period were selected for the second stage of MDAnalysis. In the second, refinement stage, each of the five snapshots is minimized by performing a short MD simulation at  $T_1$ , and selecting the lowest energy snapshot, again by simulated annealing, excluding snapshots from the beginning equilibration period. This process is repeated until the energy obtained was no longer lower than the one obtained in the previous step. Finally, the lowest energy snapshot was determined and written out to a separate file for further analysis. (Ernits, Bolotakis et al. 2016)

Simulations with  $T_0 = 298\text{K}$  and  $T_1 = 400\text{K}$  were performed with the reasoning that running at room temperature would probe the phase space and then running at the high temperature for a short period of time would help to overcome potential barriers, and hopefully make it possible to get to the global energy minimum structure. This approach was later abandoned in favor of the one described in the next sentence, because the results of the 400K simulations would always produce higher energy configurations than room temperature ones, and therefore their results were never used.

In cases where  $T_0 = 315\text{K}$  and  $T_1 = 298\text{K}$ , the goal was to simulate the system close to the melting temperature of P3W, which was determined to be between 315K and 330K by running simulations at both temperatures, and seeing that the protein would lose its secondary structure at the higher temperature, probing the phase space and finally running at room temperature to allow the system to settle towards lower energy configurations, hoping to find the global minimum.

### Restraining potential

Distance restraints were created in YASARA using the *RestrainDis* command, which creates distance restraint which adds a potential to the force field in the form of a Soft-square potential, described in the YASARA manual. In some cases, the potential was scaled by a constant. If the name of an experiment contains a suffix in the form *s<number>*, it should be read to mean that the restraint potentials in that experiment were scaled by the number. For example “Ca<sup>2+</sup> s25” refers to an experiment involving Ca<sup>2+</sup> where restraint potentials were scaled by a factor of 25. The force scaling is only applied to the restraints during the initial phase of the MDAnalysis, and restraints are reverted to unscaled versions during the refinement phase.

Restraints were generally made so that the ion being bound was restrained to stay between 2 and 3 Å of single Oδ atoms of the asparagine residues at numbers 9, 11 and 13, the backbone carbonyl

## Materials and methods

oxygen of threonine 15 and both O $\epsilon$  atoms of glutamine 20. These restraints were determined to be correct by Joel T. Welch et al. (2003) and the distances were determined to be valid for all ions studied in this report by Zheng, Chruszcz et al. (2008). These ions are Ca<sup>2+</sup>, Mg<sup>2+</sup>, Fe<sup>2+</sup>, Fe<sup>3+</sup> and Eu<sup>3+</sup>.

### Distance measurements

Distances were measured in YASARA by highlighting the atoms whose distances were being measured, and noting down the distance that automatically appeared on the screen. The resulting value marks the distance between the center points of the atoms. All distances in this report are center to center distances.

### FRET efficiency

The efficiency of FRET was calculated using the following formula:  $E_{\text{FRET}} = R_0^6 / (R_0^6 - r^6)$  where  $R_0$  is the Förster radius, where  $E_{\text{FRET}}$  reaches 50%, and  $r$  is the radius between the fluorophores. (Hussain 2012)

In the case of GFP and mCherry, which is a kind of red fluorescent protein, this value is around 51 Å. (Albertazzi, Arosio et al. 2009)

### Binding energy

Binding energy in YASARA was calculated by adding all potential energies together in the case where the ion is in place, and when the ion is moved infinitely far away from the solution. More positive binding energy means that there is higher affinity. This is done automatically by invoking the *BindEnergyObj* command. For this to work, the simulation cell is first converted into a cubic cell with wall boundaries.

### Restraint energy

Restraint energy in YASARA was calculated by invoking the *RestEnergy* command. This returns the current summed up energy of all restraints in the soup.

# Results and discussion

## Homology modeling

The first step was to create a 3D model of the protein, and this was done via homology modeling. The sequence was modeled with three different tools— YASARA, Swissmodel, and RaptorX — the results of which were compared and one of which was chosen. A comparison of the three models can be seen in figure 3. All of these models appear to have the same basic structure and the differences seem to be relatively minor, especially in the helical regions. The model created by YASARA was chosen for subsequent simulations.

The YASARA model was created by supplying the algorithm with two homologous protein structures called “Solution structure of Engrailed homeodomain L16A mutant”, PDB code *1ztr* and “Engrailed homeodomain helix-turn-helix motif” with PDB code *2p8l*. The model resulting from *2p8l* was chosen as the result. It was chosen because its overall quality Z-score of -0.054 was better than the score of -1.314 for the model created from *1ztr*. YASARA also tried to create a hybrid model, but did not manage to improve the quality that way. The alignment of these sequences with that of *p3w* looks like this:

```
>P3W
-----TERRRQQLDKDGDGTIDEREIKIWFQNKRAKIK--
>2P81:A|PDBID|CHAIN|SEQUENCE
AKREFNENRYL-----TERRRQQLSS--ELGLNEAQIKIWFQNKRAKIKKS
>1ZTR:A|PDBID|CHAIN|SEQUENCE
GDEKRPRTAFSSEQLARAKREFNENRYLTERRRQQLSS--ELGLNEAQIKIWFQNKRAKIRRS
```

Swissmodel used a different template, “Solution structure of Engrailed homeodomain WT” with PDB code *2jwt.1.A*, to create its model. (Arnold, Bordoli et al. 2006) (Biasini, Bienert et al. 2014) (Benkert, Biasini et al. 2011)

```
P3W      TERRRQQLDKDGDGTIDEREIKIWFQNKRAKIK
2jwt.1.A  TERRRQQLS--SELGLNEAQIKIWFQNKRAKIK
```

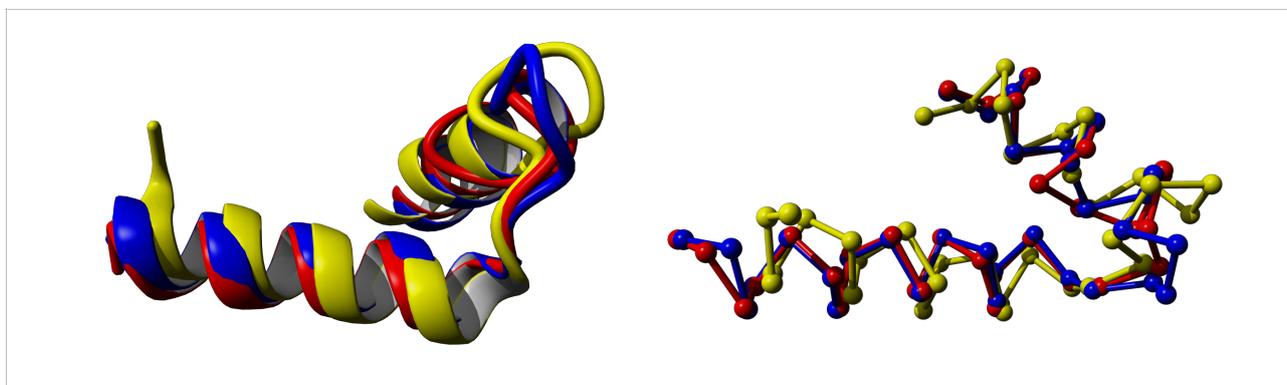
Finally, RaptorX used PDBs with identifiers *1b8iA*, *1ftzA*, *2r5yA*, *4cycA* and *9antA* to create the template for its model. These identifiers correspond to PDB Entries titled “STRUCTURE OF THE HOMEOTIC UBX/EXD/DNA TERNARY COMPLEX”, “NUCLEAR MAGNETIC RESONANCE SOLUTION STRUCTURE OF THE FUSHI TARAZU HOMEODOMAIN FROM DROSOPHILA AND COMPARISON WITH THE ANTENNAPEDIA HOMEODOMAIN”, “Structure of Scr/Exd complex bound to a consensus Hox-Exd site”, “CRYSTAL STRUCTURE OF A UBX-EXD-DNA COMPLEX INCLUDING THE HEXAPEPTIDE AND UBDA MOTIFS” and “ANTENNAPEDIA HOMEODOMAIN-DNA COMPLEX” in the same order.

An alignment of these template protein sequences is shown below.

## Results and discussion

```
>P3W
-----TERRRQLDKDGDGTIDEREIKIWFQNKRAKIK---
>1b8iA
EKEFHINHYLTRRRRIEMAHAL--SLTERQIKIWFQNRMKLKKEI
>1ftzA
EKEHFNRYITRRRRIDIANAL--SLSERQIKIWFQNRMKSKKDRTLDSPE
>2r5yA
EKEHFNRYLTRRRRIEIAHAL--SLTERQIKIWFQNRMKWKKEHK
>4cycA
EKEFHINHYLTRRRRIEMAHAL--CLTERQIKIWFQNRMKLKKEIQAIK
>9antA
EKEHFNRYLTRRRRIEIAHAL--SLTERQIKIWFQNRMKWKKEN
```

A visual comparison of the three resulting models is shown in figure 3. These alignments were done by aligning the models produced by SWISS-MODEL and RaptorX with the one that was produced by YASARA.



**Figure 3:** Comparison of homology models from different sources. Yellow - YASARA, red - Swissmodel, blue - RaptorX. Ribbon representation shown on the left and a trace through C $\alpha$  atoms shown on the right. Alignments were done so that RaptorX and SWISS-MODEL models were both aligned with the YASARA model. The models appear to agree quite well, the yellow model was chosen as the base for all subsequent simulations.

The output of all three possible pairwise alignments are shown in table 1. The RMSD values in parentheses represent the RMSD that was calculated based on all C $\alpha$  atoms, whereas the ones without parentheses were calculated based on the aligned residues shown in the last column of the table. The alignments as well as the images demonstrate that the binding loop region between the helices is difficult to predict as all three models produced significantly differing results in that area. None of the models have the  $\beta$ -strand from residues 14 to 17 that was determined via NMR and the E-helix starts already from residue number 17 instead of 21 as has been determined experimentally. (Joel T. Welch 2003)

The E-helix starting too early in the chain causes the binding loop to become crumpled, making it difficult to form a  $\beta$ -strand. This error was initially not noticed and that is why there was no attempt to search for alternative templates with high homology that have a  $\beta$ -strand in the area, however, the best overall sequence homology is necessarily with *engrailed*, from which the majority of the sequence, namely the helices, are taken from.

## Results and discussion

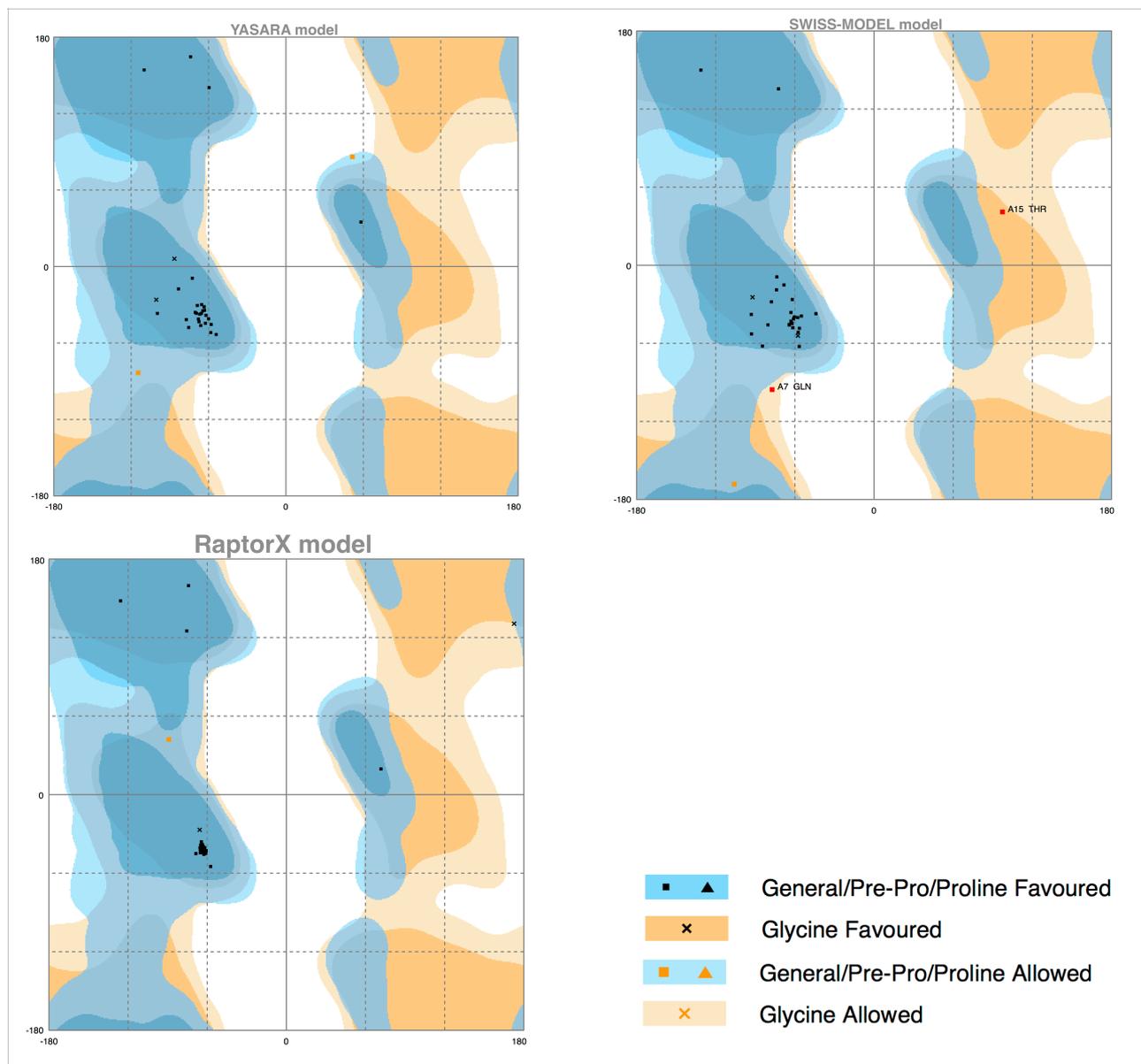
### Homology model backbone RMSD

	RMSD (Å)	Structural alignment (cutoff distance 3.75 Å)
<b>YASARA and SWISS-MODEL</b>	1.548 Å (2.865 Å)	Y: 1 T-ERRRQQLDKDG-----DGTIDEREIKIWFQNKRAKIK-- 34 ERRRQQ DGTIDEREIKIWFQNKRAK S: 1 -TERRRQQ-----LDKDG DGTIDEREIKIWFQNKRAK--IK 33
<b>YASARA and RaptorX</b>	1.497 Å (2.604 Å)	Y: 1 T-ERRRQQLDKDGDGT-----IDEREIKIWFQNKRAKIK-- 34 ERRRQQLDK IDEREIKIWFQNKRAK R: 1 -TERRRQQLDK-----DGDGTIDEREIKIWFQNKRAK--IK 33
<b>RaptorX and SWISS-MODEL</b>	0.536 Å (2.301 Å)	R: 1 TERRRQQLDKDGDG-----TIDEREIKIWFQNKRAKIK- 34 TERRRQQ TIDEREIKIWFQNKRAKI S: 1 TERRRQQ-----LDKDG DGTIDEREIKIWFQNKRAKI-K 33

**Table 1:** RMSD values and alignments produced by the AlignObj command in YASARA. RMSD values in parentheses are calculated separately over all 33 Ca atoms.

To validate the structures more, psi-phi plots were created using RAMPAGE, these are shown in figure 4. The worst performing model is the one created by SWISS-MODEL. It has two residues that are considered to be outliers and they both are situated in the binding loop region between the two helices. The model created using YASARA comes in second with no outliers, but two residues in the allowed-but-not-perfect region, one of these is even quite close to the edge of the allowed area. RaptorX created the most compliant model based on this test, it has one residue in the allowed areas, everything else is in the favored areas. (S.C. Lovell 2002)

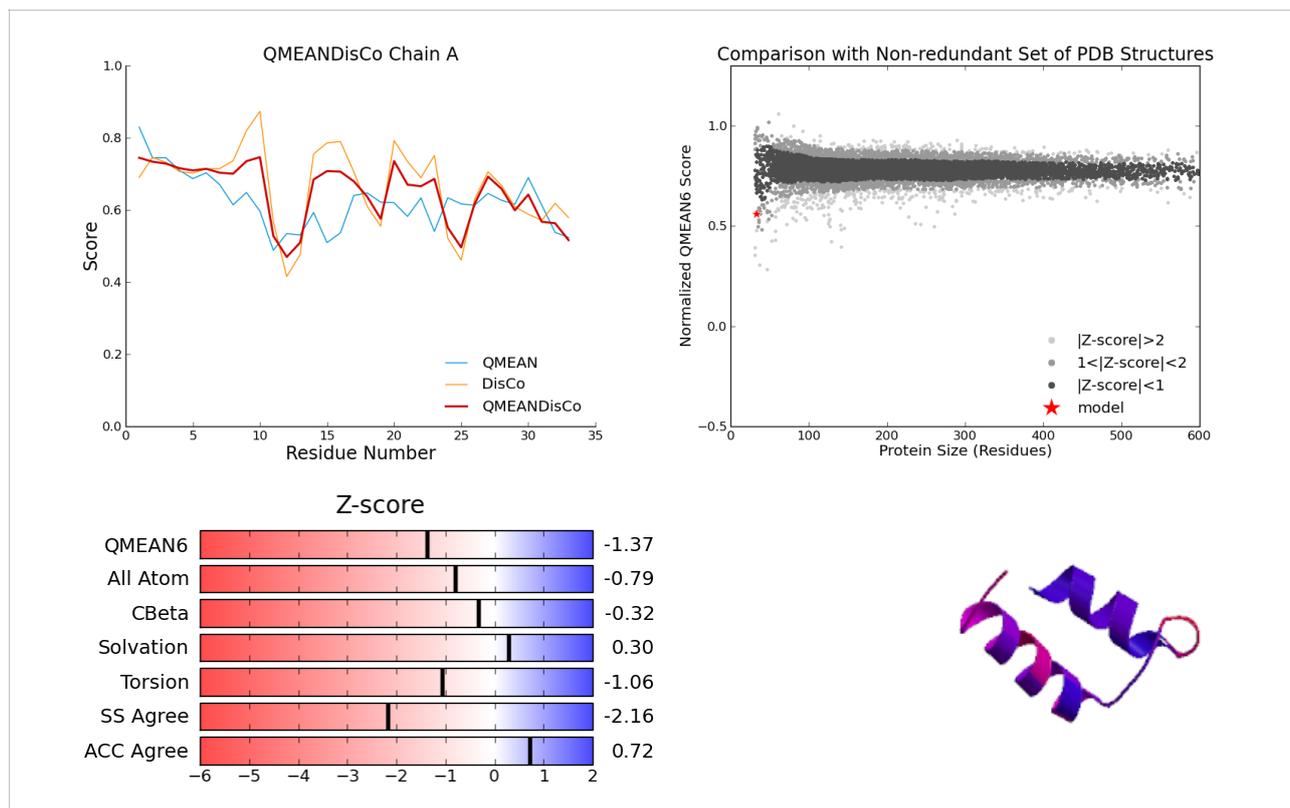
## Results and discussion



**Figure 4:** Comparison of homology models from different sources using psi-phi plots. RaptorX has the best performance in this test, with all residues except one positioned in the favored areas, there are no outliers. YASARA comes in second, it has two residues in the allowed region. Swiss-Model performed the worst in this regard, with two residues of the ion binding loop lying outside the allowed regions.

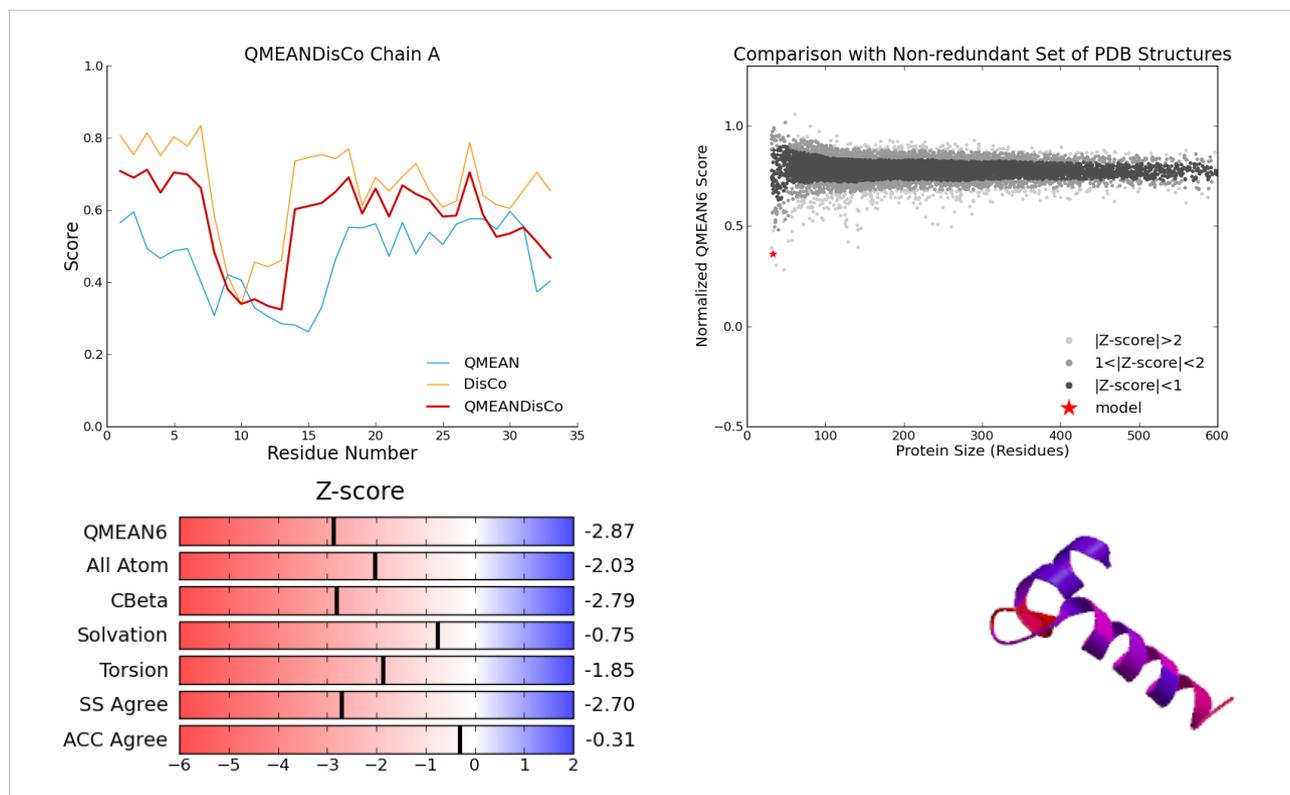
The quality of the models was further examined using the QMEAN tool, which is the one used internally by SWISS-MODEL. The results of the assessment of the model obtained from YASARA are shown in figure 5. The overall score places the model at the low end of acceptable models in terms of quality, assuming that structures in the protein databank fulfill that criterion. The local quality estimates are lowest in the binding loop area, where the homology was lowest, and towards the end of the sequence.

## Results and discussion



**Figure 5:** QMEAN scoring results for the YASARA model.

A similar overall picture appears to hold in the case of the model generated by SWISS-MODEL, seen in figure 6. In this case, however, all possible component scores are lower and the model appears to be much less adequate than the previous one.

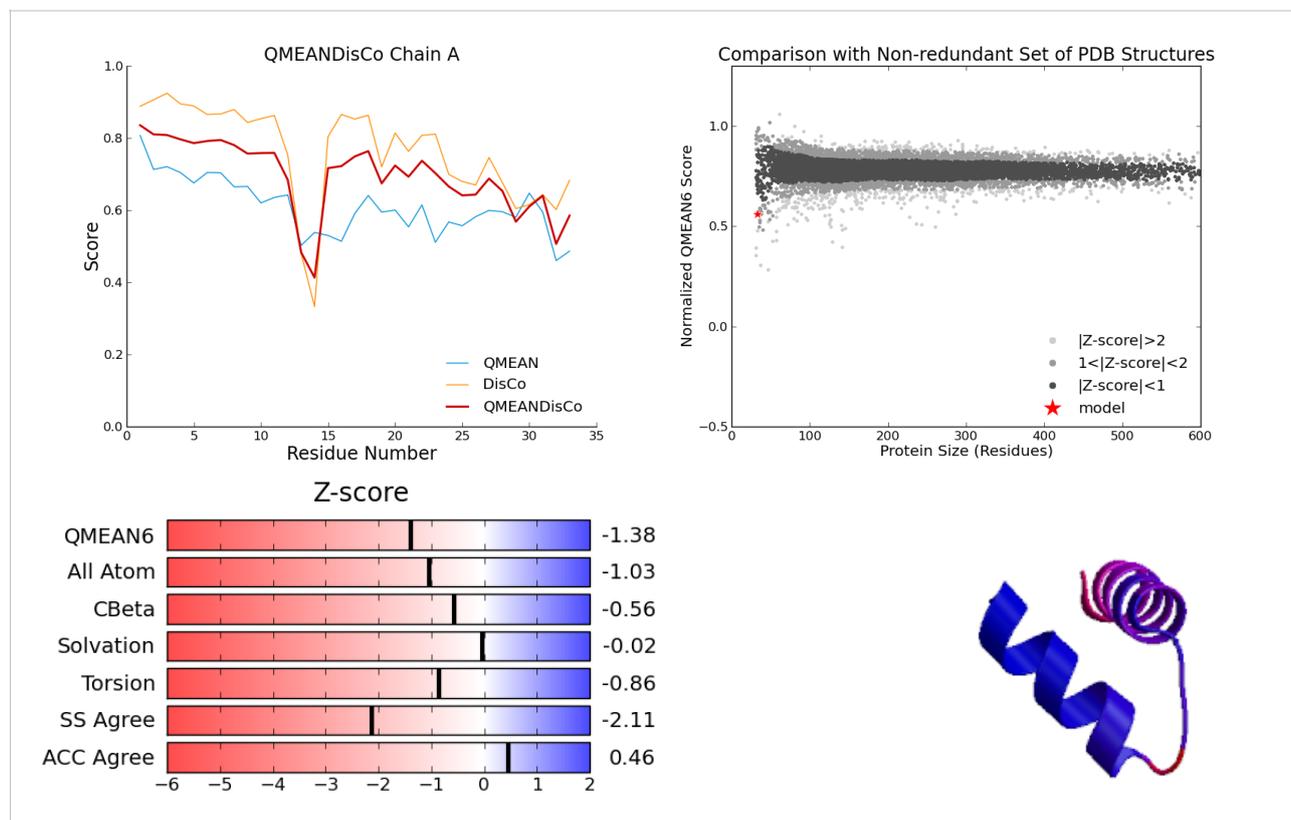


**Figure 6:** QMEAN scoring results for the SWISS-MODEL model.

## Results and discussion

Finally, the results of the QMEAN analysis of the model obtained from RaptorX are shown in figure 7. The final score is very close to the one obtained for the YASARA model. Most component scores of QMEAN6 are slightly worse, however the additional DisCo score appears to be overall somewhat better for the RaptorX model, except for residue number 14, which appears to be heavily distorted. The points that were made about the secondary structure agreement score for the YASARA model hold in this case as well.

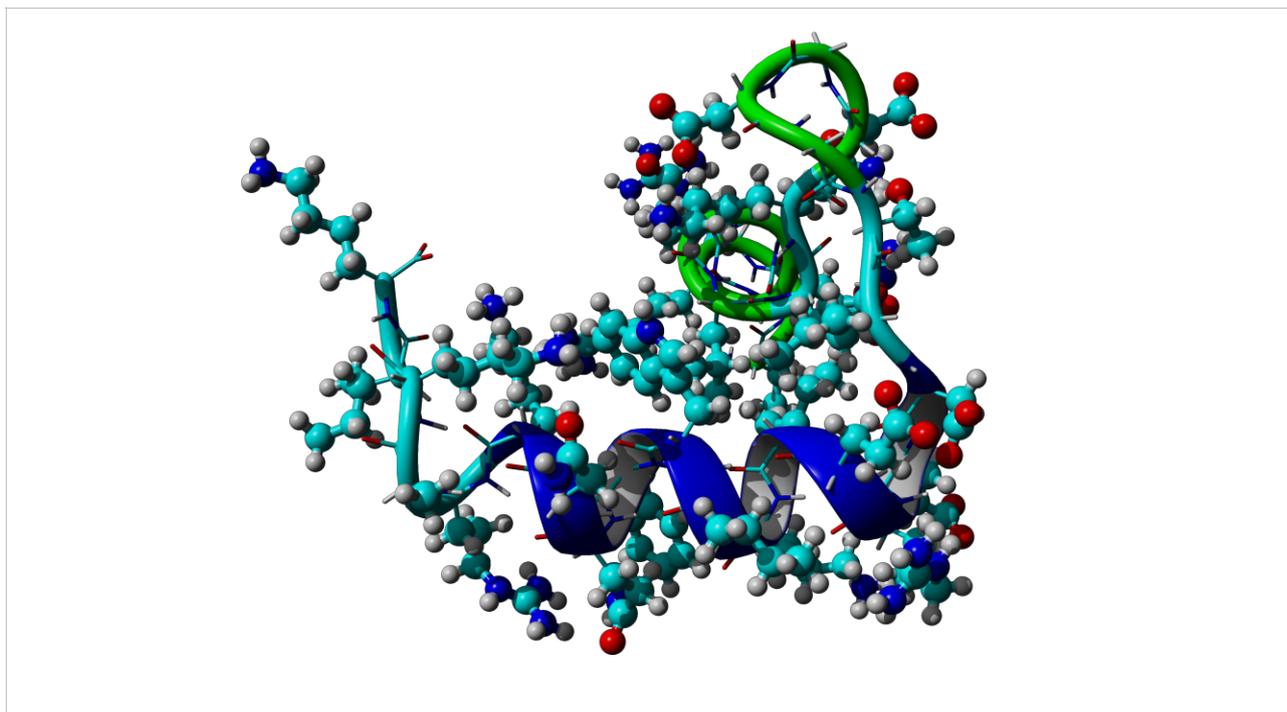
The model produced by YASARA was used in all subsequent simulations. Even though RaptorX produced a model that was slightly better by some estimates, the difference was not particularly large, and the YASARA model did perform slightly better according to QMEAN6. Results concerning the region between residues number 10 and 15 need to be treated with caution as these residues appear to have the lowest local quality. The additional fact that the models did not predict a  $\beta$ -strand between the helices and the E-helix was modeled to be too long means that all results should be viewed with caution.



**Figure 7:** QMEAN scoring results for the RaptorX model.

The lowest of the six main component scores is the secondary structure agreement score, which is explained by the PSIPRED secondary structure prediction in figure 8. Almost half of the long E-helix at the end of the sequence has been predicted to be coil instead of helix. As the protein is so small, this has a significant, nearly 1/5 effect on the component score. This does not necessarily mean that the prediction is wrong, however chemical shift analysis by Joel T. Welch (2003) does suggest that the helix structure reaches all the way to the end of the sequence, and that the QMEAN score is somewhat lower than warranted. The score in the binding loop region is unfortunately not

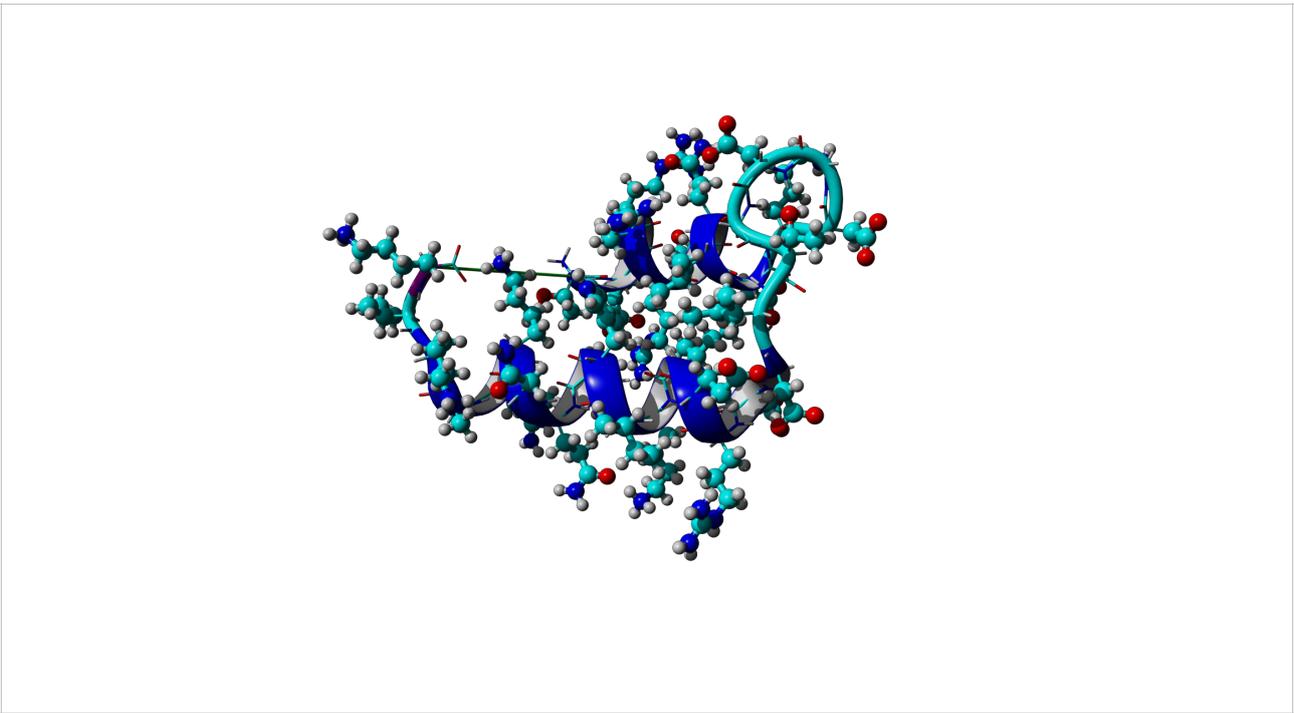




**Figure 9:** Result of the simulation of P3W with T0 set to 298K and T1 set to 400K. The structure appears to have started losing its secondary structure. The E helix has actually partly unwound to the extent predicted by PSIPRED, the F-helix is deformed but is maintaining a semblance of a helix.

### P3W restrained 298K

Next, the same kind of simulation was performed, only this time the distance between the C $\alpha$  atoms of the termini were affected with a restraining potential with parameters  $d = 16 \text{ \AA}$ ,  $d_{\text{minus}} = 15 \text{ \AA}$ ,  $d_{\text{plus}} = 17 \text{ \AA}$ , and the force was left unscaled. This means that the potential was keeping the termini between 1 and 32  $\text{\AA}$  apart. As the results show, this potential did not really get to have any significant effect on the conformation, as this time, the protein appears to have stayed very close to the original conformation. This is shown in figure 10.

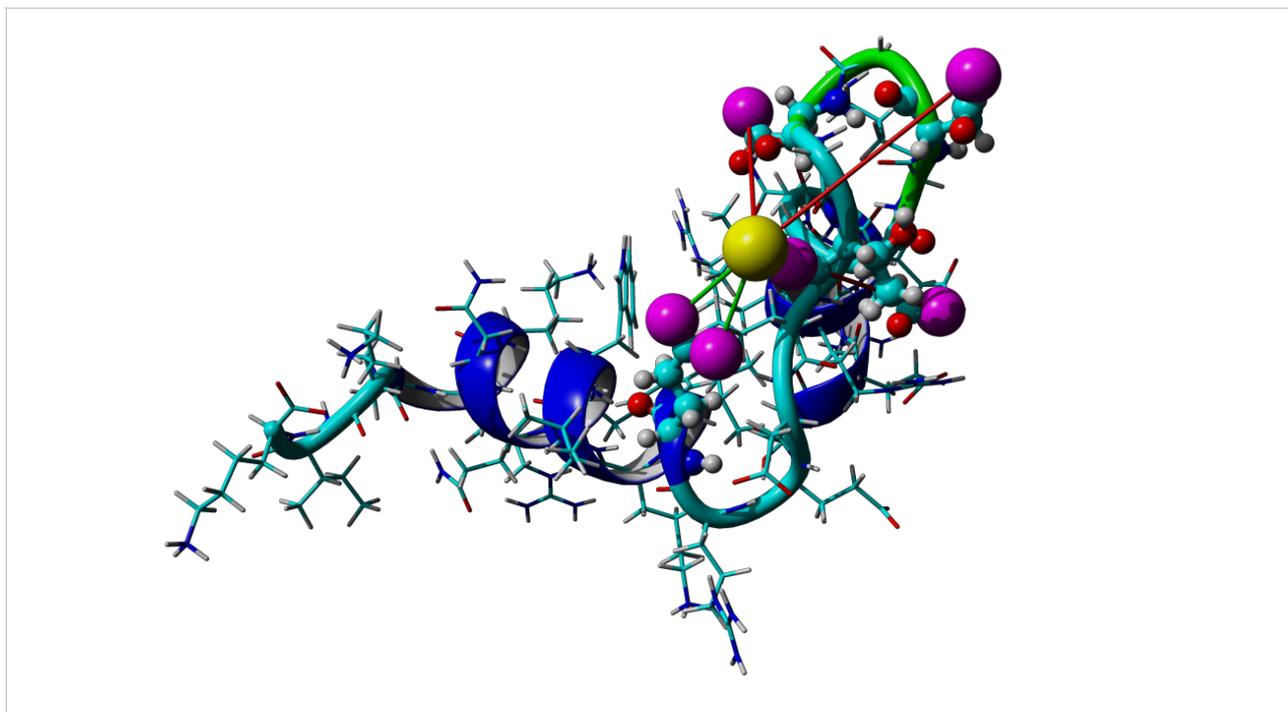


**Figure 10:** Result of the simulation with restrained end residues, T0 was set to 298K and T1 set to 400K. There appear to be no significant changes to the structure compared to the homology model.

### Ca<sup>2+</sup>

After this baseline setting, the next step was to start testing how the protein would bind ions. The first of these ions simulated is Ca<sup>2+</sup>, the result is shown in figure 11. The ion was restrained to the relevant oxygen atoms with restraint parameters  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5 \text{ \AA}$ . These restraints had quite a wide potential well, where there was no force acting on the atoms, and their strength was unscaled, making it so that they only kept the ion from escaping from the binding loop.

The ion appears to have only made an ionic bond to the backbone carbonyl oxygen of Thr 15, all the other restrained oxygens are more than  $3 \text{ \AA}$  away from the ion which is too far to be considered bound. The overall structure of the protein has not changed very much, the F-helix appears to be slightly more perpendicular to the E-helix than in the case of the homology model.

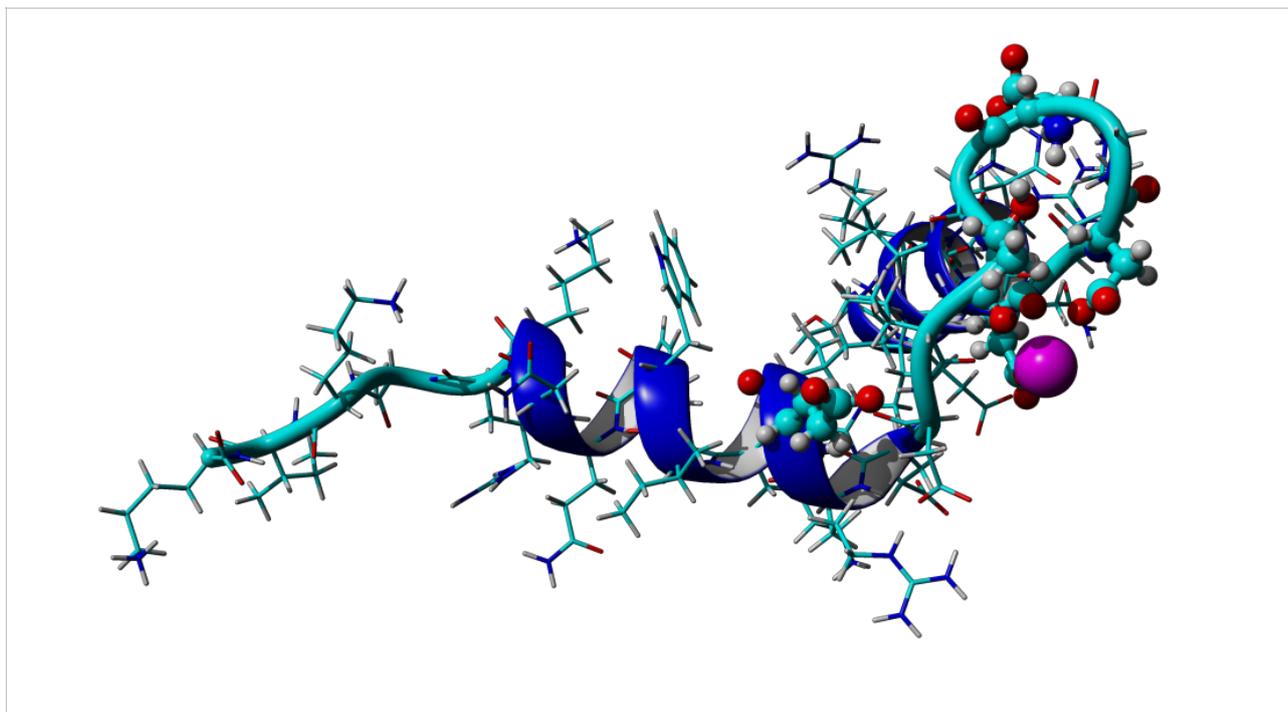


**Figure 11:** Result of the Ca<sup>2+</sup> simulation with restraints. Restraint  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5 \text{ \AA}$ , making the total width of the soft square well  $5 \text{ \AA}$ . The overall structure from the homology models seems to be retained, but the restraints have not had a significant effect.

### Fe<sup>3+</sup>

Fe<sup>3+</sup> was also tried, the result is shown in figure 12. The simulation was run with all the same parameters as the previous one. This ion had significantly more interactions with the protein. There were five oxygen atoms from the protein within  $3 \text{ \AA}$  of the ion, three of which were restrained to be there. The oxygens in the vicinity of the ion were OD1 and O of Asp 9, OD1 of Asp 11, O of Thr 15 and O of Ile 16. Restrained atom names have been bolded. This suggests that the amount of positive charge on the ion has a significant effect on binding.

Otherwise, the general structure of the protein remains largely unchanged from the homology model.



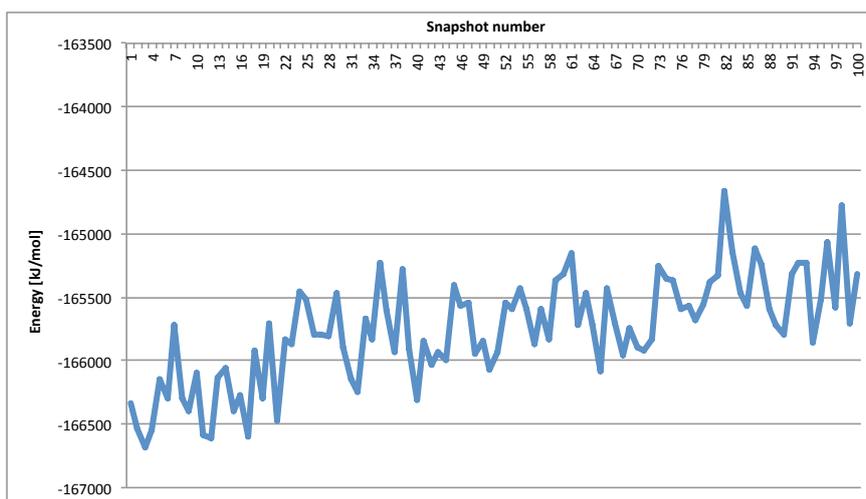
**Figure 12:** Result of the Ca<sup>2+</sup> simulation with restraints. Restraint  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5 \text{ \AA}$ , making the total width of the soft square well  $5 \text{ \AA}$ . The overall structure from the homology models seems to be retained but the restraints have not had a significant effect.

Next, simulations involving Eu<sup>3+</sup> were desired, because this could enable comparisons with existing results published by Joel T. Welch et al. (2003), however it turned out that the version of YASARA used up until this point did not have proper support for Eu<sup>3+</sup>, but the newest available version did, so the program was updated.

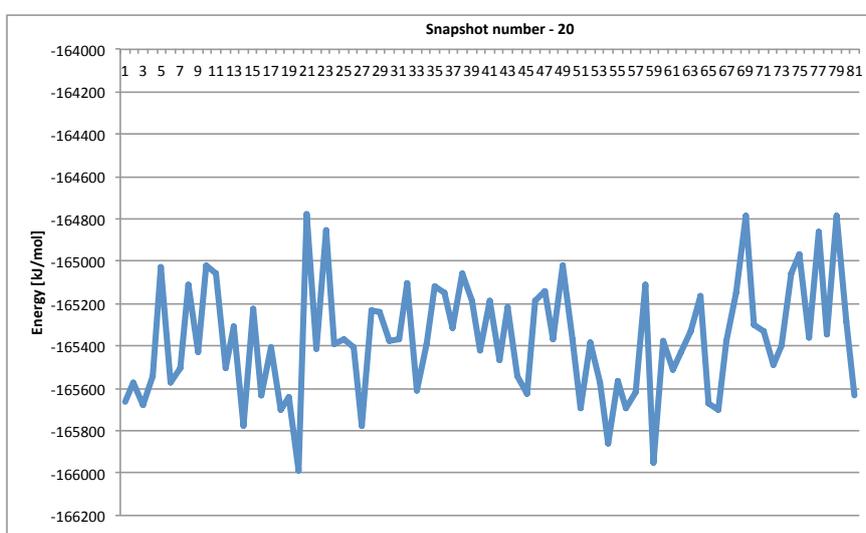
In addition to updating the program, the whole MDAnalysis strategy was reconsidered. It was realized that running at 298K first and 400K second was likely not the optimal way to probe the phase space, and it was decided to first run at some temperature that is close to the melting / denaturing temperature of P3W, to probe the phase space and only then run at 298K to try and settle towards a low energy state.

Further, the duration of simulation for the refinement simulations at T1 was increased from 0.5 ns to 2 ns, and a cutoff was added to exclude snapshots from the equilibration period from analysis. This move was driven by the discovery that 0.5 ns was not enough time for the system to equilibrate to the simulation temperature, and produce enough useful data from the plateau, an example of this is shown in figure 13.

## Results and discussion



(a)

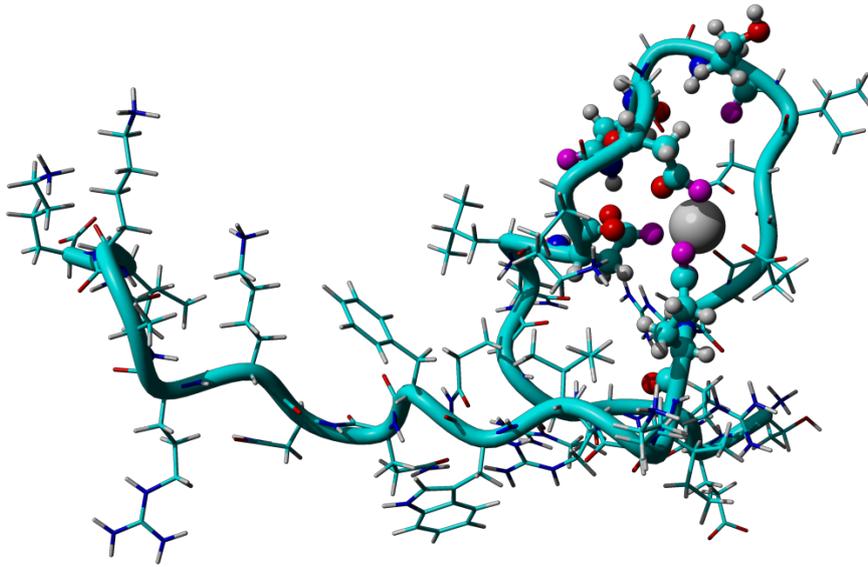


(b)

**Figure 13:** An example of the equilibration period at the start of a T1 simulation. (a) The simulation was 0.5 ns long and 100 snapshots were taken. For each snapshot, energy was minimized and the resulting energies were plotted on this chart. It can be seen that energy is starting to plateau around snapshot 80. (b) The simulation was 2 ns long and 100 snapshots were taken. For each snapshot after the first 19, energy was minimized and the resulting energies were plotted on this chart. It seems that the equilibration period has been excluded. This was obtained from a simulation with a distance restraint between the termini of the protein  $T_0 = 315\text{K}$  and  $T_1 = 298\text{K}$ .

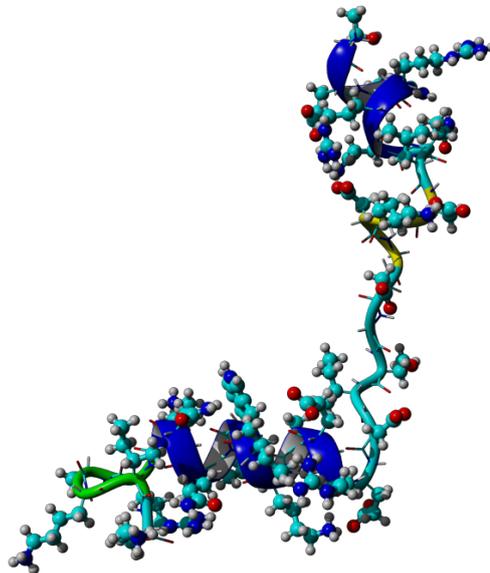
### Eu3+ 330K

To find the correct  $T_0$  temperature, simulations were run. The first of these simulations was run with  $T_0 = 330\text{K}$  and  $T_1 = 298\text{K}$ . The results of this test are shown in figure 14. This simulation included Eu3+ with restraints where  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5$  and the force was left unscaled. The protein appears to have lost all of its secondary structure. This points to the fact that 330K is above the denaturation temperature.



**Figure 14:** Result of the Eu3+ simulation with T0 set to 330K. The structure appears to have completely lost its secondary structure.

#### P3W 325K



**Figure 15:** Result of the simulation with T0 set to 325. The structure appears to have started losing its secondary structure.

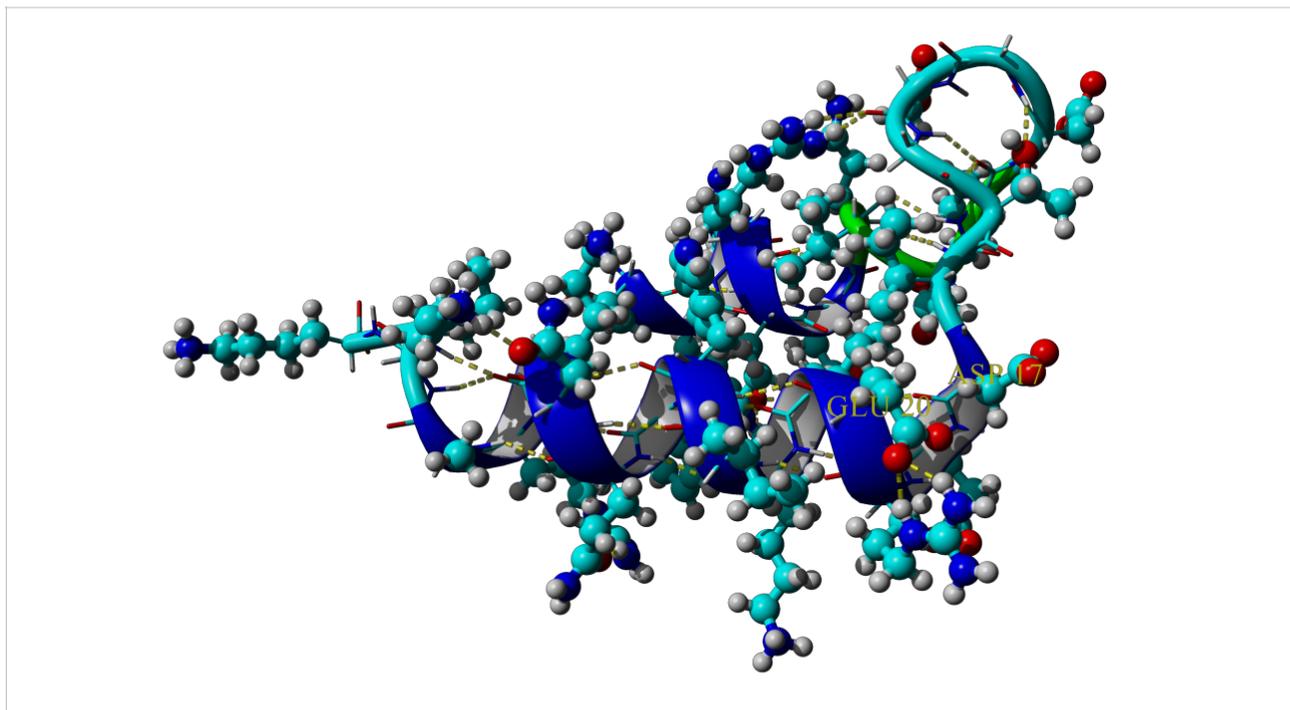
Next, a T0 value of 325K was tried, this time no ion was included in the simulation. The results of this simulation are shown in figure 15. The protein seems to have started losing its secondary structure, but most obviously the helices have split apart and the molecule has started to stretch out.

## Results and discussion

This suggests that 325K is very close to the melting temperature, therefore 315K was selected for further simulations.

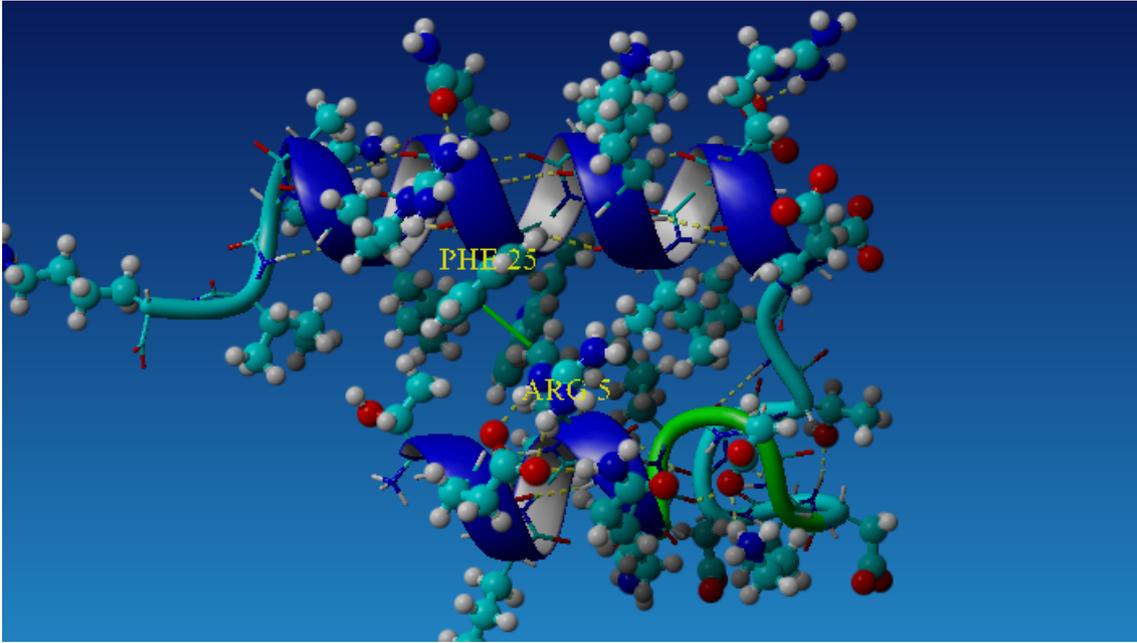
### P3W 315K

Using this new approach, a second attempt was made to simulate the protein with the first and the last residue restrained with parameters  $d = 16 \text{ \AA}$ ,  $d_{\text{minus}} = 15 \text{ \AA}$ ,  $d_{\text{plus}} = 17 \text{ \AA}$ , and unscaled forces. The result can be seen in figure 16. The result looks very similar to the original terminus distance restraining one, the helices are positioned almost in parallel.



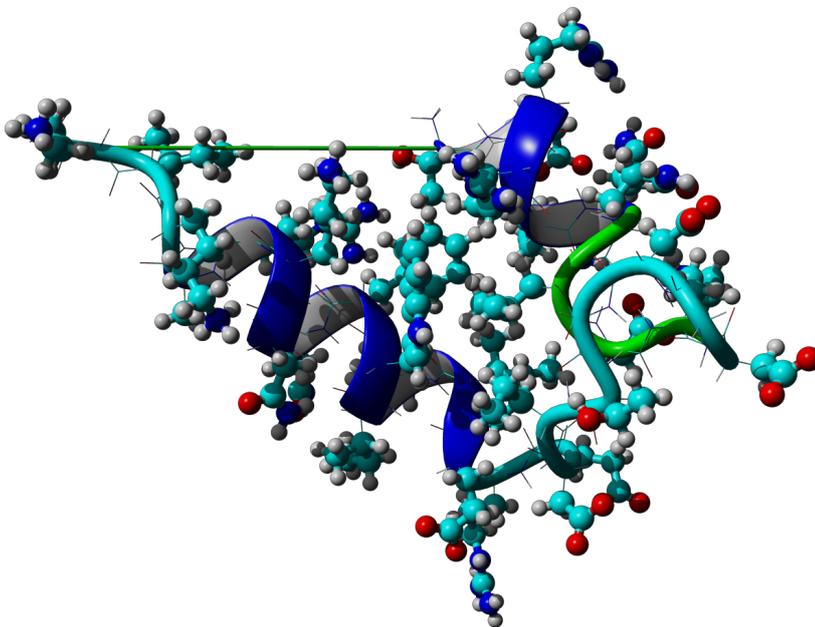
**Figure 16:** Result of the simulation with T0 set to 315K, T1 set to 298K. The structure appears to have the same structure as the original simulation with restrained termini above. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

Hydrophobic interactions and hydrogen bonds are shown in figure 17. Only one hydrophobic interaction was detected by YASARA. There are several hydrogen bonds that are stabilizing the binding loop.



**Figure17:** Result of the simulation with T0 set to 315K, T1 set to 298K. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. Showing the hydrophobic interaction between the helices.

#### P3W restrained 315K



**Figure 18:** Result of the simulation with restrained end residues and T0 set to 315K, T1 set to 298K. The structure appears to have the same structure as the original simulation with restrained termini above. The green line indicates the restraint.

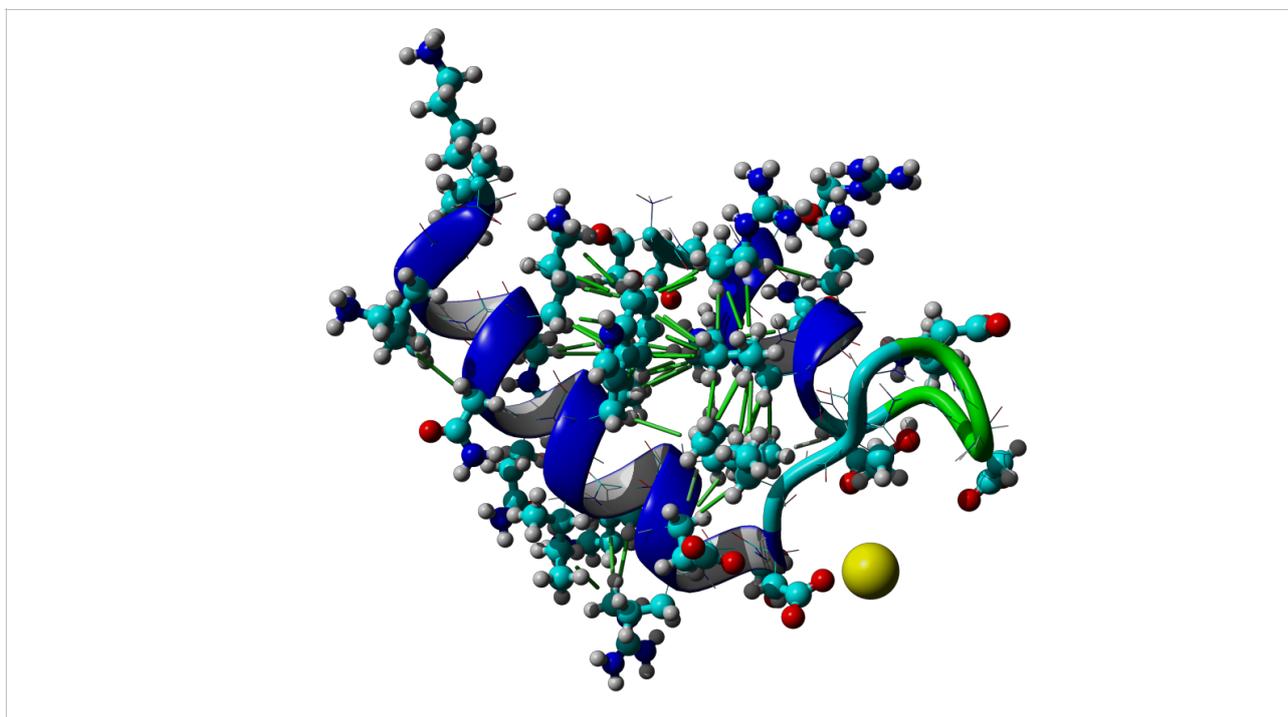
Using this new approach, a second attempt was made to simulate the protein with the first and the last residue restrained with parameters  $d = 16 \text{ \AA}$ ,  $d_{\text{minus}} = 15 \text{ \AA}$ ,  $d_{\text{plus}} = 17 \text{ \AA}$ , and unscaled forces.

## Results and discussion

The result can be seen in figure 18. The result looks very similar to the original terminus distance restraining one.

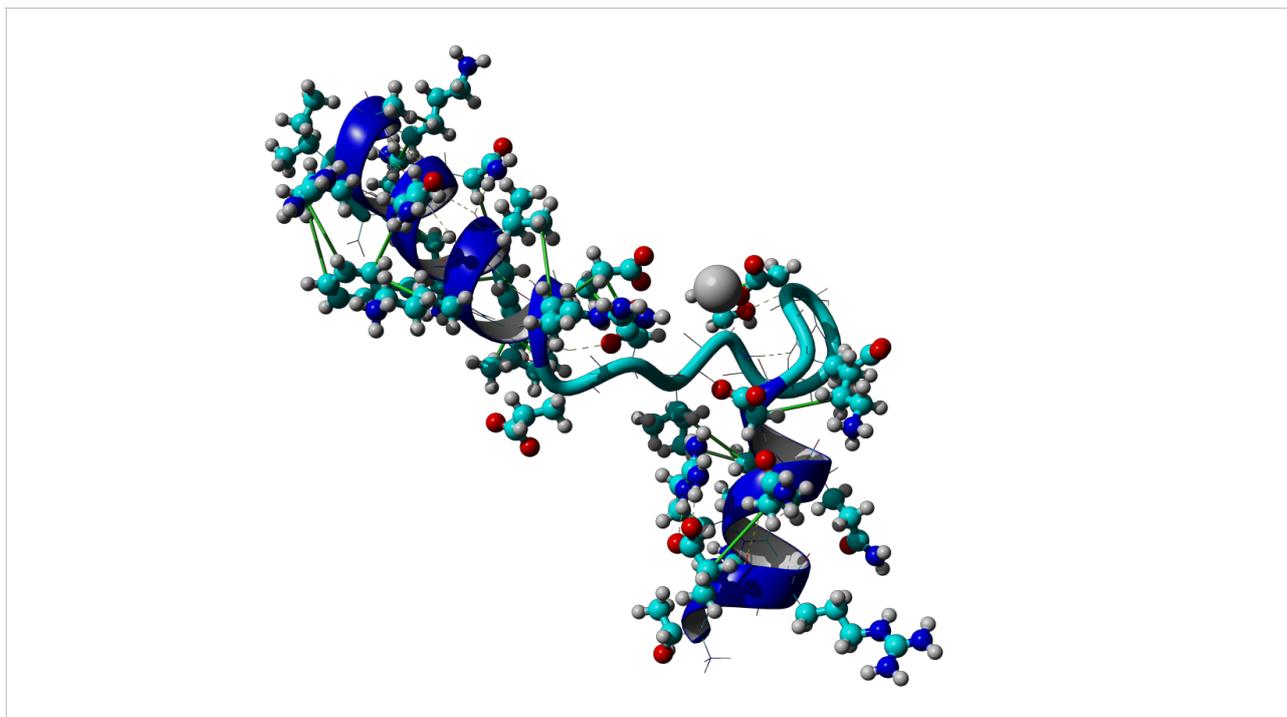
### Ca<sup>2+</sup> Wide restraint 315K

Next, it was time to try and see how the ion would behave when binding different ions. The first of these simulations was done with Ca<sup>2+</sup> to have an opportunity to compare with existing results, the result is shown in figure 19. Restraint parameters here were  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5 \text{ \AA}$ , and the force was again left unscaled. There does not seem to be many significant differences from the previous simulation, where there was no ion present, but the termini were restrained. The calcium has made only one ionic bond, that is with Asp 17, which was not even restrained to do so. This suggests that there might be something wrong with the restraints in this case.



**Figure 19:** Result of the simulation with Ca<sup>2+</sup> restrained,  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5 \text{ \AA}$ , T<sub>0</sub> was set to 315K and T<sub>1</sub> set to 298K. The structure appears to have the same structure as the simulation with restrained end residues above. The ion has not bound to any of the restrained residues. Green lines indicate hydrophobic interactions.

It was decided to give this configuration of restraints one more try with a different ion, Eu<sup>3+</sup>, to better understand the situation. The result is shown in figure 20. Restraint parameters here were again  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5 \text{ \AA}$ , and the force was again left unscaled. The overall structure of the protein has changed considerably. The two helices are no longer touching, and the F-helix has started turning away from the E-helix. The ion has made ionic bonds with both O $\delta$  atoms of Asp 11. This time the ion at least attached itself to a restrained residue, but these results still suggest that something important is wrong with the restraints.



**Figure 20:** Result of the simulation with  $\text{Eu}^{3+}$  restrained with  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5 \text{ \AA}$ , making the total width of the soft square well  $5 \text{ \AA}$ , forces were unscaled.  $T_0$  was set to 315K,  $T_1$  set to 298K. Green lines indicate hydrophobic interactions.

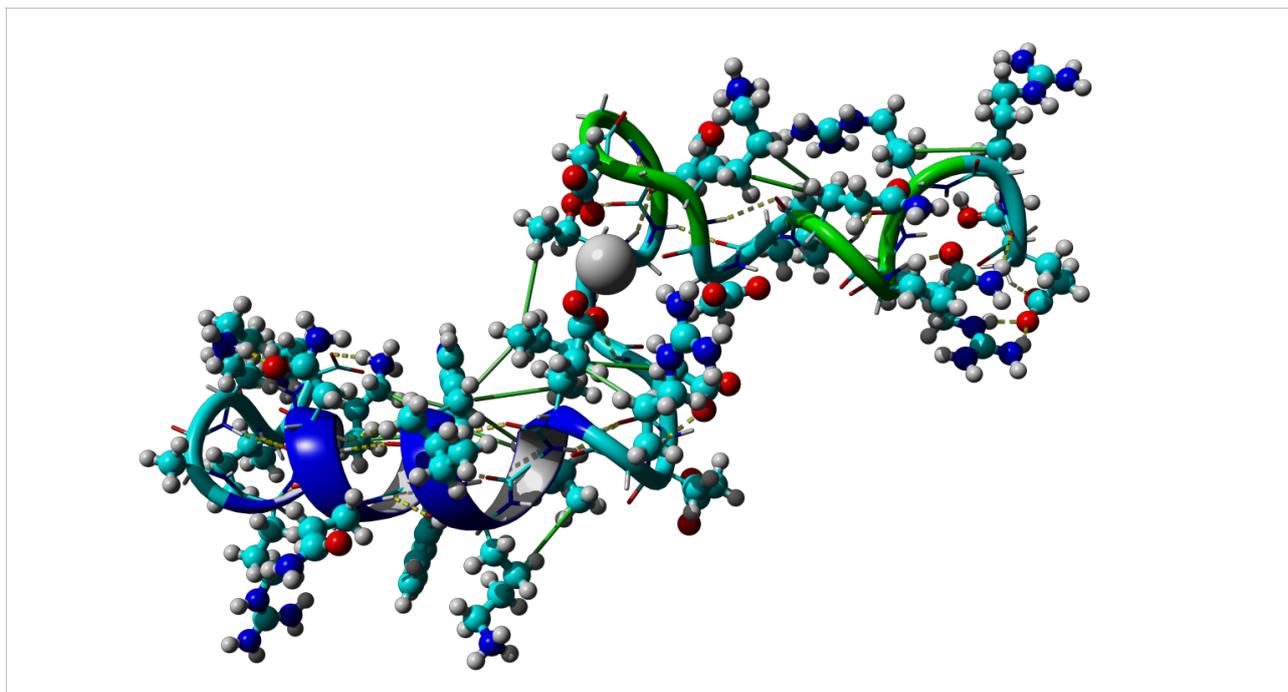
Some time was then taken to really understand the way these restraints work, and it was discovered that the parameters  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 1.5 \text{ \AA}$ ,  $d_{\text{plus}} = 3.5 \text{ \AA}$  would produce a potential well spanning from  $1 \text{ \AA}$  to  $6 \text{ \AA}$  of distance between the atoms. Previously, it was thought that the potential well would span from  $1.5 \text{ \AA}$  to  $3.5 \text{ \AA}$  with these parameters. From this point on, all distance restraints between ions and oxygen atoms were configured so that  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 0.5 \text{ \AA}$  and  $d_{\text{plus}} = 0.5 \text{ \AA}$ , creating a potential well spanning from  $2 \text{ \AA}$  to  $3 \text{ \AA}$  of distance between the restrained atoms.

Simulations from this point on are named according to the scheme described earlier, in all cases  $T_0 = 315\text{K}$  and  $T_1 = 298\text{K}$ .

### **Eu3+ s1**

This simulation was run with temperatures  $T_0 = 315\text{K}$  and  $T_1 = 298\text{K}$ . The binding ion was  $\text{Eu}^{3+}$ , and six restraining potentials were used to help pull oxygen atoms close to it. The restraints were created so that they would try to keep their respective atoms between 2 and 3  $\text{\AA}$  apart from each other. The overall structure of the protein resulting from this simulation is shown in figure 21.

The smaller alpha helix seems to have unwound somewhat, but there is still a curly shape. The short alpha helix also turned to point away from the longer one, similarly to the previously shown case of  $\text{Eu}^{3+}$  binding using the erroneously wide restraints.



**Figure 21:** Result of the Eu<sup>3+</sup> simulation with unscaled restraints. The overall shape of the protein has changed quite a lot, the termini have separated a lot. One of the helices is losing its shape. Some restraints were satisfied, but not all. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

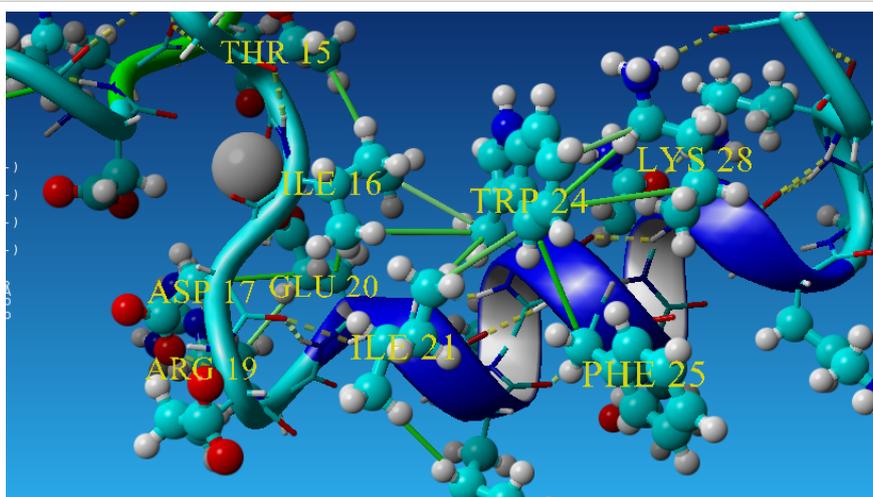
The energies determined from these results are shown in table 2. Restraint energy is relatively large, meaning that the system is unstable in this configuration.

Eu<sup>3+</sup> s1 P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	191.538 kJ/mol
<b>P3W</b>	64.495 kJ/mol
<b>Energy from binding to water</b>	127.043 kJ/mol
<b>Restraint energy</b>	61.722 kJ/mol

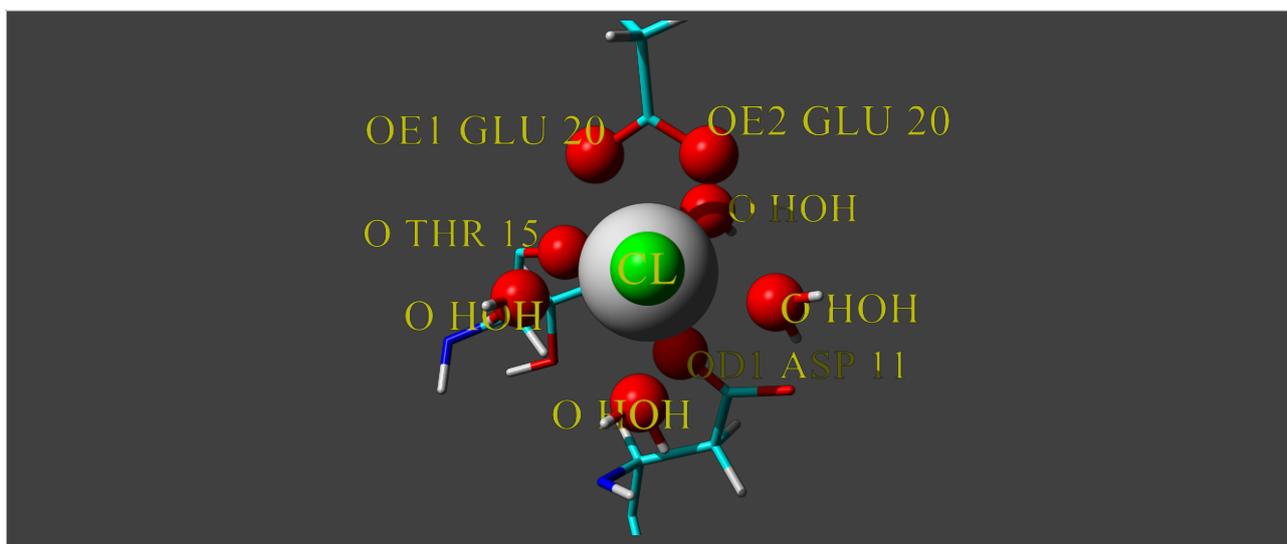
**Table 2:** Energies for Eu<sup>3+</sup>+s1. Restraint energy is quite high, meaning that the system is unstable in this configuration.

A closer look at the stabilizing hydrophobic pocket with Trp 24 is shown in figure 22. The hydrophobic pocket appears to include Phe 25, Lys28, Ile 21, Ile 16 and even Glu 20, Thr 15, Asp 17 and Arg 19. The pocket does appear to control the angle between the helices as has been noted before. (Joel T. Welch 2003)



**Figure 22:** Result of the Eu<sup>3+</sup> simulation with restraints unscaled. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. View of the hydrophobic pocket including the key stabilizing residue Trp 24.

Oxygens within 3 Å of the Eu<sup>3+</sup> are displayed in figure 23. There is a total of 8 oxygen atoms and one Cl<sup>-</sup> ion in that area. Four of the oxygens belong to water molecules, the rest are restrained oxygens from Asp 11, Thr 15, and Glu 20. Distances from the ion to restrained atoms and Oδ of Asp 17, which has been suggested as a possible ligand for the ion, are shown in table 3. Two of the six restraints were left unsatisfied, suggesting that the restraints might be too weak to be able to make the protein cross the kinds of potential barriers at the given temperatures and timescales. The backbone carbonyl oxygens of residues 10 and 13 were observed to be forming hydrogen bonds with residues 13 and 15 respectively, creating a potential barrier that is unlikely to be crossed without additional energy from a restraint during a 10 ns simulation.



**Figure 23:** Result of the Eu<sup>3+</sup> simulation with unscaled restraints. Ionically bound oxygen atoms, with less than 3 Å distance from Eu<sup>3+</sup>, are shown as balls and labeled. Labels are formatted as “<atom name> <residue name>”. There are 8 oxygen atoms and one chlorine atom from the solvent in that range and therefore the coordination number appears to be 9.

## Results and discussion

### Eu3+ P3W distances, Restraint scale 1

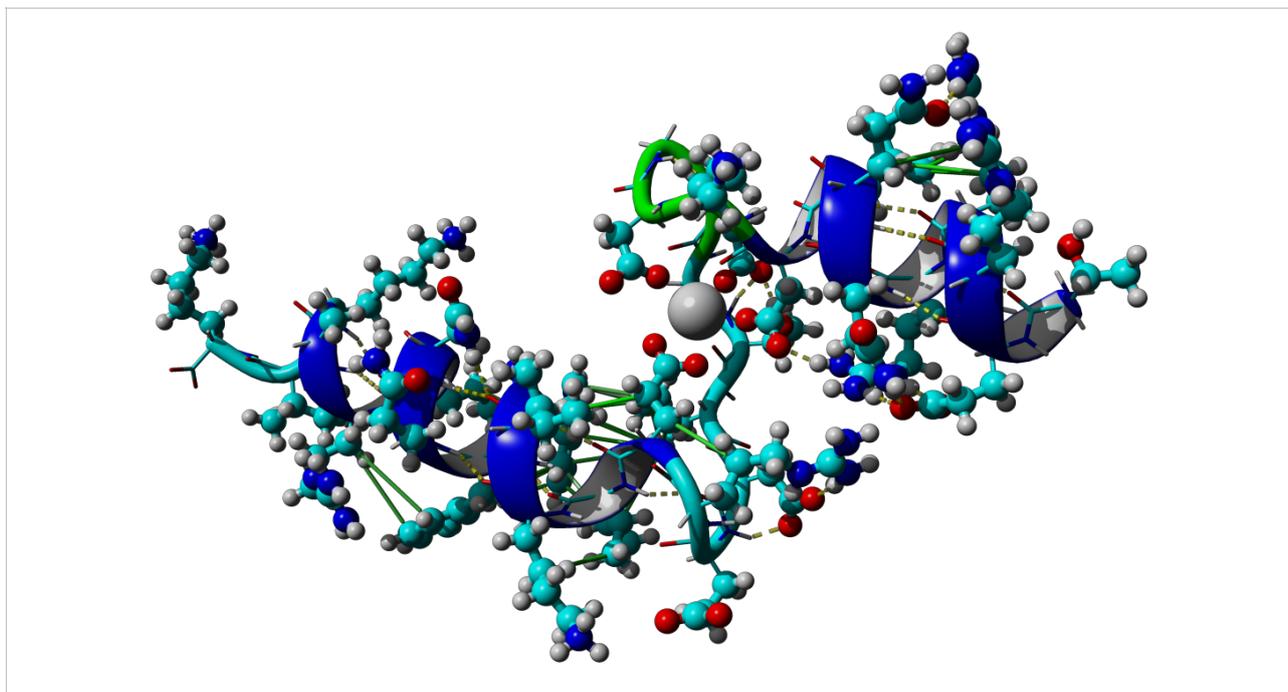
Residue (Atom name)	Distance to ion(Å)	Note
Asp 9 (OD)	4.869 Å	Too far
Asp 11 (OD)	2.340 Å	Correct
Asp 13 (OD)	9.507 Å	Too far
Thr 15 (O)	2.483 Å	Correct
Asp 17 (OD) (Unrestrained)	7.246 Å	Too far (unrestrained)
Glu 20 (OE1)	2.330 Å	Correct
Glu 20 (OE2)	2.455 Å	Correct

**Table 3:** Distances from key oxygens to the Eu3+ ion with restraint strength unscaled. Restraint  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 0.5 \text{ \AA}$ ,  $d_{\text{plus}} = 0.5 \text{ \AA}$ .

### Eu3+ s25

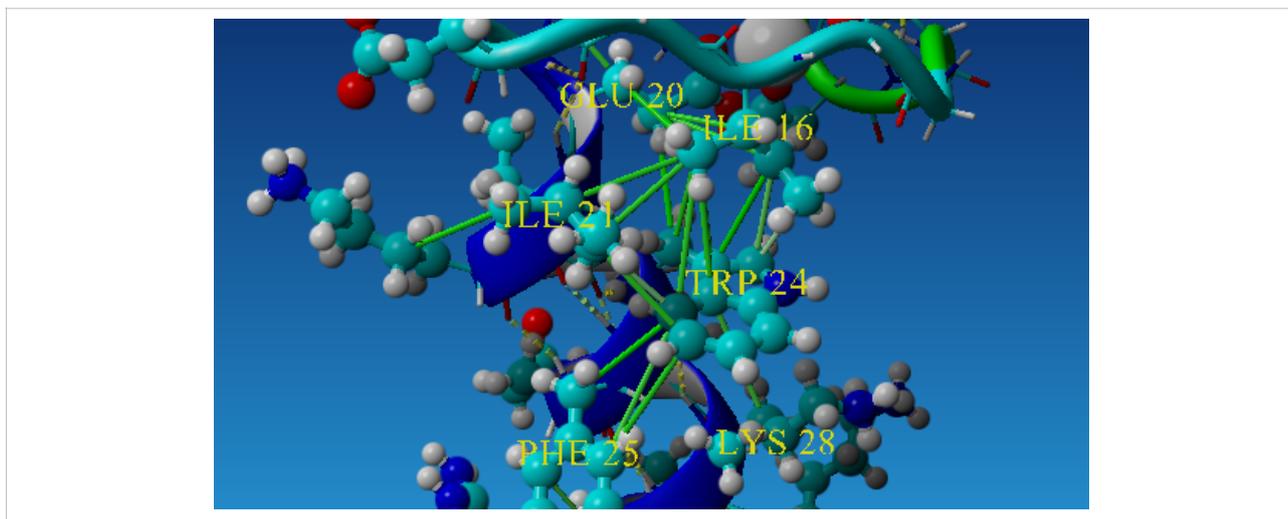
This simulation was run with temperatures  $T_0 = 315\text{K}$  and  $T_1 = 298\text{K}$ . The binding ion was Eu3+, and six restraining potentials were used to help pull known binding atoms close to it. The restraints were created so that they would try to keep their respective atoms between 2 and 3 Å apart from each other, and the potentials were scaled by a factor of 25. This was done hoping that the increased forces would be able to overcome any potential barriers that had prevented the system from reaching a configuration satisfying the restraints.

The overall structure at the end of the simulation is shown in figure 24. This time, both helices are left intact, the basic principle behind the configuration, however, remains the same as with weak restraints. This configuration does not agree with experiment in that the helices are not interacting at all, it has been determined that Trp 24 should interact with a hydrophobic pocket to stabilize the structure



**Figure 24:** Result of the Eu<sup>3+</sup> simulation with restraints scaled to 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

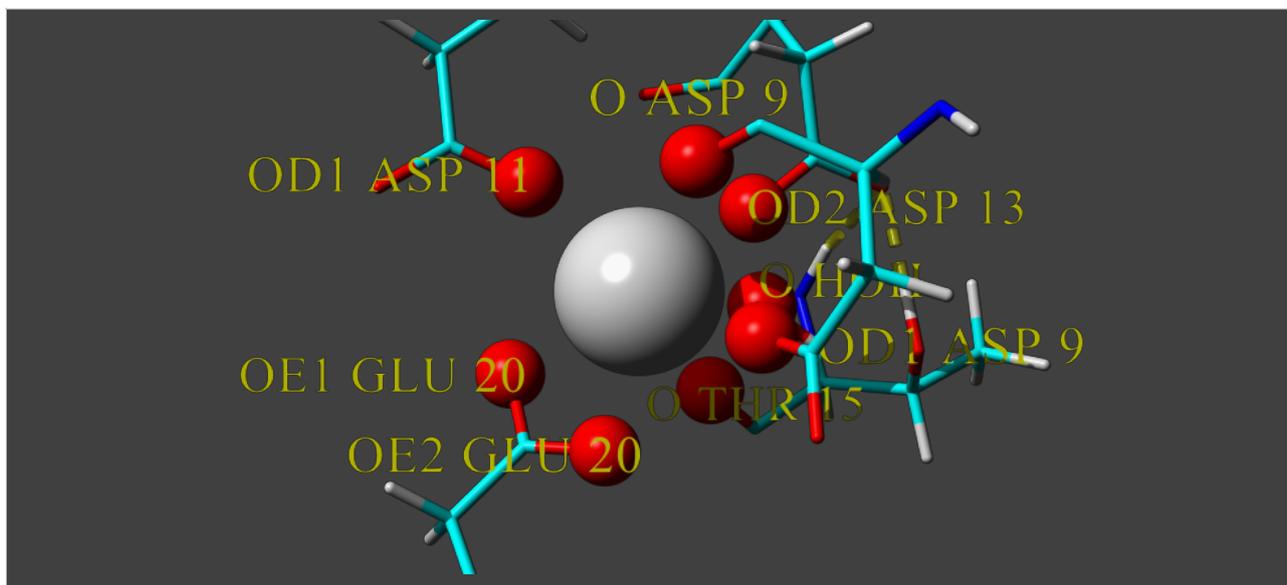
Experimental data has said that Trp 24 and Phe 25 are involved in a hydrophobic pocket that serves to stabilize the orientation of the helices. A close up view of the area around the residue is shown in figure 25. In this case, the hydrophobic pocket appears to include Phe 25, Lys28, Ile 21, Ile 16 and Glu 20. (Joel T. Welch 2003)



**Figure 25:** Result of the Eu<sup>3+</sup> simulation with restraints scaled to 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. View of the hydrophobic pocket including the key stabilizing residue Trp 24.

## Results and discussion

At the end of the simulation, it appears that the coordination number is 8, as there is one unrestrained backbone carbonyl oxygen and one water molecule that have bound in addition to the restrained ones, this is shown in figure 26. Comparing this to experimentally gathered knowledge of the binding reveals that the coordination number appears to be correct, however there should be two water molecules. The other water molecule may have been excluded by the ionic binding of the unrestrained backbone carbonyl oxygen from Asp 9.



**Figure 26:** Result of the Eu<sup>3+</sup> simulation with restraints scaled by 25, starting from the result of Eu<sup>3+</sup> s25. Ionically bound oxygen atoms, with less than 3 Å distance from Eu<sup>3+</sup>, are shown as balls and labeled. Labels are formatted as “<atom name> <residue name>” Coordination number is 8, the ion has bound 7 oxygens from the protein and one from water. The dotted yellow line indicates a hydrogen bond.

Restraint energy is zero, and this means that the restraints are satisfied perfectly without input from the restraints. This is shown in table 4. Since P3W was designed to be able to bind this ion, this result is not too surprising, however it does give some additional confidence that the simulation methods used are viable.

Eu<sup>3+</sup> s25 P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	234.567 kJ/mol
<b>P3W</b>	94.029 kJ/mol
<b>Energy from binding to water</b>	140.538 kJ/mol
<b>Restraint energy</b>	0.000 kJ/mol

**Table 4:** Energies for Eu<sup>3+</sup>+s25. Restraint energy is zero, meaning that the system is stable.

## Results and discussion

In the case of strengthened restraints, all distances between the restrained atoms are in the area where there is no artificial force acting upon them due to the restraints. This can be seen in table 5. This suggests that strong restraints should be used to overcome energy barriers and to reach proper binding configurations.

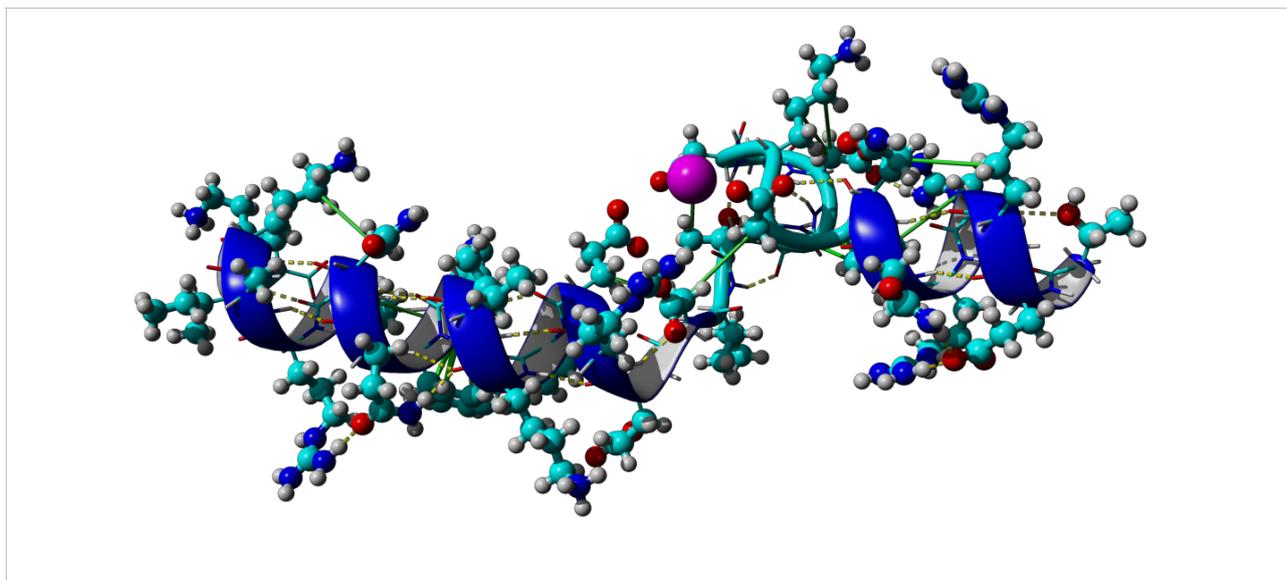
Eu3+ P3W distances, Restraint scale 25

Residue (Atom name)	Distance to ion(Å)	Note
Asp 9 (OD)	2.267 Å	Correct distance
Asp 9 (O) (unrestrained)	2.564 Å	Correct distance (unrestrained)
Asp 11 (OD)	2.299 Å	Correct distance
Asp 13 (OD)	2.312 Å	Correct distance
Thr 15 (O)	2.475 Å	Correct distance
Asp 17 (OD) (Unrestrained)	9.564 Å	Too far (unrestrained)
Glu 20 (OE1)	2.348 Å	Correct distance
Glu 20 (OE2)	2.445 Å	Correct distance

**Table 5:** Distances from key oxygens to the Eu3+ ion with restraint strength scaled by 25 times. Restraint  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 0.5 \text{ \AA}$ ,  $d_{\text{plus}} = 0.5 \text{ \AA}$ . There is one oxygen, the backbone carbonyl oxygen of Asp 9, that was not restrained and has bound spontaneously. Altogether, the ion has bound 7 oxygens from the protein.

### Fe3+ s1

This simulation was run with temperatures  $T_0 = 315\text{K}$  and  $T_1 = 298\text{K}$ . The binding ion was Fe3+, and six restraining potentials were used to help pull oxygen atoms close to it. The restraints were created so that they would try to keep their respective atoms between 2 and 3 Å apart from each other. The overall structure of the protein resulting from this simulation, which is quite similar to what was obtained from simulations with Eu3+, is shown in figure 27.



**Figure 27:** Result of the Fe<sup>3+</sup> simulation with unscaled restraints. The overall shape resembles that obtained from the Eu<sup>3+</sup> simulations. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

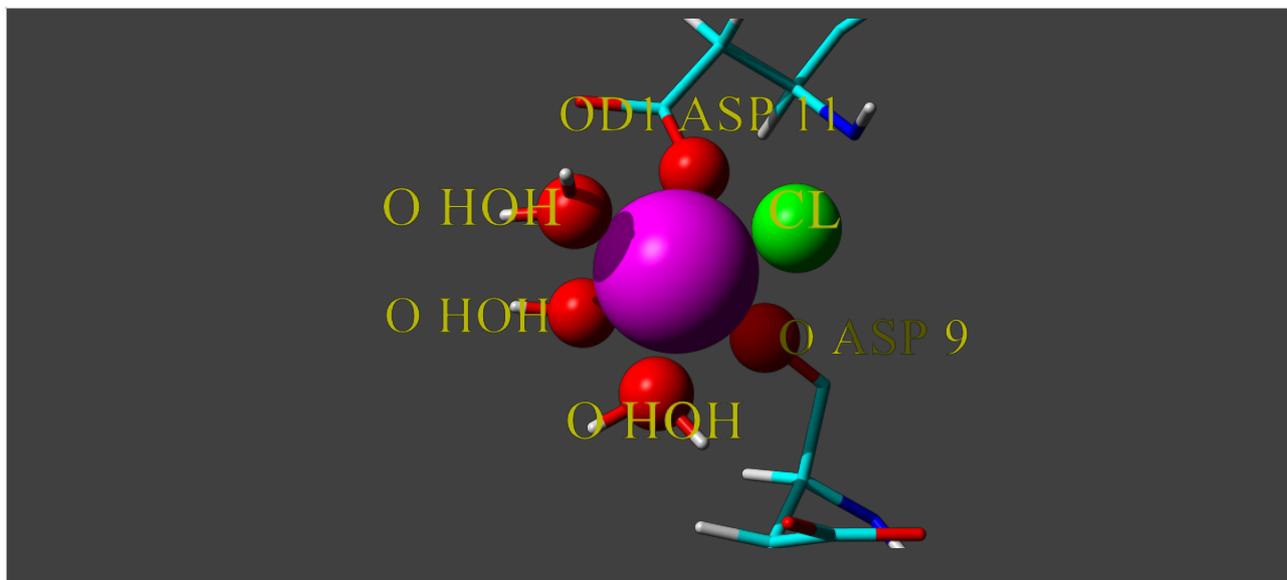
The energies calculated based on this result are shown in table 6. The restraint energy is again quite high, meaning that the system would quickly move into a different configuration without the presence of the restraints, that it is unstable without the restraints.

Fe<sup>3+</sup> s1 P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	240.863 kJ/mol
<b>P3W</b>	87.606 kJ/mol
<b>Energy from binding to water</b>	153.257 kJ/mol
<b>Restraint energy</b>	68.868 kJ/mol

**Table 6:** Energies for Fe<sup>3+</sup> s1. Restraint energy is quite high, meaning that the system is unstable in the given configuration without restraints.

The close-up view showing oxygens within 3 Å of the ion is shown in figure 28. The protein has only contributed two oxygen atoms as ligands for the ion and only one of those was restrained. This, again, suggests that the unscaled restraint potentials may have not been strong enough to cross the potential barriers on the way to the bound configuration. In addition to the two oxygens belonging to the protein, three water molecules and one Cl<sup>-</sup> ion have attached to the Fe<sup>3+</sup>.



**Figure 28:** Result of the Fe<sup>3+</sup> simulation with unscaled restraints. Close-up of the binding site shows only water molecules and residues in the immediate vicinity of the ion. Only two oxygens from the protein have formed ionic bonds in addition to three water molecules and one Cl<sup>-</sup> ion.

The distances to key oxygens of restrained residues and Asp 17 are shown in table 7. Only one of the oxygens is within 3 Å of the ion. It is actually located closer than 2 Å from the ion, meaning that the restraints have actually had to start pushing it away from the ion.

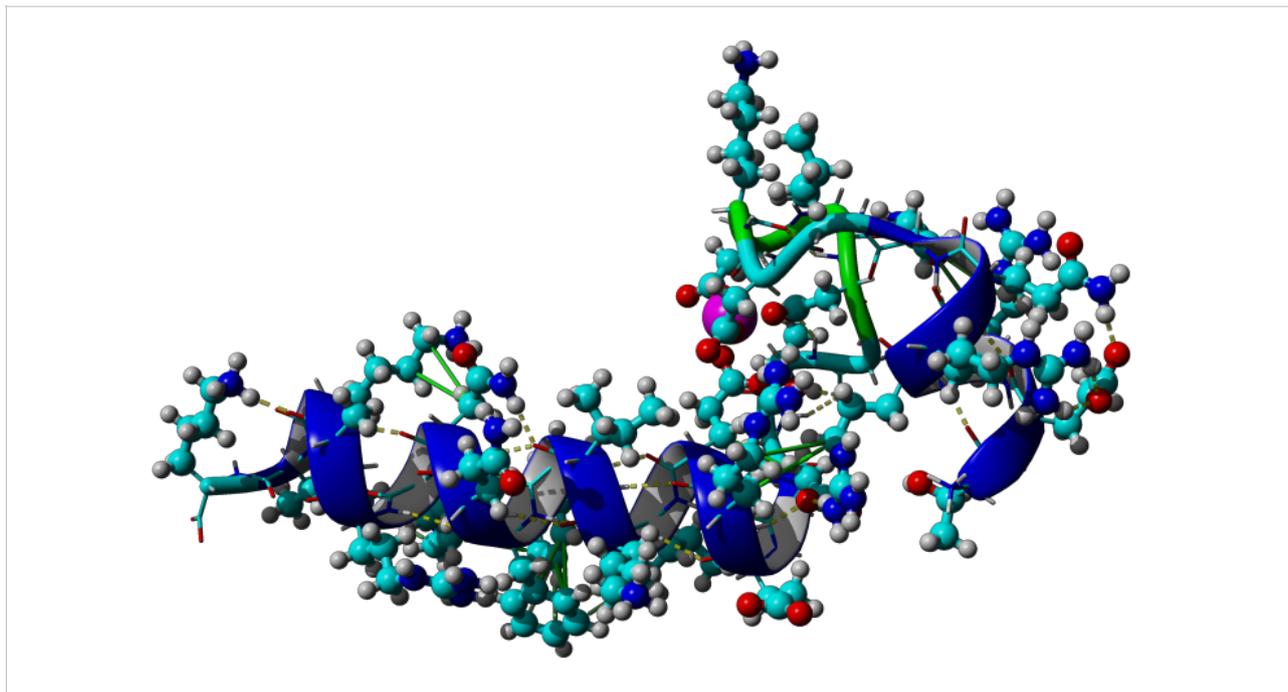
Fe<sup>3+</sup> P3W distances, Restraint scale 1

Residue (Atom name)	Distance to ion(Å)	Note
Asp 9 (OD)	4.092 Å	Too far
Asp 11 (OD)	1.873 Å	Too close
Asp 13 (OD)	9.387 Å	Too far
Thr 15 (O)	4.126 Å	Too far
Asp 17 (OD) (Unrestrained)	6.358 Å	Too far (unrestrained)
Glu 20 (OE1)	3.812 Å	Too far
Glu 20 (OE2)	4.287 Å	Too far

**Table 7:** Distances from key oxygens to the Fe<sup>3+</sup> ion with restraint strength unscaled. Restraint  $d = 2.5$  Å,  $d_{\text{minus}} = 0.5$  Å,  $d_{\text{plus}} = 0.5$  Å.

### Fe<sup>3+</sup> s25

Next, the same ion, Fe<sup>3+</sup>, was simulated using strong restraints where the force was scaled by a factor of 25. The simulation was run with temperatures T<sub>0</sub> = 315K and T<sub>1</sub> = 298K. Six restraining potentials were used to help pull known binding atoms close to the ion. The restraints were created so that they would try to keep their respective atoms between 2 and 3 Å apart from each other.



**Figure 29:** Result of the Fe<sup>3+</sup> simulation with restraints scaled by 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

The overall structure of the result can be seen in figure 29. Compared to the Eu<sup>3+</sup> result, the F-helix has turned so that its terminus is now closer to the beginning of the E-helix. This appears to be caused by the fact that all bound atoms are positioned about 0.5 Å closer to the ion than in the case of Eu<sup>3+</sup>, causing the backbone to constrict more. This is also the probable reason why one of the O $\delta$  atoms of Glu 20 did not reach within 3 Å of the ion, because the side chain could not reach the ion with both atoms, as might be concluded from table 8 and figure 30. Another explanation for this fact could be crowding around the area where the oxygen atom would have entered, and thus the atom may have simply been pushed out.

There appears to be one water molecule within 3 Å of the ion, which might support the first explanation, however the water molecule is positioned in a way that it is not clear that its removal would help the excluded oxygen get close to the ion.

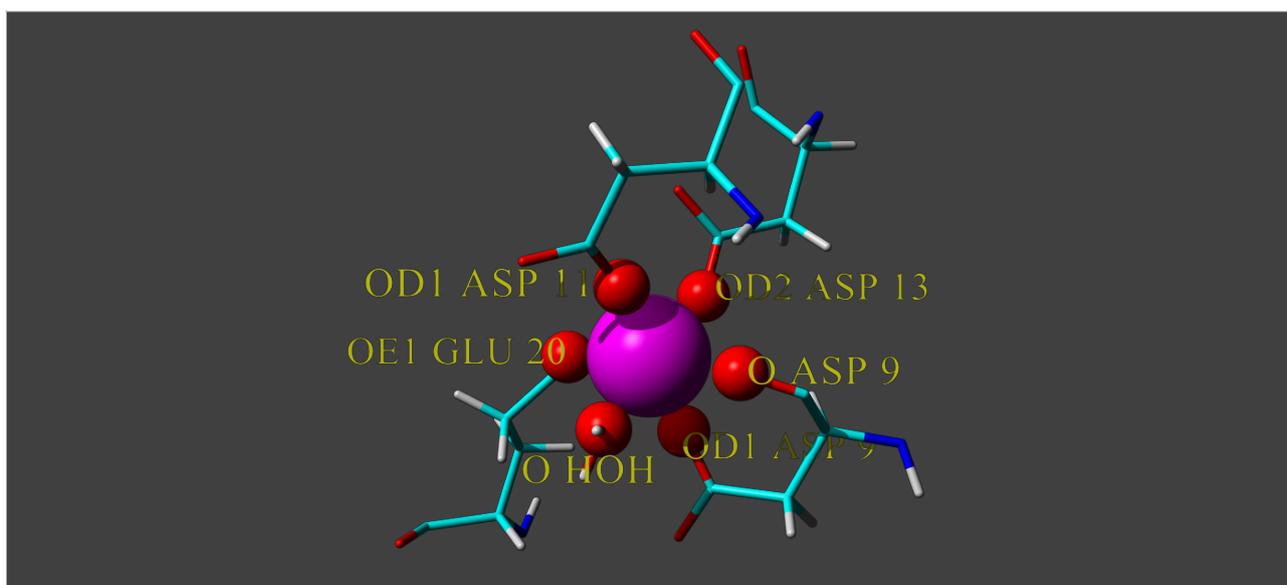
The exclusion of the oxygen is likely the primary reason why the restraint energy is as high as it is, as seen in table 9.

## Results and discussion

### Fe3+ P3W distances, Restraint scale 25

Residue (Atom name)	Distance to ion(Å)	Note
Asp 9 (OD)	1.893 Å	Too close
Asp 11 (OD)	1.875 Å	Too close
Asp 13 (OD)	1.870 Å	Too close
Thr 15 (O)	4.804 Å	Too far
Asp 17 (OD) (Unrestrained)	9.903 Å	Too far (unrestrained)
Glu 20 (OE1)	1.868 Å	Too close
Glu 20 (OE2)	3.174 Å	Too far (maybe pushed out)

**Table 8:** Distances from key oxygens to the Fe3+ ion with restraint strength scaled by 25 times. Restraint  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 0.5 \text{ \AA}$ ,  $d_{\text{plus}} = 0.5 \text{ \AA}$ .



**Figure 30:** Result of the Fe3+ simulation with restraints scaled by 25. Close up of the binding site showing only water molecules and residues in the immediate vicinity of the ion. There were a total of four water molecules with their oxygen atoms in the area, however three of those were separated from the ion by amino acid atoms and are not shown.

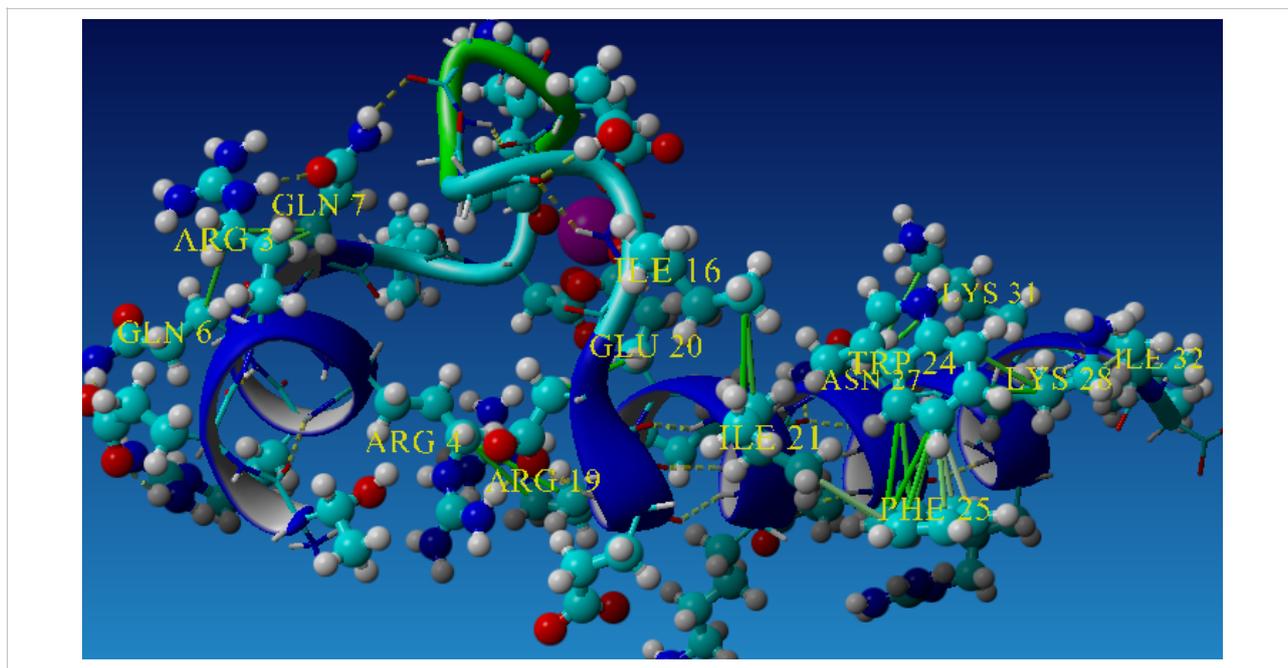
## Results and discussion

### Fe<sup>3+</sup> s25 P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	145.808 kJ/mol
<b>P3W</b>	47.680 kJ/mol
<b>Energy from binding to water</b>	98.128 kJ/mol
<b>Restraint energy</b>	11.295 kJ/mol

**Table 9:** Energies for Fe<sup>3+</sup> s1. Restraint energy is not zero, but considerably lower than with weak restraints, meaning that the system has become more stable.

Hydrophobic interactions and hydrogen bonding is shown in figure 31. Trp 24 and Phe 25 are strongly attracted to each other, Phe 25 is pulling at Ile 21, which is pulling in Ile 16. There is one hydrogen bond between Arg 4 and Arg 19 in addition to some hydrophobic attraction. This appears to be the key interaction orienting the helices. The E-helix is the same length as it was in the homology model. Other interactions affecting the orientation of the F-helix are a chain of hydrogen bonds and hydrophobic attraction running from Gly 12 through Gln 7 to Arg 3. There is also a number of hydrogen bonds in the loop that are helping to keep the ligating residues close to the ion.



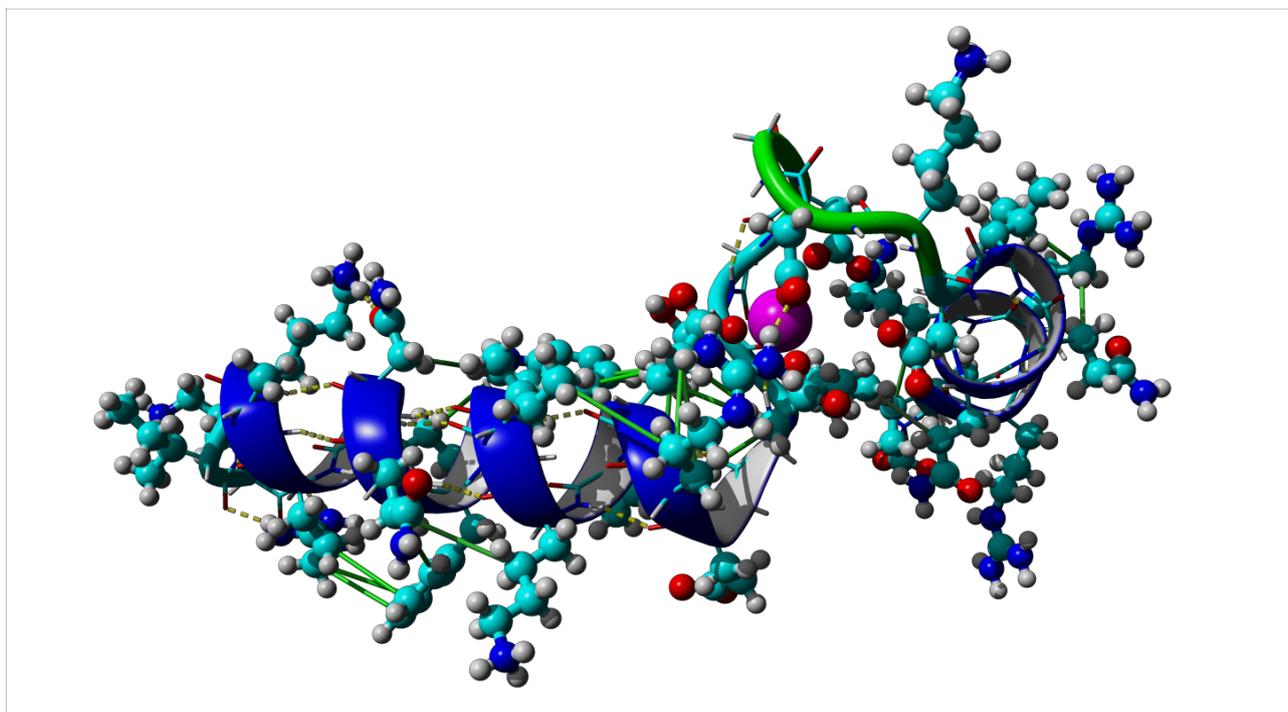
**Figure 30:** Result of the Fe<sup>3+</sup> simulation with restraint strength scaled by 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. In the largest hydrophobic pocket, the key stabilizing residue Trp 24 appears to be strongly interacting with Phe 25 which is pulling in Ile 21, which in turn is pulling Ile 16. There is both hydrophobic and hydrogen bonding interaction between Arg 4 and arg 19.

## Results and discussion

This result again supports the idea that weak restraints are not sufficient to pull the system to the correct conformation, and that restraints that have their force multiplied by 25 are at least better if not definitely sufficient to achieve that goal. Further simulations were all performed with just the strong restraints.

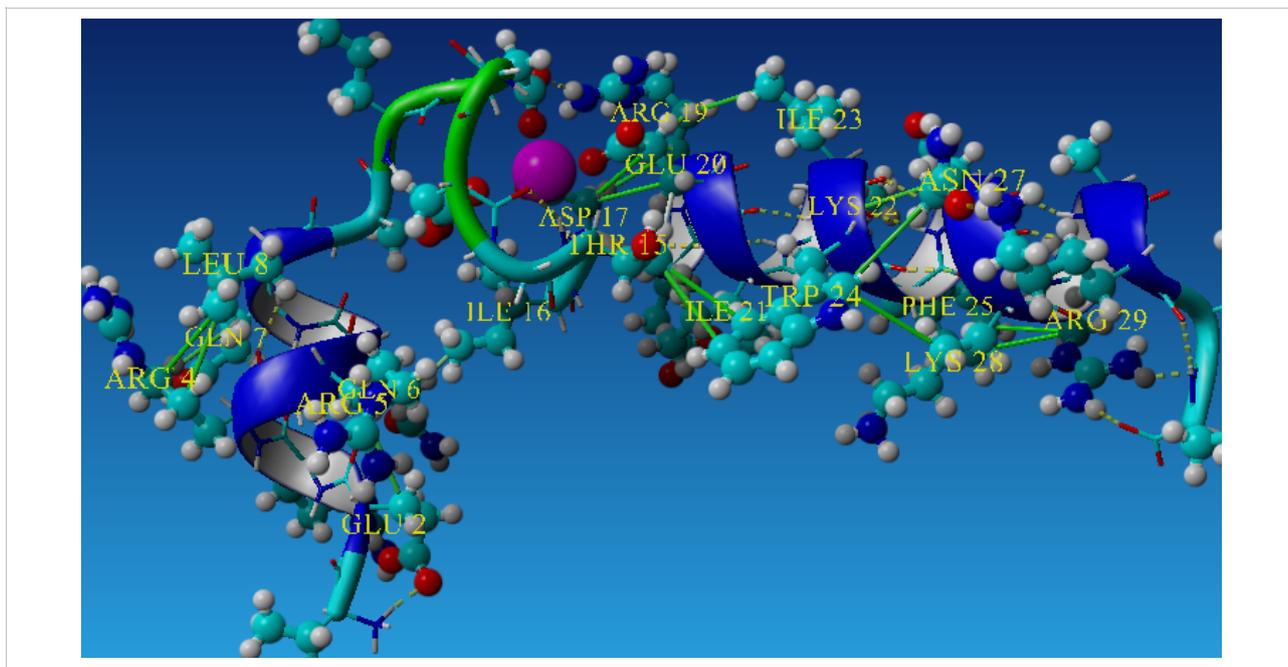
### Fe<sup>2+</sup> s25

In order to get some data about the importance of charge on the binding, Fe<sup>2+</sup> was also simulated. This simulation was run with temperatures T<sub>0</sub> = 315K and T<sub>1</sub> = 298K. The binding ion was Fe<sup>3+</sup>, and six restraining potentials were used to help pull oxygen atoms close to it. The restraints were created so that they would try to keep their respective atoms between 2 and 3 Å apart from each other. The large scale structure of the result is shown in figure 31. The F-helix appears to have turned to point away from the camera when looking from such an angle where Glu 20 is above the E-helix axis. This structure implies right away that Asp 9 has not bound to the ion.

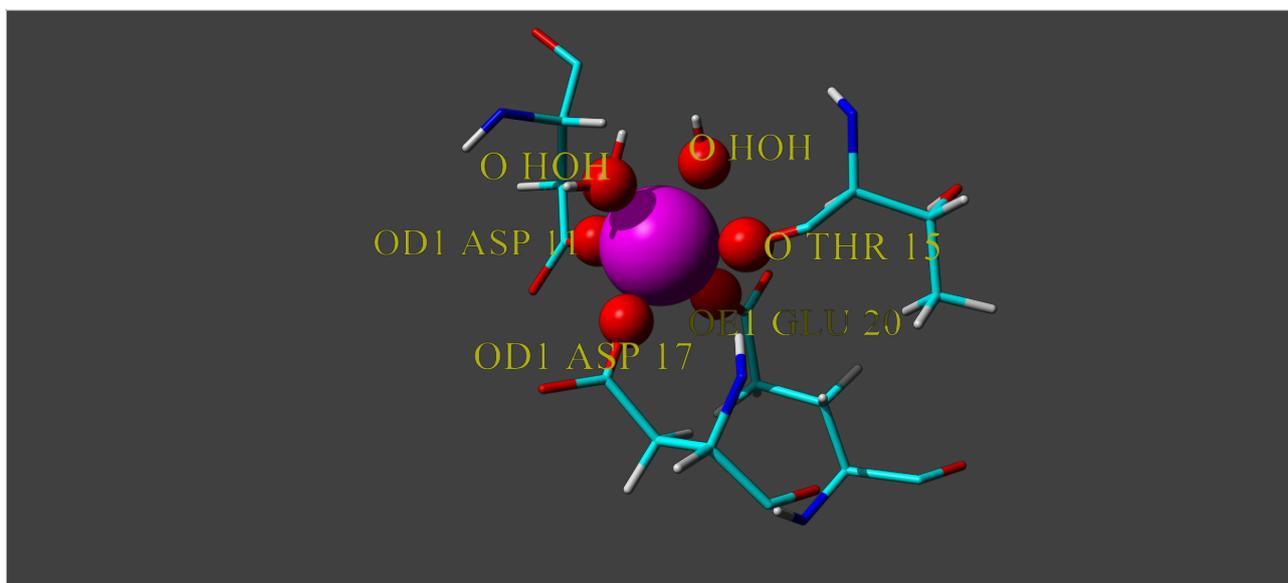


**Figure 31:** Result of the Fe<sup>2+</sup> simulation with restraints scaled by 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

Hydrophobic and hydrogen bond interactions are shown in figure 32. Interactions of the E-helix appear to have split into chains. Phe 25 is interacting along the axis of the helix with Arg 29. Trp 24 is interacting with Thr 15, Asn 27 and Lys 28, which is also roughly along the helix axis. Arg 19, Glu 20 and Asp 17 have formed a triangle of hydrophobic attraction. There is a weak attraction between Ile 16 and Gln 6. Arg 19 and Asp 11 have formed a hydrogen bond, which is also helping to stabilize the loop.



**Figure 32:** Result of the Fe<sup>2+</sup> simulation with restraint strength scaled by 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. The hydrophobic pocket including the key stabilizing residue Trp 24 appears to have split into slices as Phe 25 has its own pocket under the helix and Trp 24 has moved to interact with Thr 15. There is a weak interaction between Ile 16 and Gln 6.



**Figure 33:** Result of the Fe<sup>2+</sup> simulation with restraints scaled by 25. Close-up of the binding site shows only water molecules and residues in the immediate vicinity of the ion. It looks like two water molecules are located in the region, bringing the coordination number up to six.

A close inspection of the vicinity of the ion is shown in figure 33. The figure shows oxygen atoms within 3 Å of the ion. Four oxygens from the protein are within that range, three of them constrained with restraints. The unrestrained oxygen belongs to Asp 17, which was suggested as a possible ligand by Joel T. Welch (2003). In addition, two water molecules can be seen to be

## Results and discussion

interacting with the ion from the general direction of Asp 13, which would be expected interact somewhere between residues 11 and 15.

The binding energies calculated from this result are shown in table 10. The restraint energy is close to what was calculated for Fe<sup>3+</sup>, even though the number of unsatisfied restraints is higher, suggesting that these restraints are on average closer to being satisfied than in the case of Fe<sup>3+</sup>.

Further investigation after all simulations of this report were already done revealed that performing further/repeated energy minimization experiments could make the binding energy turn from negative to positive, suggesting that the energy minimization method used may have been inadequate, however, conclusions that are drawn based on geometric considerations and large scale topology should still hold some value.

Fe<sup>2+</sup> s25 P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	-59.506 kJ/mol
<b>P3W</b>	-64.557 kJ/mol
<b>Energy from binding to water</b>	5.051 kJ/mol
<b>Restraint energy</b>	9.792 kJ/mol

**Table 10:** Energies for Fe<sup>2+</sup> s25. Restraint energy is not zero, but still relatively low. For the first time, binding energy was found to be negative, suggesting insufficient energy minimization.

The distances between the restrained atoms along with one O $\delta$  from Asp 17 are shown in table 11. Three restrained atoms are close enough to be considered to have bound, the atom from Asp 17 has also made an ionic bond with Fe<sup>2+</sup>.

Fe<sup>2+</sup> P3W distances, Restraint scale 25

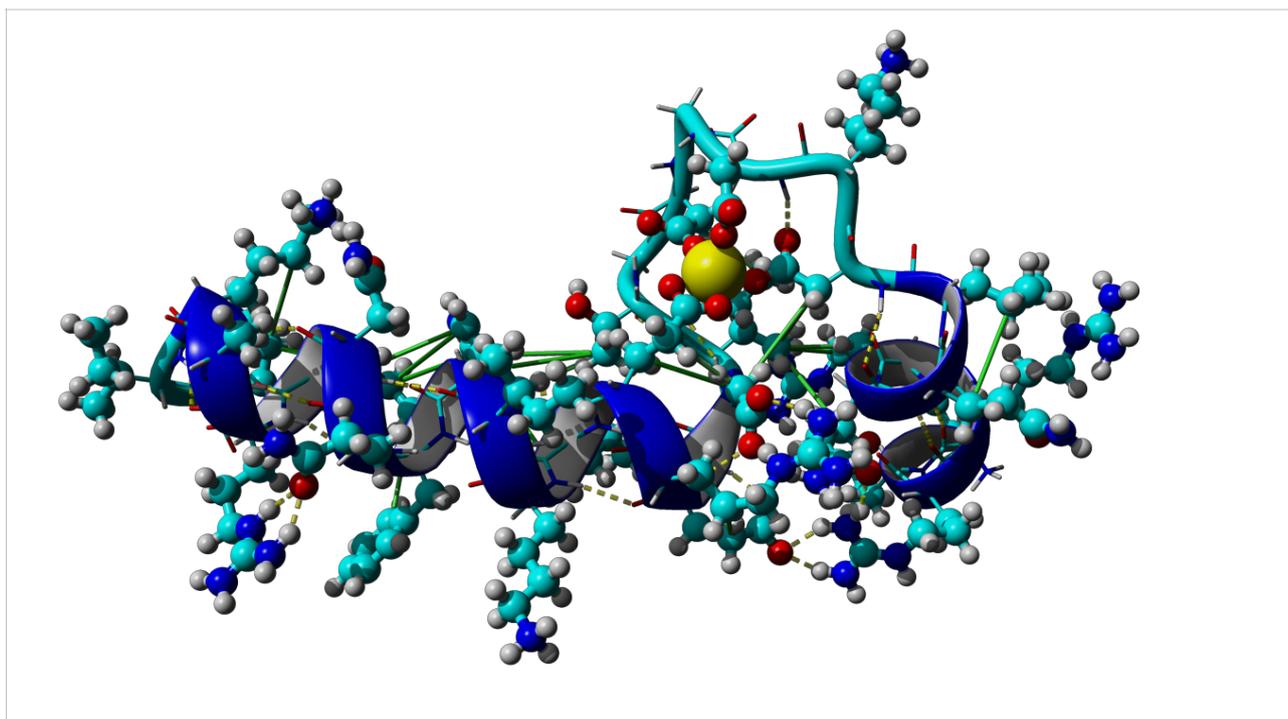
Residue (Atom name)	Distance to ion(Å)	Note
<b>Asp 9 (OD)</b>	4.168 Å	Too far
<b>Asp 11 (OD)</b>	1.806 Å	Too close
<b>Asp 13 (OD)</b>	3.656 Å	Too far
<b>Thr 15 (O)</b>	2.157 Å	Correct
<b>Asp 17 (OD) (Unrestrained)</b>	1.807 Å	Too close <b>(unrestrained)</b>
<b>Glu 20 (OE1)</b>	1.811 Å	Too close
<b>Glu 20 (OE2)</b>	3.708 Å	Too far (Pushed out)

**Table 11:** Distances from key oxygens to the Fe<sup>2+</sup> ion with restraint strength scaled by 25 times. Restraint  $d = 2.5 \text{ \AA}$ ,  $d_{\text{minus}} = 0.5 \text{ \AA}$ ,  $d_{\text{plus}} = 0.5 \text{ \AA}$ .

## Results and discussion

### Mg<sup>2+</sup> s25

Another type of ion that was tried is Mg<sup>2+</sup>. This element was chosen because its weight and size are even smaller than those of Ca<sup>2+</sup>. This simulation was run with temperatures  $T_0 = 315\text{K}$ ,  $T_1 = 298\text{K}$ , and six restraining potentials were used to help pull oxygen atoms close to it. The restraints were created so that they would try to keep their respective atoms between 2 and 3 Å apart from each other. The large scale structure of the result is shown in figure 34. The orientation of the helices appears to be almost the same as it was in the result of Fe<sup>2+</sup> s25.



**Figure 34:** Result of the Mg<sup>2+</sup> simulation with restraints scaled by 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

Calculated energies are shown in table 12. Binding energies are again negative, suggesting that the energy minimization may have been insufficient. Restraint energy is almost zero, suggesting that the system should be more or less stable in this configuration.

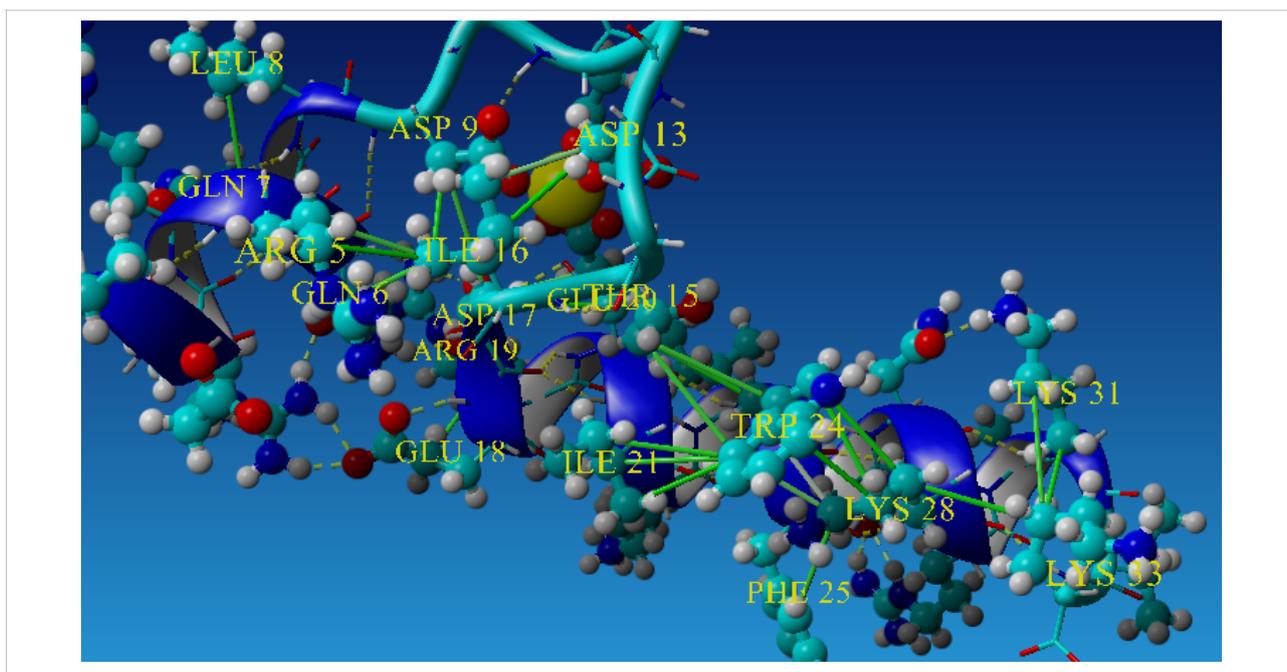
Mg<sup>2+</sup> s25 P3W binding energies

	Energy (kJ/mol)
P3W and water	-240.454 kJ/mol
P3W	-113.514 kJ/mol
Energy from binding to water	-126.940 kJ/mol
Restraint energy	0.686 kJ/mol

**Table 12:** Energies for Mg<sup>2+</sup> s25. Restraint energy is essentially zero. Binding energy was again found to be negative, suggesting insufficient energy minimization.

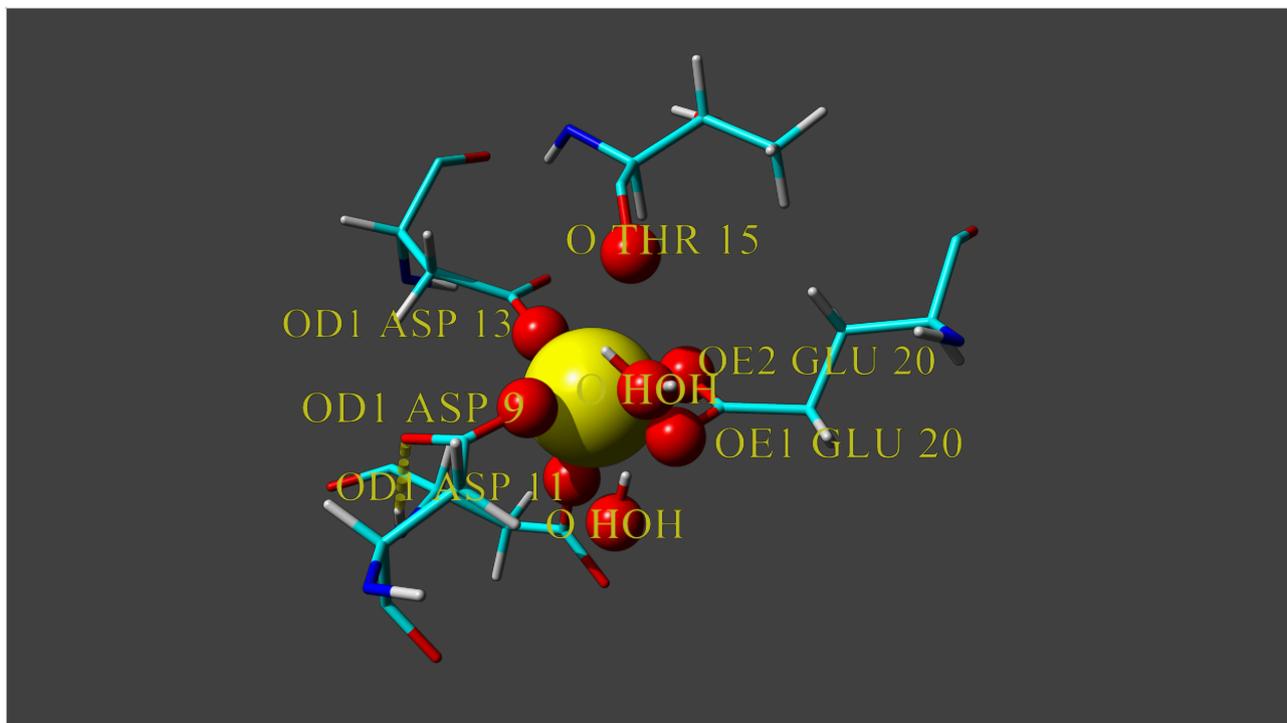
## Results and discussion

Hydrophobic interactions are shown in figure 35. The pocket involving Trp 24 appears to be strongly established, even though the E-helix is quite long. It is pulling Thr 15 and forcing the binding loop to move further to the left than it would otherwise be. This in turn enables Ile 16 to interact with Asp 9, Arg 5 and Gln 6, orienting the F-helix. In this case, there is also a chain of interactions from Asp 9 through Asp 17 to Glu 20 keeping the helices close. It is also important that there are two hydrogen bonds between Glu 18 and Arg 3 and one between Asp 17 and Gln 6 helping to keep the helices together. There is also a number of hydrogen bonds in the loop that are helping to keep the ligating residues close to the ion.



**Figure 35:** Result of the Mg<sup>2+</sup> simulation with restraint force scaled by 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. View of the hydrophobic pocket including the key stabilizing residue Trp 24, which does not seem to affect the loop area very much in this case. Hydrophobic interactions appear to be stabilizing the whole structure.

Oxygens within 3 Å of the ion are shown in figure 36. All restrained atoms have been pulled into this area. In addition, there are two water molecules in the region, one of which has its oxygen just 1.56 Å from the ion. The other water molecule has its oxygen 2.28 Å from the ion, but is pointing it away from the Mg<sup>2+</sup> and therefore should likely not be considered to be bound.



**Figure 36:** Result of the Mg<sup>2+</sup> simulation with restraints scaled by 25. Close-up of the binding site shows only water molecules and residues in the immediate vicinity of the ion. The dotted yellow line indicates a hydrogen bond.

The measured distances between all restrained oxygens and one oxygen from Asp 17 are shown in table 13.

Mg<sup>2+</sup> P3W distances, Restraint scale 25

Residue (Atom name)	Distance to ion(Å)	Note
Asp 9 (OD)	1.784 Å	Too close
Asp 11 (OD)	1.795 Å	Too close
Asp 13 (OD)	1.792 Å	Too close
Thr 15 (O)	2.781 Å	Correct
Asp 17 (OD) (Unrestrained)	5.772 Å	Too far (unrestrained)
Glu 20 (OE1)	1.870 Å	Too close
Glu 20 (OE2)	1.877 Å	Too close

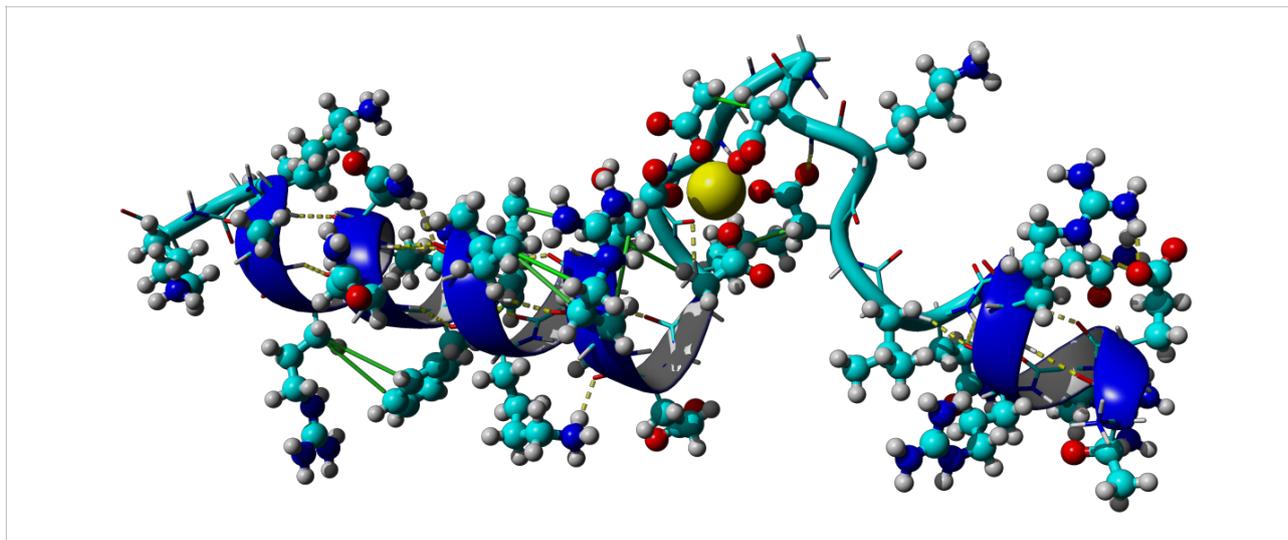
**Table 13:** Distances from key oxygens to the Mg<sup>2+</sup> ion with restraint strength scaled by 25 times. Restraint  $d = 2.5$  Å,  $d_{\text{minus}} = 0.5$  Å,  $d_{\text{plus}} = 0.5$  Å.

### Ca<sup>2+</sup> s25

The last type of ion that was tried is Ca<sup>2+</sup>. This simulation was again run with temperatures  $T_0 = 315\text{K}$ ,  $T_1 = 298\text{K}$ , and six restraining potentials were used to help pull oxygen atoms close to it. The restraints were created so that they would try to keep their respective atoms between 2 and 3 Å

## Results and discussion

apart from each other. The large scale structure of the result is shown in figure 37. This conformation appears to have the largest distance between the termini of all the ones encountered thus far, the protein seems to be stretched out.



**Figure 37:** Result of the Ca<sup>2+</sup> simulation with restraints scaled to 25. The protein seems to be stretched out. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

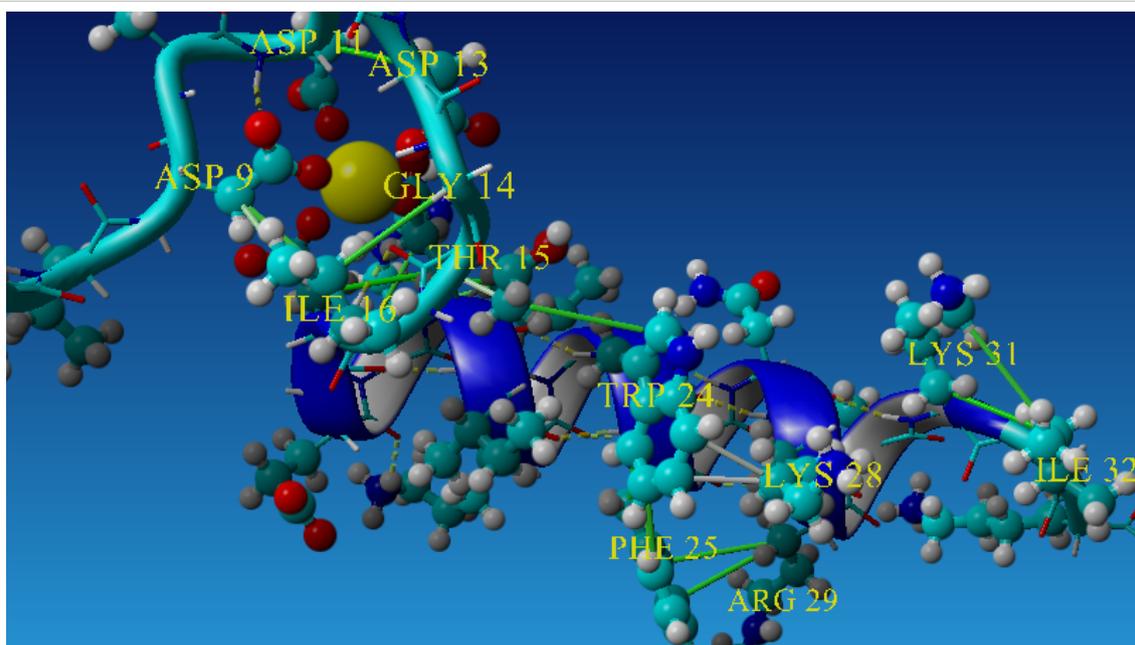
Calculated binding energies are shown in table 14. Some of the binding energies are once again negative. The restraint energy is quite low, indicating that the system would likely stay in a similar conformation if the restraints were to be removed.

Ca<sup>2+</sup> s25 P3W binding energies

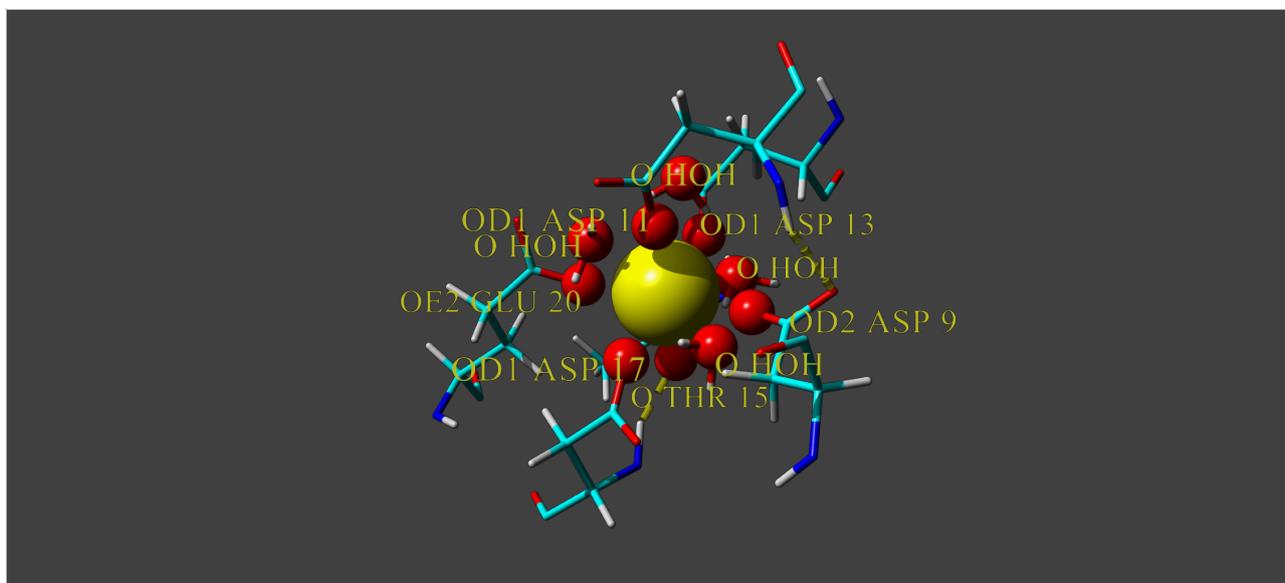
	Energy (kJ/mol)
<b>P3W and water</b>	-9.947 kJ/mol
<b>P3W</b>	79.129 kJ/mol
<b>Energy from binding to water</b>	-89.075 kJ/mol
<b>Restraint energy</b>	3.004 kJ/mol

**Table 14:** Energies for Ca<sup>2+</sup> s25. Restraint energy is close to zero. Binding energy was again found to be negative, suggesting insufficient energy minimization.

Hydrophobic interactions are shown in figure 38. The pocket around Trp 24 appears to have lost its strength due to an elongation of the E-helix. The F-helix had no detectable hydrophobic interactions. There is a system of hydrophobic interactions in the loop area, such as the ones involving Ile 16, most notably with Asp 9, however they do not appear to be the primary force governing the structure in this case.



**Figure 38:** Result of the Ca<sup>2+</sup> simulation with restraints strength scaled by 25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. The hydrophobic pocket including the key stabilizing residue Trp 24, does not seem to affect the loop area very much in this case. There is a series of hydrophobic interactions that were detected in the loop region, including between Ile 16 and Asp 9, however these are not very significant in this case. No hydrophobic interactions were detected in the F-helix.



**Figure 39:** Result of the Ca<sup>2+</sup> simulation with restraints scaled by 25. Ionically bound oxygen atoms, with less than 3 Å distance from Ca<sup>2+</sup>, are shown as balls and labeled. Labels are formatted as “<atom name> <residue name>”. Residue There are 10 oxygen atoms in the region, however one of them belongs to a water molecule that is actually pointing it away from the ion (top on image) and therefore the coordination number appears to be 9.

## Results and discussion

A closer view of oxygen atoms within 3 Å of the ion is shown in figure 39. There is a total of 10 oxygen atoms in the area, one of those is part of a water molecule that is pointing away from the ion, therefore the coordination number is 9. Five of the six restraints have been satisfied, in addition, an unrestrained oxygen atom from Asp 17 is bound, and one of the restrained oxygens of Glu 20 appears to have been pushed out due to the crowding.

In this case, as can be seen in table 15, the potential barriers have been crossed, and an energy minimum where all the restraints are satisfied has been reached. The fact that the restraints are unscaled during the second stage of the MD-analysis, and that they are still all satisfied, supports the idea that the simulations with unscaled restraints are hampered by potential barriers.

Ca<sup>2+</sup> P3W distances, Restraint scale 25

Residue (Atom name)	Distance to ion(Å)	Note
Asp 9 (OD)	2.162 Å	Correct
Asp 11 (OD)	2.170 Å	Correct
Asp 13 (OD)	2.185 Å	Correct
Thr 15 (O)	2.352 Å	Correct
Asp 17 (OD) (Unrestrained)	2.186 Å	Correct (unrestrained)
Glu 20 (OE1)	3.847 Å	Too far
Glu 20 (OE2)	2.147 Å	Correct

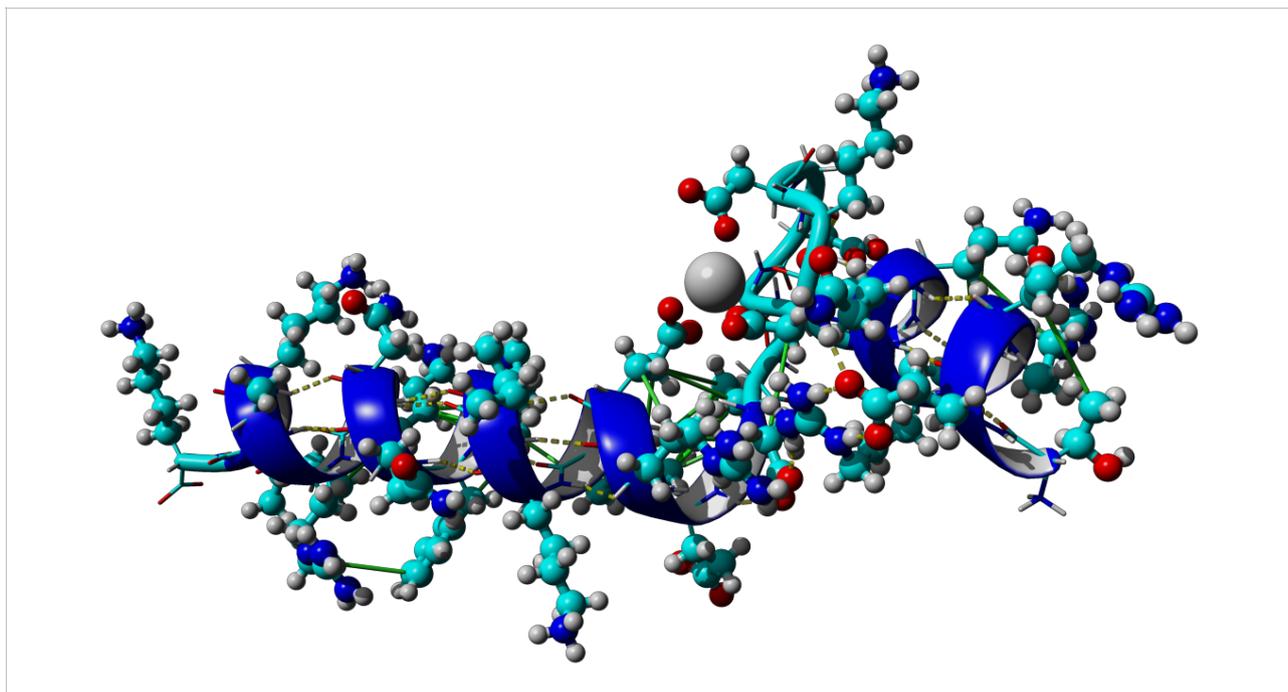
**Table 15:** Distances from key oxygens to the Ca<sup>2+</sup> ion with restraint strength scaled by 25 times. Restraint  $d = 2.5$  Å,  $d_{\text{minus}} = 0.5$  Å,  $d_{\text{plus}} = 0.5$  Å.

### Simulations to check the stability of the reached configurations

To study the stability of the resultant configurations, new simulations were run with all the same parameters as before, continuing from the results shown previously and used for the calculation of FRET efficiency, only this time the restraining potentials were removed.

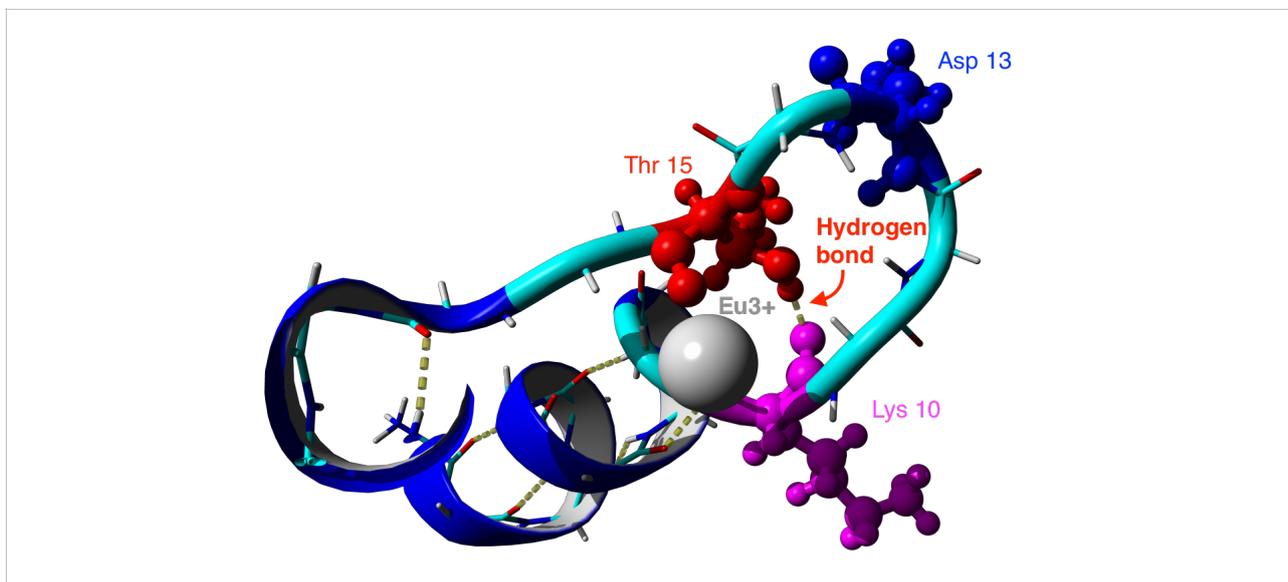
#### Eu<sup>3+</sup> (no restraints) continuing from Eu<sup>3+</sup> s25

Removing the restraints from the result of Eu<sup>3+</sup> s25 led to a configuration where Asp 13 is no longer bound to the ion, and the orientation of alpha helices is similar to the result of Fe<sup>3+</sup> s25, demonstrated in figure 40.



**Figure 40:** Result of the Eu<sup>3+</sup> simulation with restraints removed, starting from the result of Eu<sup>3+</sup> s25. Hydrophobic interactions shown as green lines, hydrogen bonds shown as yellow dotted lines.

Lys 10 appears to have formed a hydrogen bond with Thr 15, causing Asp 13 to move away from the ion, this is shown in figure 41.

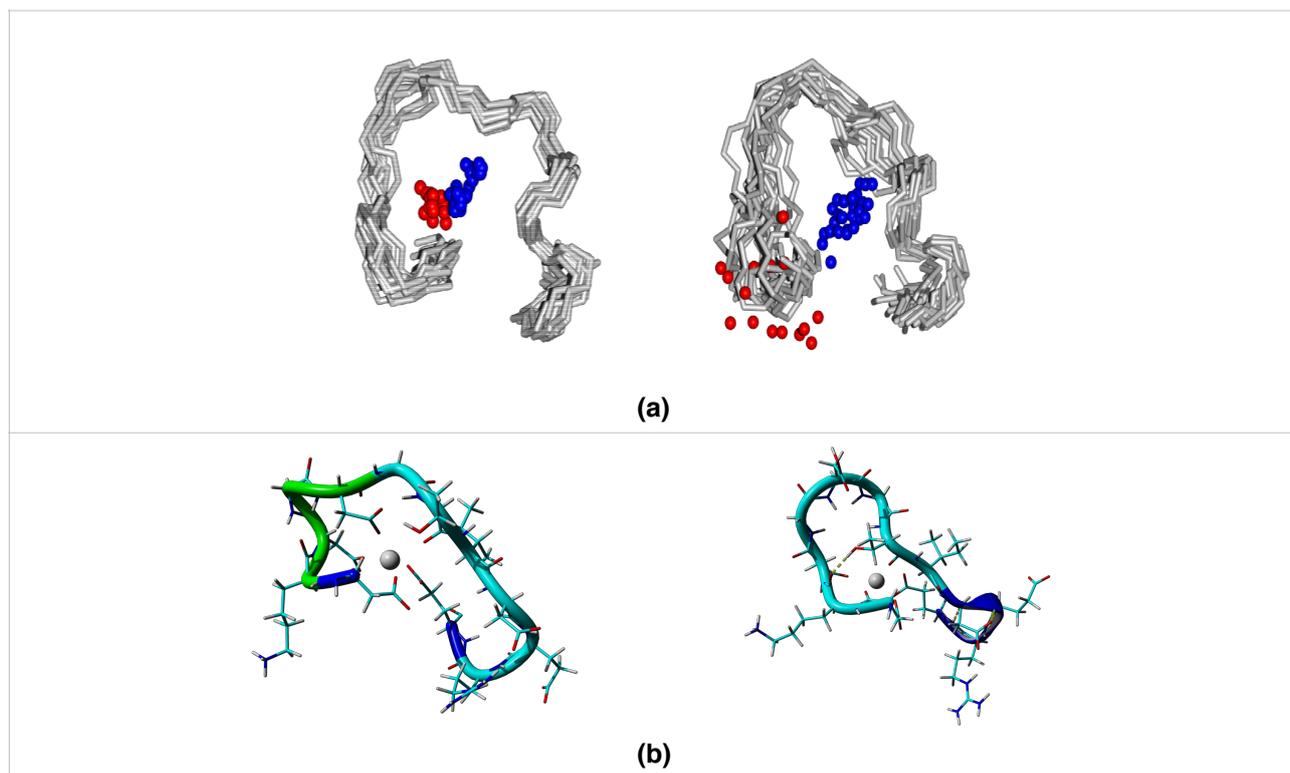


**Figure 41:** Result of the Eu<sup>3+</sup> simulation with restraints removed, starting from the result of Eu<sup>3+</sup> s25. Lys 10 and Thr 15 have created a hydrogen bond, keeping Asp 13 away from Eu<sup>3+</sup>. Dotted yellow lines indicate hydrogen bonds.

A comparison was made between the restrained and unrestrained results by Joel T. Welch et al. (2003) and this report. This is shown in figure 42. As noted in the source, the overall fold of the protein does not greatly depend on the presence of restraints. The results obtained in this report

## Results and discussion

show a somewhat flatter shape of the binding loop. In both cases, the result obtained without restraints appears to have the area of Asp 13 further away from the ion than when the restraints are present, although the effect seems more pronounced in the results of this report. This certainly does not mean that the simulations are correct, however it does suggest that that residue might be less firmly bound. The primary reason behind the loss of Asp 13 in the unrestrained case could be the fact that, as pointed out above, the E-helix was modeled as starting from residue number 17 instead of 21 and thus deformed the binding loop.



**Figure 42:** Solution structure of P3W with Eu<sup>3+</sup>. (a) Backbone with Eu<sup>3+</sup> with restraints of the left and without restraints on the right, taken from (Joel T. Welch 2003). (b) Residues 9 to 20 from Eu<sup>3+</sup> s25 on the left and its unrestrained continuation simulation on the right. The overall fold of the molecule is similar in all cases, however both molecules in (b) are somewhat more elongated than in (a). In both cases, the location of Asp 13 seems to have retreated a bit from the ion, although to a lesser extent in (a).

Calculated binding energies are shown in table 16. Compared to the restrained version, the total calculated binding energy has increased significantly, the binding energy without the presence of water molecules is about two thirds of what it was with restraints. The reduction of energy from binding to the protein is consistent with the fact that Asp 13 is no longer bound, although the magnitude of the change is surprisingly large. The large increase of binding energy with the water component is likely due to the Cl<sup>-</sup> ion, which has a negative charge, and therefore interacts very strongly with the Eu<sup>3+</sup>. The Cl<sup>-</sup> ion might be the reason why the binding energy with the protein has decreased to such a large extent due to the negatively charged ion, decreasing the electric field strength in the whole area.

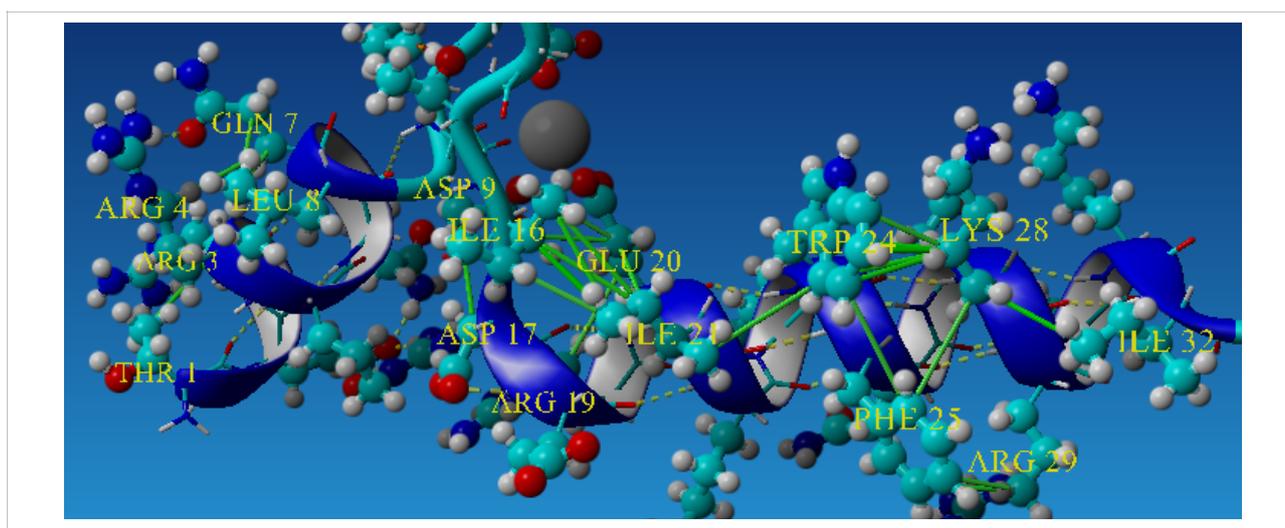
## Results and discussion

### Eu3+ P3W continuation binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	147.091 kJ/mol
<b>P3W</b>	66.893 kJ/mol
<b>Energy from binding to water</b>	80.198 kJ/mol

**Table 16:** Calculated binding energies.

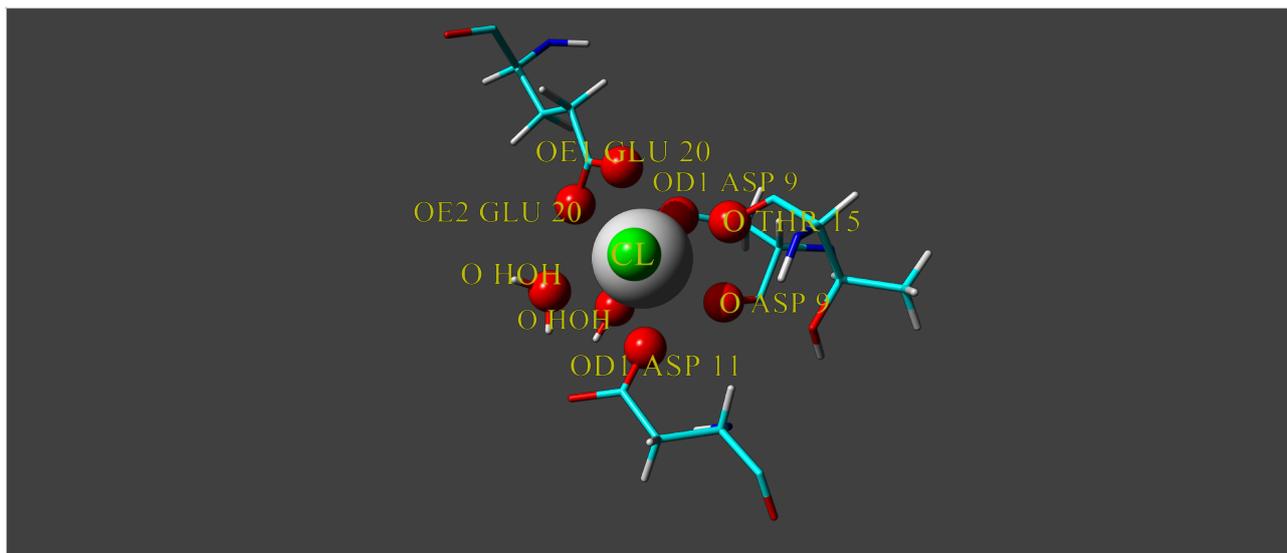
The hydrophobic pocket around Trp 24 appears to have split into two parts due to an elongation of the E-helix into the binding loop compared to the restrained version as can be seen in figure 43. This might suggest an error because Trp 24 has been proven to be important for the stabilization of the relative orientations of the helices and in this case it is only connected to the part of the binding pocket that is interacting with the binding loop through one hydrophobic interaction. The two halves of the protein structure are connected by one hydrophobic interaction between Ile 16 and Asp 9, which is not enough to really stabilize the orientations.



**Figure 43:** Result of the Eu3+ simulation with restraints removed. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. View of the hydrophobic interactions network and the pocket including the key stabilizing residue Trp 24.

A close view of oxygen atoms within 3 Å of the ion is shown in figure 44. Six of the seven ionic bonds from the restrained results have been preserved, there is one Cl<sup>-</sup> ion and two water ions in the area. Based on this, the coordination number of the Eu3+ ion appears to be 9. Experimental data has suggested that the number should be 8, with six protein oxygens and two water molecules. The Cl<sup>-</sup> is therefore extra and suggests that the binding pocket has become unnaturally open.

## Results and discussion



**Figure 44:** Result of the Eu<sup>3+</sup> simulation with restraints removed, starting from the result of Eu<sup>3+</sup> s25. Ionically bound oxygen atoms, with less than 3 Å distance from Eu<sup>3+</sup>, are shown as balls and labeled. Labels are formatted as “<atom name> <residue name>”

The measured distances from the ion to the oxygens that belong to the protein and are within the 3 Å area are shown in table 17. Note that repeated energy minimization experiments would cause the oxygens to recede somewhat from the ion, this is a cause for concern.

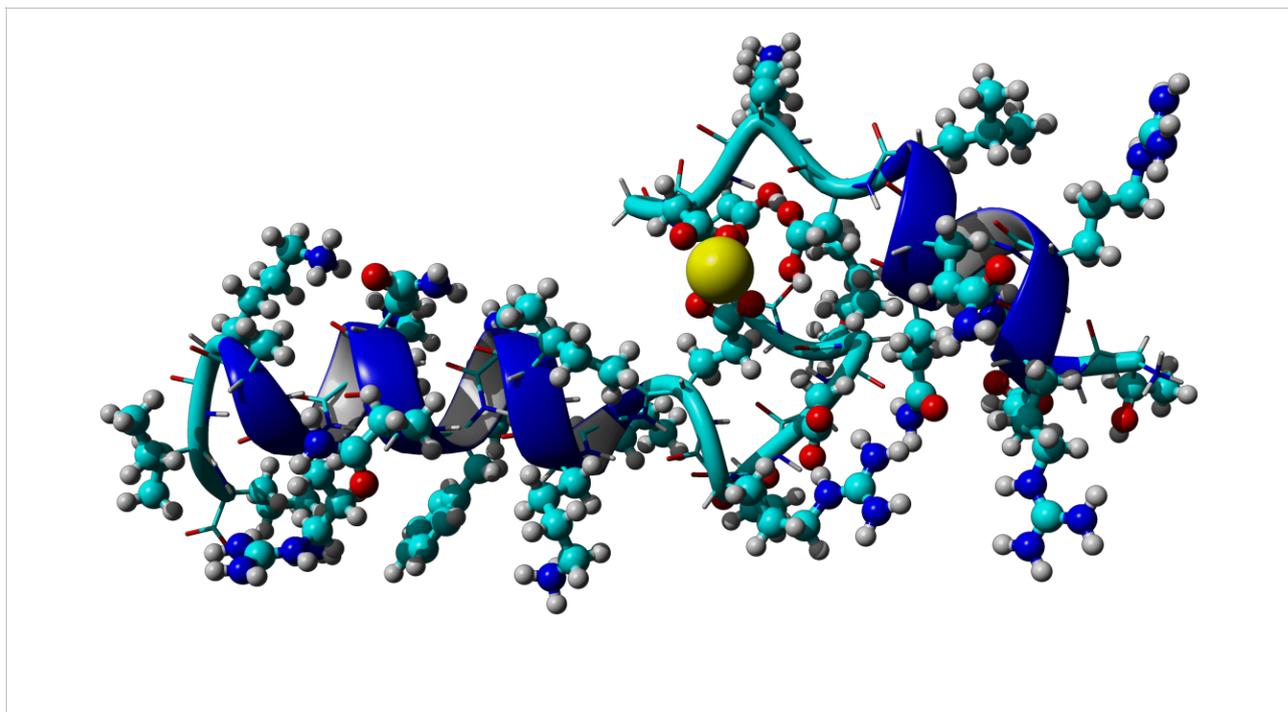
Eu<sup>3+</sup> P3W distances, oxygens closer than 3 Å

Residue (Atom name)	Distance to ion(Å)	Qualitative change from restrained version
Asp 9 (OD)	2.323 Å	Unchanged
Asp 9 (O)	2.597 Å	Unchanged
Asp 11 (OD)	2.291 Å	Unchanged
Thr 15 (O)	2.456 Å	Unchanged
Glu 20 (OE1)	2.344 Å	Unchanged
Glu 20 (OE2)	2.456 Å	Unchanged

**Table 17:** Distances from key oxygens to the Eu<sup>3+</sup> ion without restraints, starting from the result of Eu<sup>3+</sup> s25. Restraint  $d = 2.5$  Å,  $d_{\text{minus}} = 0.5$  Å,  $d_{\text{plus}} = 0.5$  Å. Compared to Eu<sup>3+</sup> s25, the only significant change is that Asp 13 is no longer bound.

### Mg<sup>2+</sup> (no restraints) continuing from Mg<sup>2+</sup> s25

Removing the restraints from the result of Mg<sup>2+</sup> s25 led to the configuration shown in figure 45. The orientation of helices is similar to that of the unrestrained Eu<sup>3+</sup> result.



**Figure 45:** Result of the Mg<sup>2+</sup> simulation with restraints removed, starting from the result of Mg<sup>2+</sup> s25.

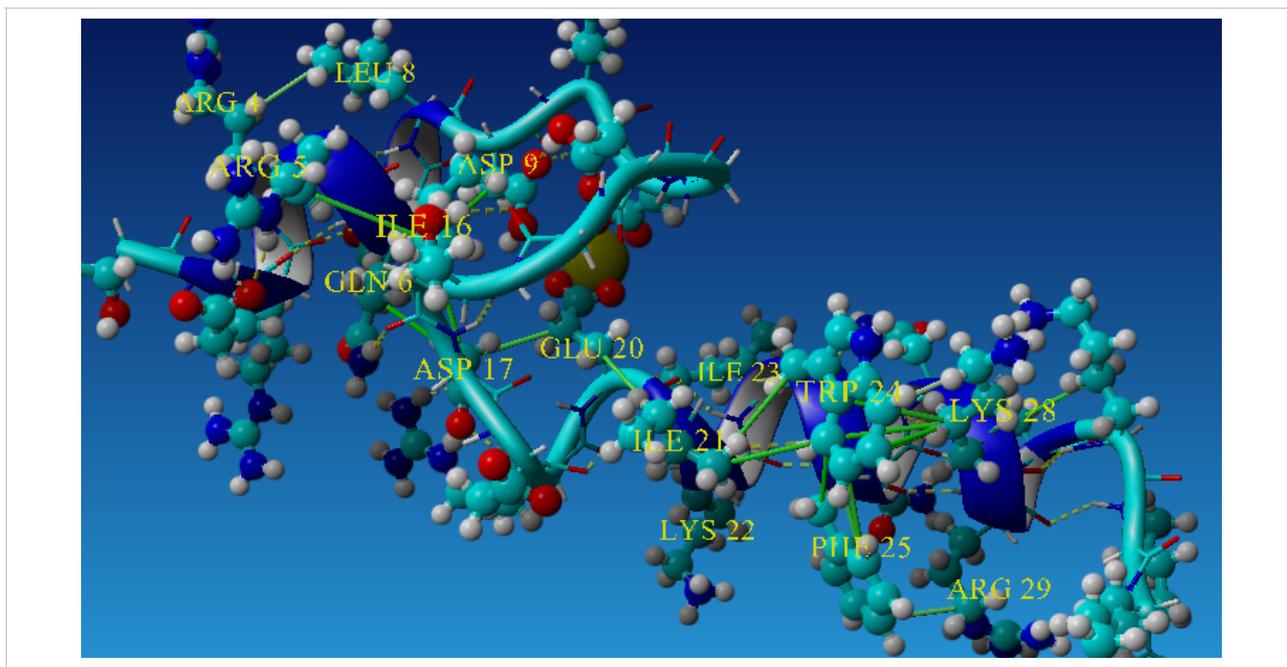
The calculated binding energies, which are negative again, are shown in table 18.

Mg<sup>2+</sup> P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	-19.076 kJ/mol
<b>P3W</b>	-91.331 kJ/mol
<b>Energy from binding to water</b>	72.256 kJ/mol

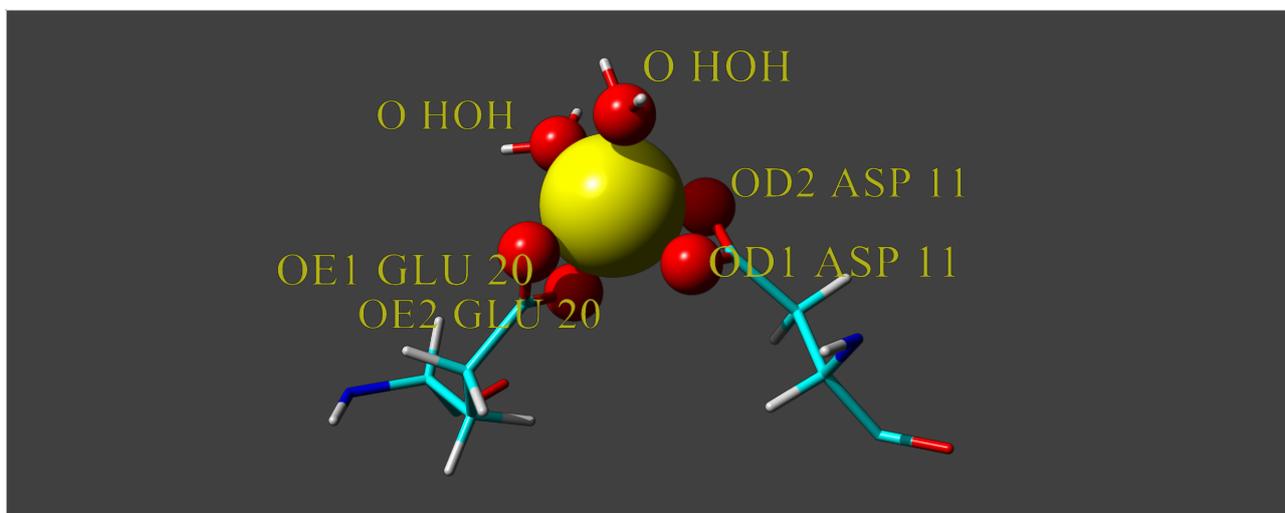
**Table 18:** Calculated binding energies.

The hydrophobic interaction network is shown in figure 46. Trp 24 appears to be playing an important role in the stabilization of the E-helix, but does not appear to have a huge effect on the binding loop. Even though the E-helix has receded and is now starting from residue number 21, the binding loop has not yet had the opportunity to reorganize into the experimentally determined structure. The orientation of the F-helix seems to be dictated by Ile 16, Asp 17, Gln 6, Asp 9 and Arg 5 forming a network of hydrophobic interactions and a hydrogen bond between Ile 16 and Gln 6.



**Figure 46:** Result of the Mg<sup>2+</sup> simulation with restraints removed. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. View of the hydrophobic pocket including the key stabilizing residue Trp 24, which does not seem to affect the loop area very much in this case. Hydrophobic interactions appear to be stabilizing the whole structure.

The six oxygens that were found to be within 3 Å of the ion are shown in figure 47. Two of these oxygens belong to water molecules, three are previously restrained and bound atoms, and one is a new ligand for the ion. Distances to these ions are shown in table 19. All these distances are below 2 Å, even though the distance from a magnesium to an oxygen that is connected to a carbon atom should be around 2.1 Å, based on the findings by Zheng, Chruszcz et al. (2008). This might be due to the inadequate energy minimization, which would also explain the negative binding energies. Unfortunately, the connection between these odd results was realized too late.



**Figure 47:** Result of the Mg<sup>2+</sup> simulation with restraints removed, starting from the result of Mg<sup>2+</sup> s25. Ionically bound oxygen atoms, with less than 3 Å distance from Mg<sup>2+</sup>, are shown as balls and labeled. Labels are formatted as “<atom name> <residue name>”

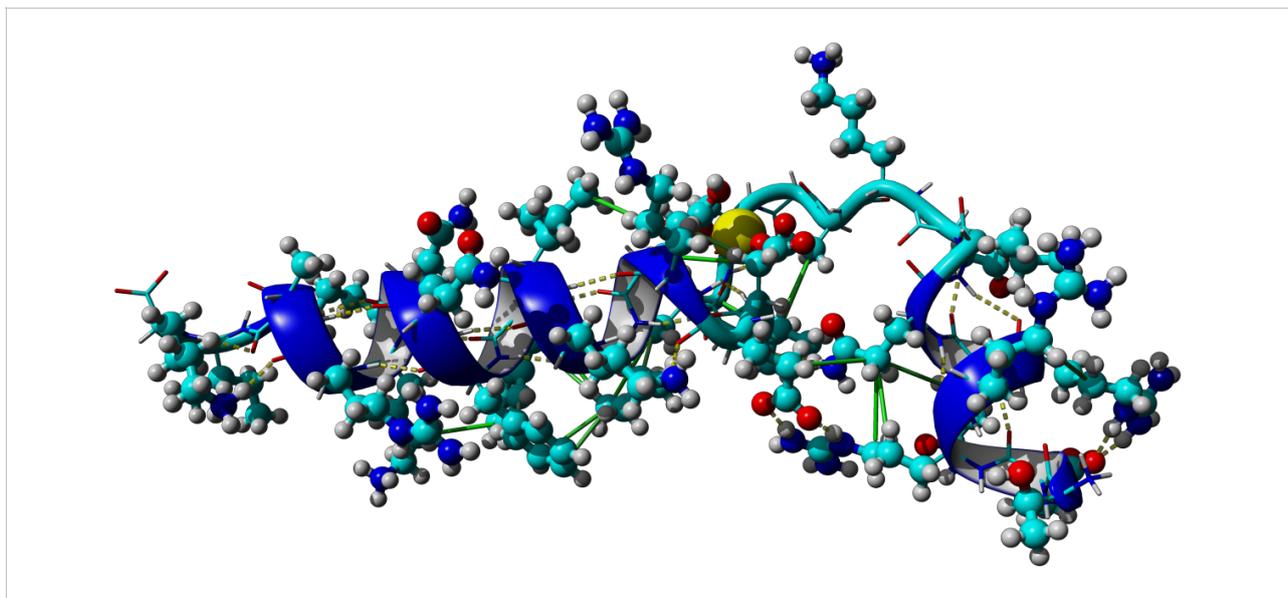
Mg<sup>2+</sup> P3W distances, oxygens closer than 3 Å

Residue (Atom name)	Distance to ion(Å)	Qualitative change from restrained version
Asp 11 (OD1)	1.878 Å	Unchanged
Asp 11 (OD2)	1.867 Å	New ligand
Glu 20 (OE1)	1.829 Å	Unchanged
Glu 20 (OE2)	1.917 Å	Unchanged

**Table 19:** Distances from key oxygens to the Mg<sup>2+</sup> ion without restraints, starting from the result of Mg<sup>2+</sup> s25. Restraint  $d = 2.5$  Å,  $d_{\text{minus}} = 0.5$  Å,  $d_{\text{plus}} = 0.5$  Å. Compared to Mg<sup>2+</sup> s25, Asp 9 and Asp 13 have separated from the ion and Asp 11 has formed a second ionic bond.

### Ca<sup>2+</sup> (no restraints) continuing from Ca<sup>2+</sup> s25

When the result of the restrained simulation Ca<sup>2+</sup> s25 was used as the starting point for a new MDAnalysis without the restraints, the system reached the configuration that can be seen in figure 48. The orientation of the helices is again similar to what was achieved in the unrestrained case for Eu<sup>3+</sup>. In this case, however, there is more distance between the helices than in the case of unrestrained Eu<sup>3+</sup> and Mg<sup>2+</sup>.



**Figure 48:** Result of the Ca<sup>2+</sup> simulation with restraints removed, starting from the result of Ca<sup>2+</sup> s25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

Calculated binding energies are shown in table 20. This time all energies are positive and relatively large. This makes sense as all the bound oxygens, are positioned between 2 and 3 Å away from the ion, as shown in table 21. There is one new ligand on the ion from Asp 11, which is now a bidentate ligand. Asp 9 has left the area around the ion and is no longer bound, allowing the F-helix to drift away from the binding site. Altogether there are still six oxygens belonging to the protein bound to the ion along with one water molecule. This is shown in figure 49.

Ca<sup>2+</sup> P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	207.444 kJ/mol
<b>P3W</b>	59.679 kJ/mol
<b>Energy from binding to water</b>	147.765 kJ/mol

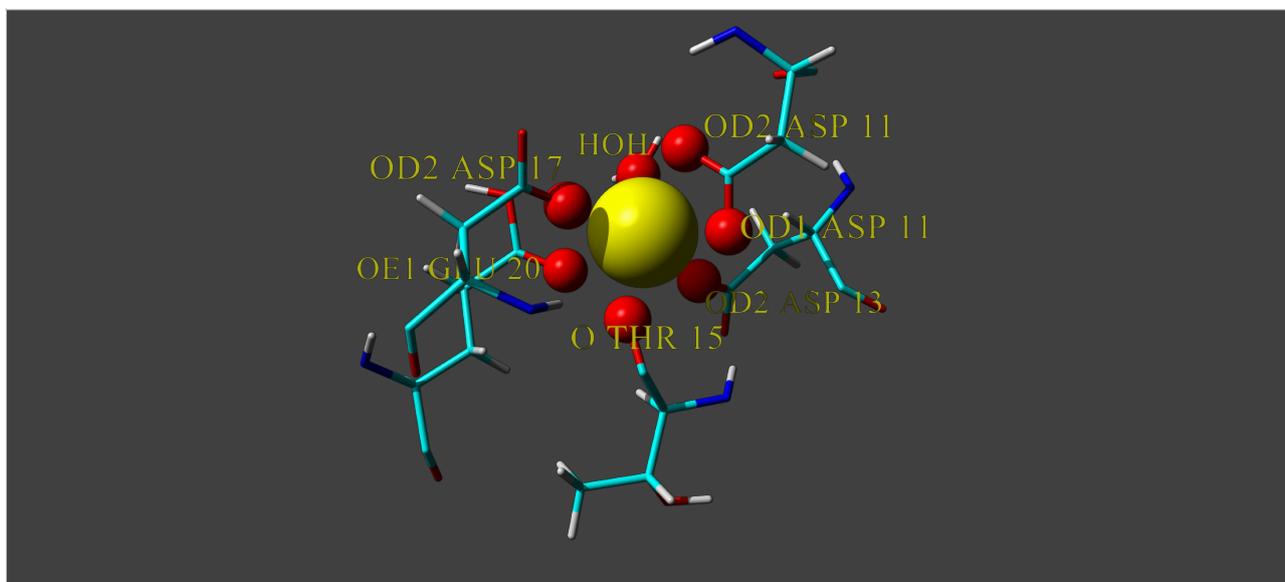
**Table 20:** Binding energies of the Ca<sup>2+</sup> ion without restraints, starting from the result of Ca<sup>2+</sup> s25.

## Results and discussion

Table Ca<sup>2+</sup> P3W distances, oxygens closer than 3 Å

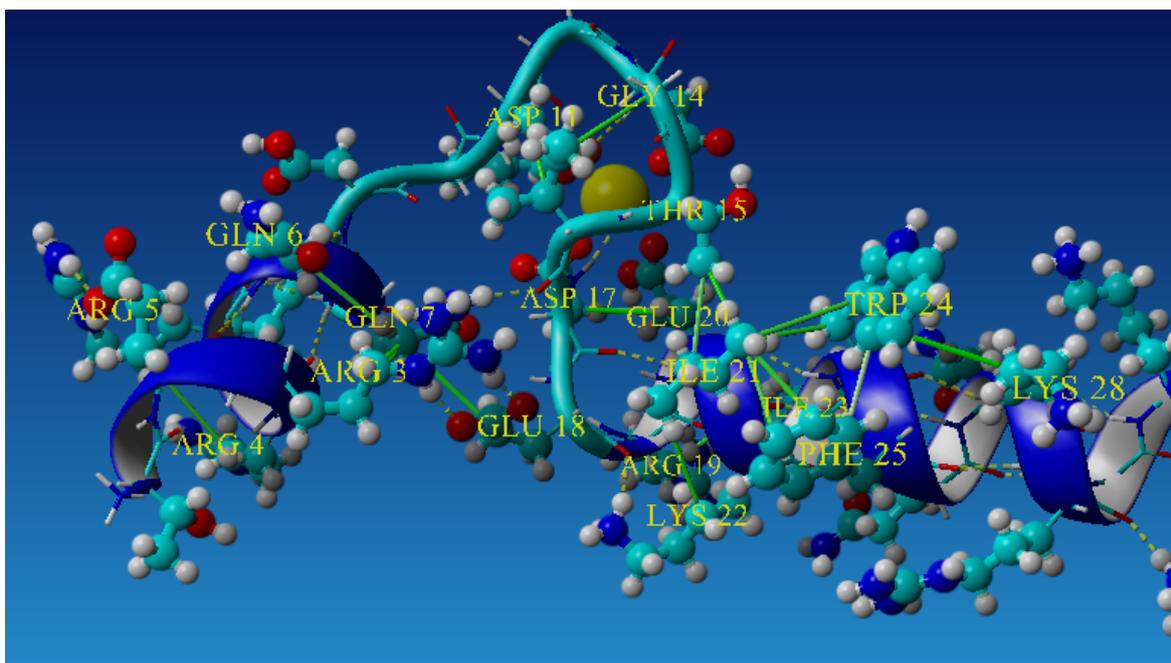
Residue (Atom name)	Distance to ion(Å)	Qualitative change from restrained version
Asp 11 (OD1)	2.261 Å	Unchanged
Asp 11 (OD2)	2.251 Å	New ligand
Asp 13 (OD)	2.182 Å	Unchanged
Thr 15 (O)	2.315 Å	Unchanged
Asp 17 (OE1)	2.166 Å	Unchanged
Glu 20 (OE1)	2.282 Å	Unchanged

**Table 21:** Distances from key oxygens to the Mg<sup>2+</sup> ion without restraints, starting from the result of Mg<sup>2+</sup> s25. Compared to Mg<sup>2+</sup> s25, Asp 9 and Asp 13 have separated from the ion and Asp 11 has formed a second ionic bond.



**Figure 49:** Result of the Ca<sup>2+</sup> simulation with restraints removed, starting from the result of Ca<sup>2+</sup> s25. Ionically bound oxygen atoms, with less than 3 Å distance from the ion, are shown as balls and labeled. Labels are formatted as “<atom name> <residue name> <residue number>”. Water is labeled as HOH.

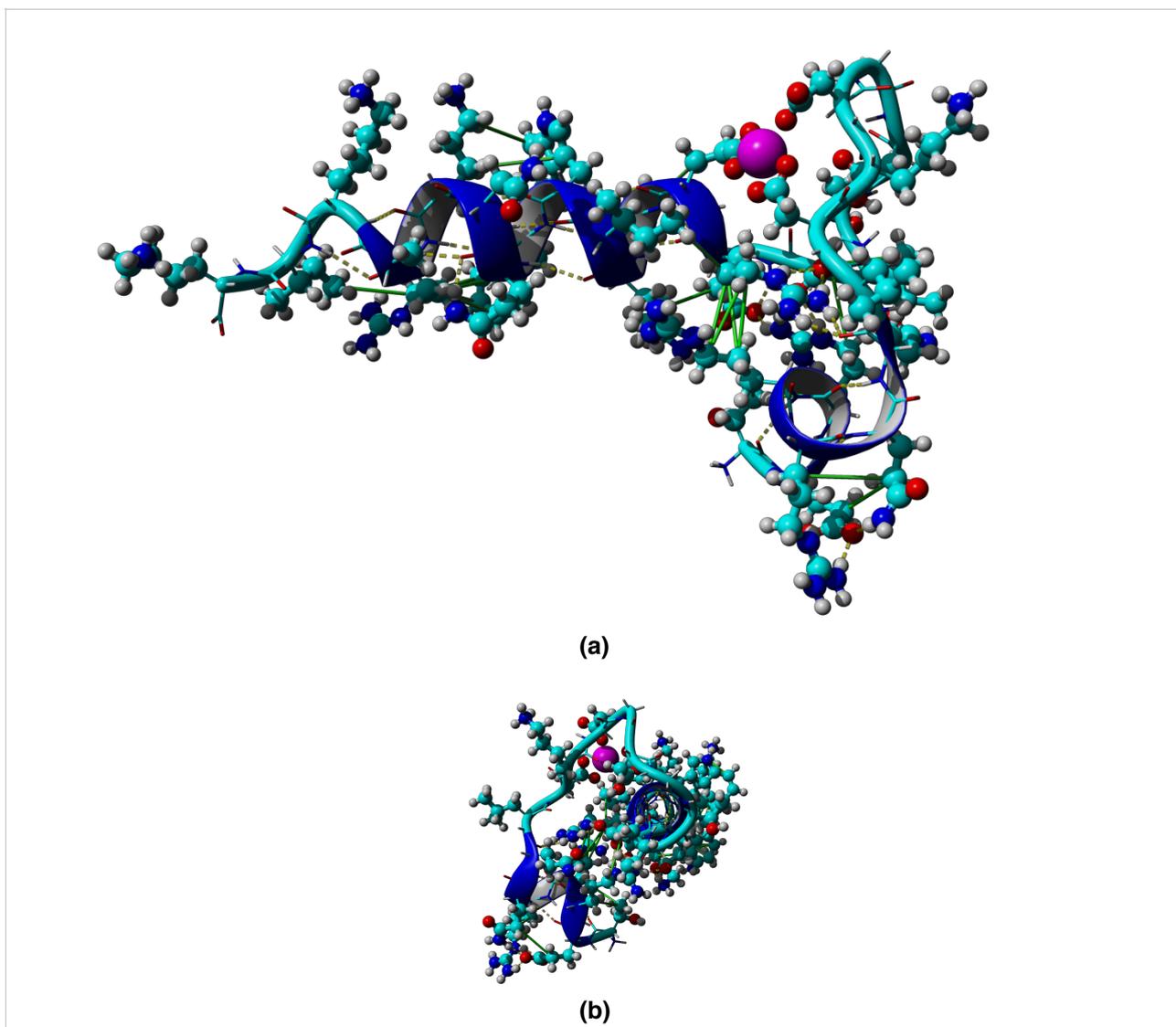
The network of hydrophobic interactions is shown in figure 50. The hydrophobic pocket involving Trp 24 appears to be in good condition. There is a large number of hydrophobic interactions scattered across the protein, including a cluster of interactions between Glu 18, Gln 7, Arg 3 and Gln 6, centered around Gln 7. The most important factor directing the orientation of the F-helix appears to be the fact that there are two hydrogen bonds between Glu 18 and Arg 3 and one between Ile 16 and Gln 7.



**Figure 50:** Result of the Ca<sup>2+</sup> simulation with restraints removed. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. Hydrophobic interactions appear to be stabilizing the whole structure. In this case, two hydrogen bonds from Glu 18 to Arg 3 and one hydrogen bond from Ile 16 to Gln 7 appear to be the key interactions governing the orientation of the F-helix.

### Fe<sup>3+</sup> (no restraints) continuing from Fe<sup>3+</sup> s25

The somewhat surprising result of the continuation simulation of Fe<sup>2+</sup> s25 without the restraints is shown in figure 51 (a). The Protein has reached a conformation where the binding loop has turned so that the binding of the ion is happening in front of the E-helix, not behind it as has been the case in every other simulation until this point. The loop makes a full turn around the axis of the E-helix if the F-helix is counted as an extension of the loop, shown in figure 51 (b).



**Figure 51:** Result of the Fe<sup>3+</sup> simulation with restraints removed, starting from the result of Fe<sup>3+</sup> s25. (a) Front view. (b) Side view showing the turn around the axis of the E-helix. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

Calculated binding energies are shown in table 22. All energies are positive, although the binding energy to the protein itself is relatively low.

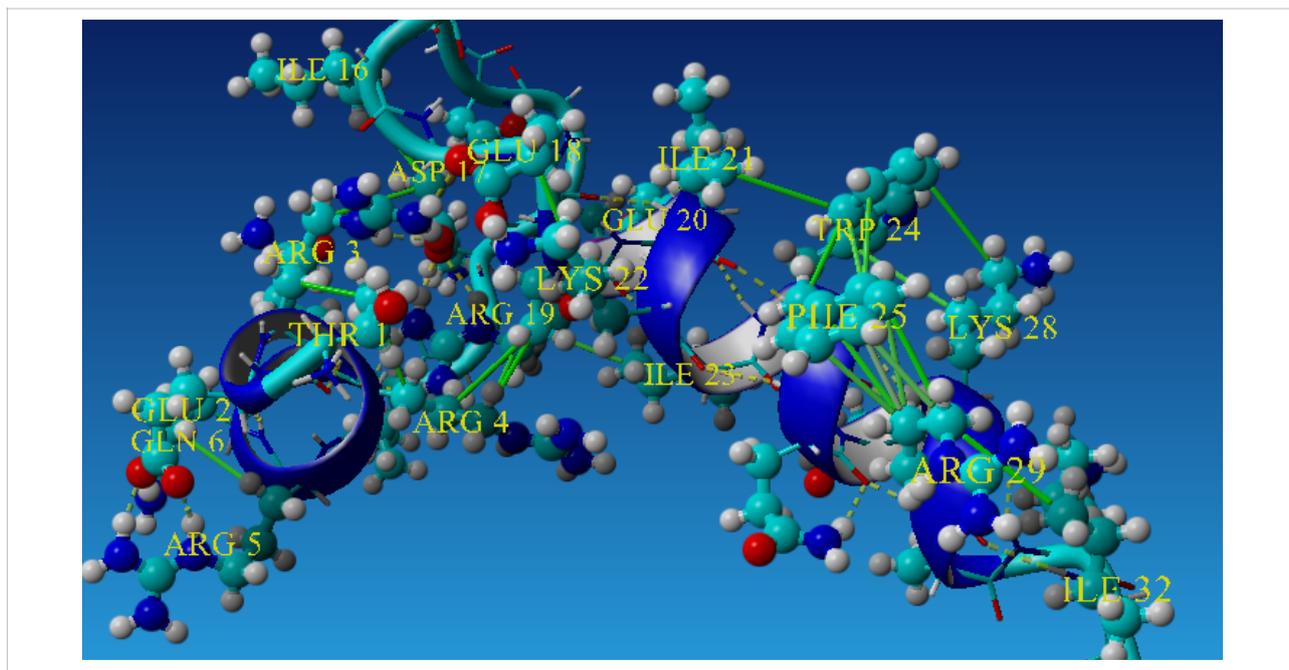
Fe<sup>3+</sup> P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	185.630 kJ/mol
<b>P3W</b>	18.239 kJ/mol
<b>Energy from binding to water</b>	167.391 kJ/mol

**Table 22:** Binding energies of the Fe<sup>3+</sup> ion without restraints, starting from the result of Fe<sup>3+</sup> s25.

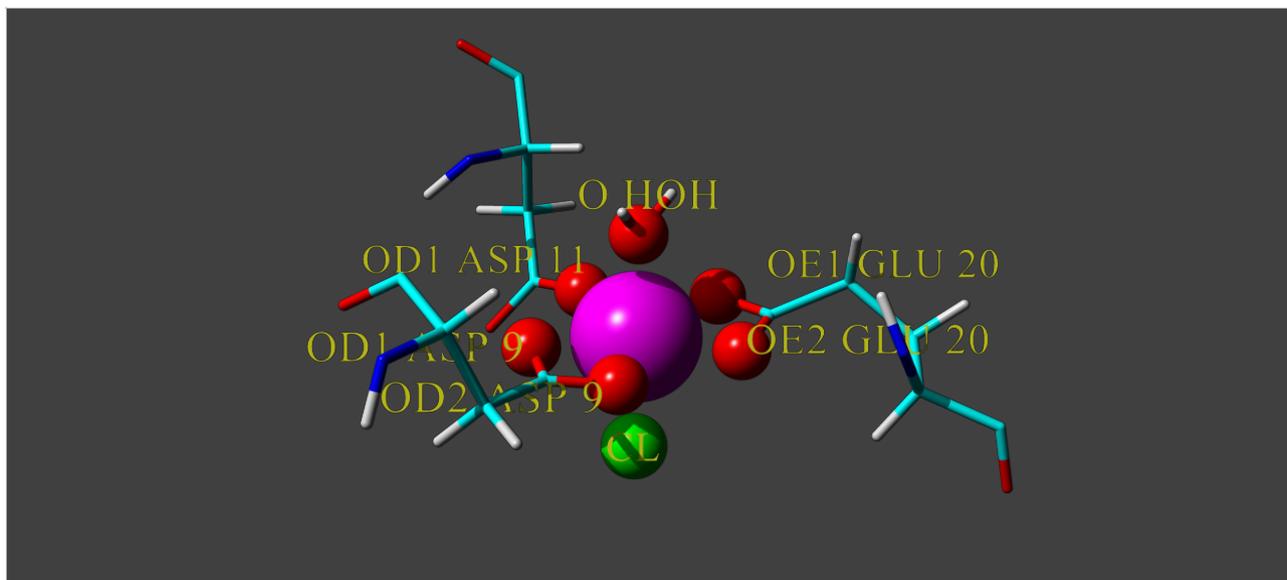
## Results and discussion

Hydrophobic interactions and hydrogen bonds are shown in figure 52. There appear to be strong hydrophobic interactions in the portion of the E-helix starting from Trp 24, however there appears to be no connection to the binding pocket. The F-helix is held in place by hydrogen bonds between Arg 19 and Asp 17. Arg 19 in turn has made a hydrogen bond to Gln 7. There is also some hydrophobic attraction between Arg 4 and Arg 19. Further, Arg 3 has made two hydrogen bonds to Glu 18 and one to asp 17. The primary stabilizing factor of this relative helix orientation comes from the numerous hydrogen bonds between side chains.



**Figure 52:** Result of the Fe<sup>3+</sup> simulation with restraints removed. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. View of the hydrophobic pocket including the key stabilizing residue Trp 24, which does not seem to affect the loop area very much in this case. Hydrophobic interactions appear to be stabilizing the whole structure.

Six oxygens and one Cl<sup>-</sup> atom that are within 3 Å of the Fe<sup>3+</sup> ion are shown in figure 53. One of the oxygens belongs to a water molecule, and five are parts of the protein. Compared to the restrained version, there have been significant changes to the binding pattern. Asp 13 is no longer bound, Asp 9 has changed from being bound through one O $\delta$  and one backbone carbonyl oxygen to being bound by two O $\delta$  atoms and Glu 20 is now bound via two atoms instead of the previous one.



**Figure 53:** Result of the Fe<sup>3+</sup> simulation with restraints removed, starting from the result of Fe<sup>3+</sup> s25. Ionically bound oxygen atoms, with less than 3 Å distance from the ion, are shown as balls and labeled. Labels are formatted as “<atom name> <residue name> <residue number>”.

The measured distances between the ion and the relevant oxygen atoms are shown in table 23. All of the bound atoms appear to be slightly too close to the ion, helping to explain the relatively low binding energy.

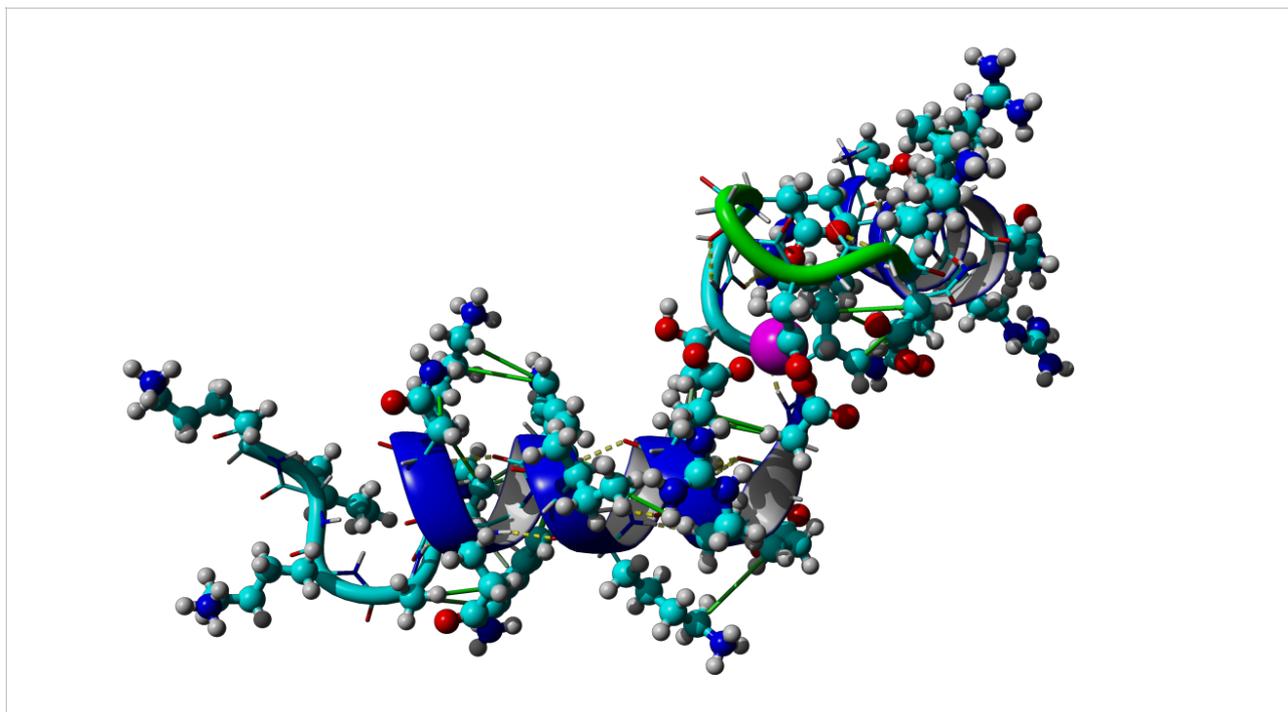
Table Fe<sup>3+</sup> P3W distances, oxygens closer than 3 Å

Residue (Atom name)	Distance to ion(Å)	Qualitative change from restrained version
Asp 9 (OD1)	1.991 Å	Unchanged
Asp 9 (OD2)	1.977 Å	New ligand
Asp 11 (OD1)	1.929 Å	Unchanged
Glu 20 (OE1)	2.005 Å	Unchanged
Glu 20 (OE2)	1.953 Å	New ligand

**Table 23:** Distances from key oxygens to the Fe<sup>3+</sup> ion without restraints, starting from the result of Fe<sup>3+</sup> s25.

### Fe<sup>2+</sup> (no restraints) continuing from Fe<sup>2+</sup> s25

The result of the continuation simulation of Fe<sup>2+</sup> s25 without the restraints is shown in figure 54. The configuration looks very similar to that of the restrained one.



**Figure 54:** Result of the Fe<sup>2+</sup> simulation with restraints removed, starting from the result of Fe<sup>2+</sup> s25. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds.

Calculated binding energies are shown in table 24. The binding energy between the protein and the ion is negative and extremely close to what was found to be the case with the restrained version. The fact that the energy is negative suggests insufficient energy minimization.

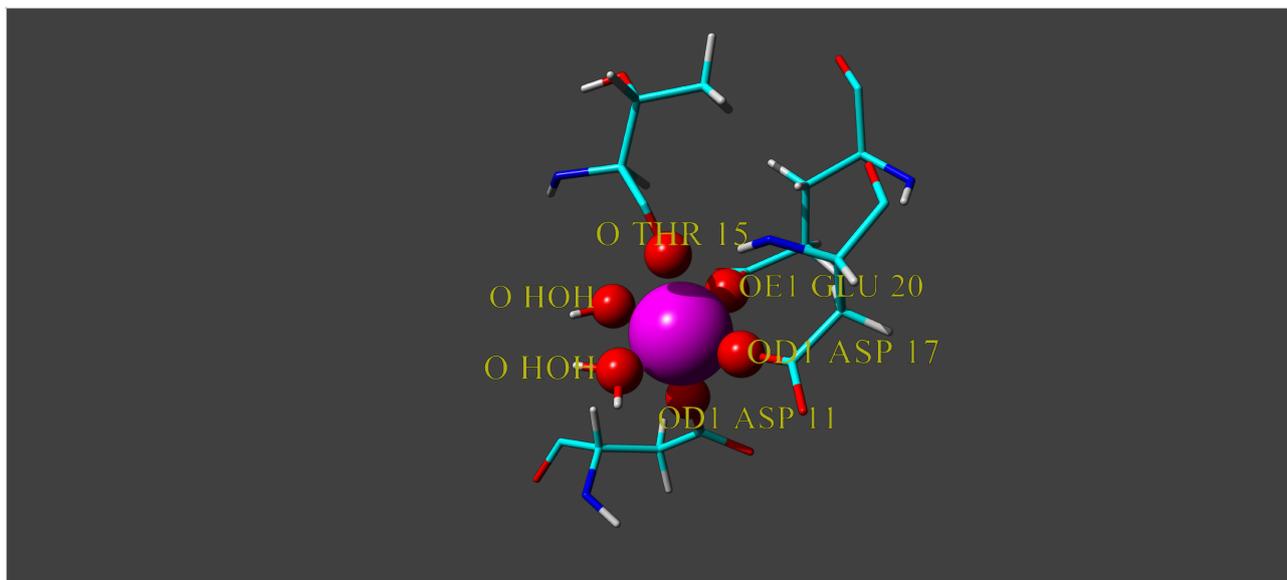
Fe<sup>2+</sup> unrestrained P3W binding energies

	Energy (kJ/mol)
<b>P3W and water</b>	-53.993 kJ/mol
<b>P3W</b>	-64.955 kJ/mol
<b>Energy from binding to water</b>	10.962 kJ/mol

**Table 24:** Binding energies of the Fe<sup>2+</sup> ion without restraints, starting from the result of Fe<sup>2+</sup> s25.

The binding pattern is unchanged from the restrained version. All the same atoms are still bound to the ion as were in the restrained version. This is shown in figure 55.

## Results and discussion



**Figure 55:** Result of the Fe<sup>2+</sup> simulation with restraints removed, starting from the result of Fe<sup>2+</sup> s25. Ionically bound oxygen atoms, with less than 3 Å distance from the ion, are shown as balls and labeled. Labels are formatted as “<atom name> <residue name> <residue number>”.

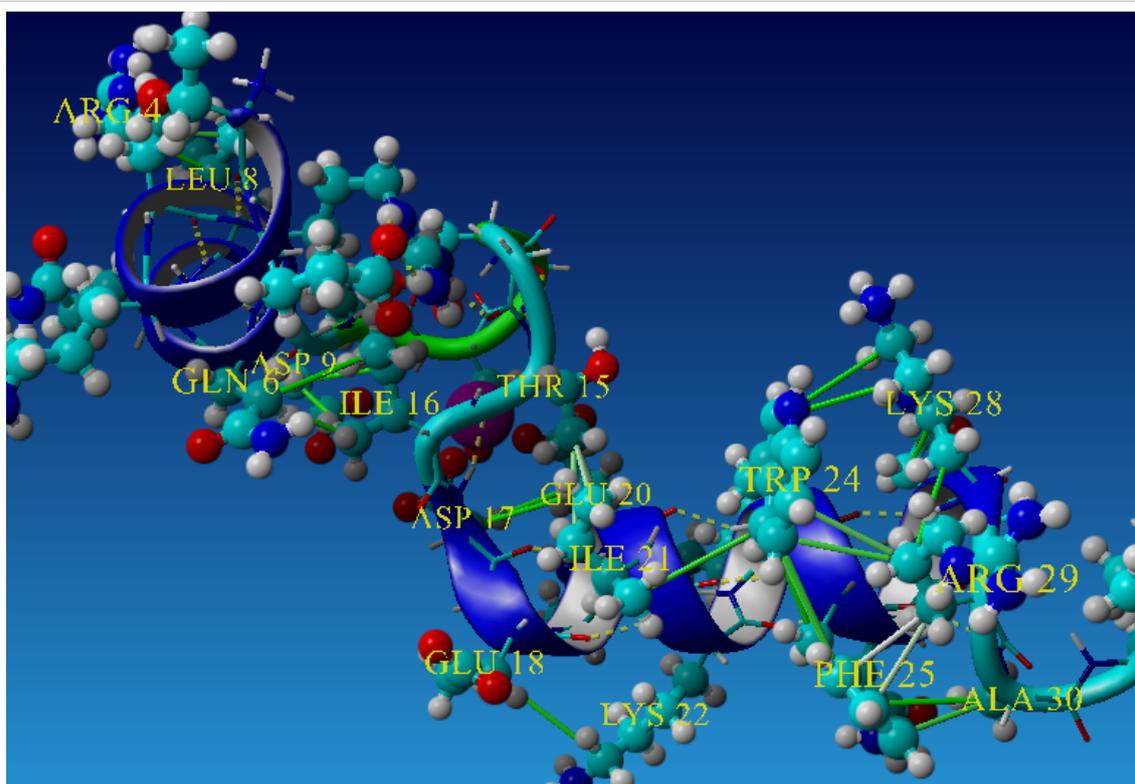
The measured distances between the ion and the relevant oxygen atoms are shown in table 25. Unsurprisingly, all distances are nearly the same as in the case of the restrained version.

Table Fe<sup>2+</sup> P3W unrestrained distances, oxygens closer than 3 Å

Residue (Atom name)	Distance to ion(Å)	Qualitative change from restrained version
Asp 11 (OD1)	1.806 Å	Unchanged
Thr 15 (O)	2.110 Å	Unchanged
Asp 17 (OD2)	1.827 Å	Unchanged
Glu 20 (OE1)	1.802 Å	Unchanged

**Table 25:** Distances from key oxygens to the Fe<sup>2+</sup> ion without restraints, starting from the result of Fe<sup>2+</sup> s25.

This result suggests that Fe<sup>2+</sup> should bind quite stably to P3W, and that it is therefore a good choice of ion. The extensive network of hydrophobic interactions is shown in figure 56. The hydrophobic pocket involving Trp 24 again appears to have been demoted in its importance due to an elongation of the E-helix, however There appears to be a strong hydrophobic pocket between Ile 16, Asp 9 and Gln 6, which might be the primary cause for the stability of this configuration. An examination of the log file from the MDAnalysis reveals that this result was obtained via four refinement steps of the 41. snapshot of the initial 10 ns simulation at a temperature of 315K. Confirming that the system had plenty of time to reorganize during the simulation.



**Figure 56:** Result of the Fe<sup>2+</sup> simulation with restraints removed. Green lines indicate hydrophobic interactions, dotted yellow lines indicate hydrogen bonds. View of the hydrophobic pocket including the key stabilizing residue Trp 24. Hydrophobic interactions appear to be stabilizing the whole structure.

### FRET efficiency calculations

The effective radius between the fluorophores was calculated by taking the radii of both fluorescent proteins and adding them to the measured N to C terminus distance, this assumes that the fluorescent proteins will be pointing in opposite directions along the line that connects the termini of P3W.

The Förster radius between GFP and mCherry, which is a kind of red fluorescent protein, is around 51 Å, meaning that FRET should only be observable in two of the cases shown in table 26. (Albertazzi, Arosio et al. 2009) The radius of GFP and mCherry is around 20 Å each, and must be taken into account. In the most distant case of Ca<sup>2+</sup> with restraints scaled by 25, the approximate distance between the centers of the fluorescent proteins would be 85 Å. This corresponds to about 4.5% efficiency of transmission. (Shaner, Patterson et al. 2007) (Hussain 2012)

## Results and discussion

### N to C terminus distances

Experiment	Distance N to C (Å)	Note
No ion	14.685 Å	$E_{\text{FRET}} = 40\%$
Eu3+ s25	39.538 Å	$E_{\text{FRET}} = 6.5\%$
Ca2+ s25	44.799 Å	$E_{\text{FRET}} = 4.5\%$
Fe3+ s25	32.966 Å	$E_{\text{FRET}} = 10\%$
Fe2+ s25	28.202 Å	$E_{\text{FRET}} = 15\%$
Mg2+ s25	32.405 Å	$E_{\text{FRET}} = 11\%$

**Table 26:** Distances from N to C terminus for experiments where restraint  $d = 2.5$  Å,  $d_{\text{minus}} = 0.5$  Å,  $d_{\text{plus}} = 0.5$  Å and  $T = 315\text{K}$  for the first 10 ns.

Measured distances and calculated FRET efficiencies for the continuation simulations are shown in table 27. The distances appear to have stayed relatively unchanged. None of the cases have caused more than a 10% change in the efficiency of the energy transfer. The largest change is in the case of Fe3+, where the efficiency increased by 7 percentage points.

### N to C terminus distances from simulations with removed restraints

Experiment	Distance N to C (Å)	Efficiency of FRET (change in brackets)	Distance change from restrained version
Eu3+ s25 Continued	39.576 Å	$E_{\text{FRET}} = 6.5\%$ (0%)	0.038 Å
Ca2+ s25 Continued	40.445 Å	$E_{\text{FRET}} = 6.1\%$ (1.6%)	-4.354 Å
Fe3+ s25 Continued	26.313 Å	$E_{\text{FRET}} = 17\%$ (7%)	-6.653 Å
Fe2+ s25 Continued	30.452 Å	$E_{\text{FRET}} = 12.6\%$ (2.6%)	2.223 Å
Mg2+ s25 Continued	38.402 Å	$E_{\text{FRET}} = 7\%$ (-4%)	5.997 Å

**Table 27:** Distances from N to C terminus for experiments where restraints were removed and  $T = 315\text{K}$  for the first 10 ns.

These calculations have an inaccuracy stemming from the fact that they do not take into account the orientation of the helices of P3W.

Drawing straight lines from the termini in directions to where the fluorophores might be positioned, and calculating efficiencies might provide more accurate predictions of the efficiency of the energy transfer. Attaching models of the actual fluorescent proteins to P3W, and running simulations to try and determine the correct location and distance between the fluorophores, would be even better.

Another source of error might be the low quality of the homology model close to the C terminus, this is in reference to residue number 32, which was observed to have allowed but unfavorable torsion angles in the homology model. The uncommon angle appeared to distort the E-helix,

## Results and discussion

making the protein slightly longer. This appears to have been the main difference between the restrained and unrestrained results of Fe<sup>2+</sup>, changing the efficiency estimate by 2.6%.

### Comparison of structures

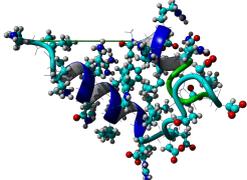
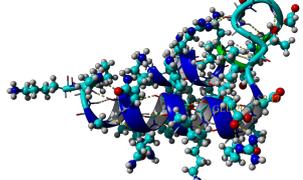
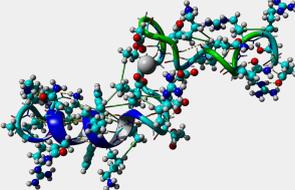
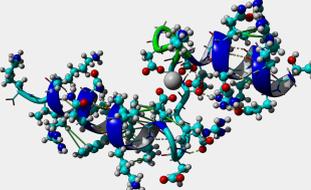
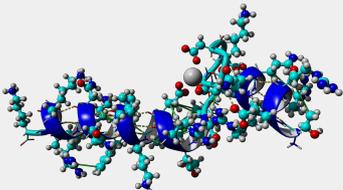
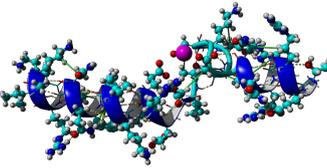
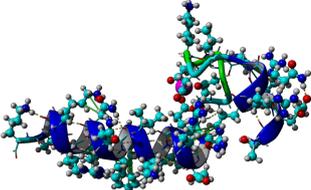
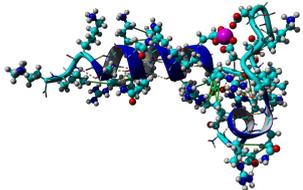
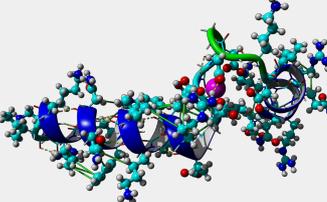
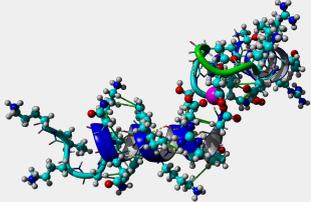
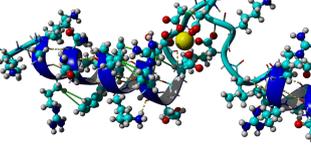
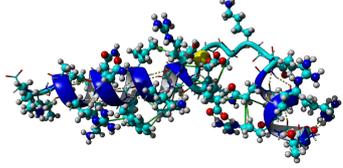
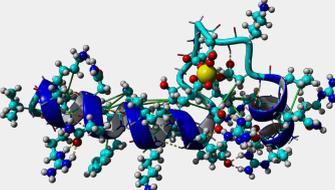
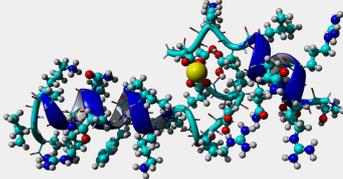
A comparison of results from simulations of P3W with all the ions simulated using the improved MDAnalysis approach where temperatures T<sub>0</sub> = 315K, T<sub>1</sub> = 298K is shown in table 28. First, the case with no ion present was used as a baseline. The results of those simulations indicate that the protein is likely to stay close to the conformation that was obtained from homology modeling if there is no ion present.

The results with strong restraints are relatively similar for Mg<sup>2+</sup>, Fe<sup>3+</sup> and Fe<sup>2+</sup>. The final structure for Eu<sup>3+</sup> can be described as slightly more open, and that of Ca<sup>2+</sup> is even more open than that. For the cases where an ion is present, removing restraints from a structure achieved with strong restraints did not lead to huge changes, except in the case of Fe<sup>3+</sup>, where the end of the helices are noticeably closer. On the other hand, Fe<sup>2+</sup> showed remarkable stability, remaining almost completely unaffected by the removal of restraints.

In the majority of cases, the helices have a similar slanted angle towards each other in the unrestrained results. In the case where no ion is bound to the protein, even though the F-helix was somewhat distorted in the case where the termini were unrestrained, the overall structure was still relatively close to the original homology model, and indeed the same is true in the case where the termini were restrained. The two images appear different because the viewing angles are different. The restrained version is viewed from such an angle that Glu 20 is between the camera and the helix axis, while in the unrestrained case it's situated above the axis.

## Results and discussion

### Comparison of resulting structures.

Ion	Weak restraints	Strong restraints	No restraints (removed)
No ion		Not simulated	
Eu3+			
Fe3+			
Fe2+	Not simulated		
Ca2+	Not simulated		
Mg2+	Not simulated		

**Table 28:** Comparison of the results of simulations where temperatures  $T_0=315\text{K}$  and  $T_1=298\text{K}$ .

### General Discussion

The three homology models that were created for the protein displayed similar structure, however one of them, the model created by SWISS-MODEL, was of significantly lower quality. Of the two remaining models, one scored better in a Psi-Phi analysis, while the other fared slightly better in a composite analysis called QMEAN. Finally, the model created by YASARA was chosen for simulations. All this was done while forgetting that the loop region of the protein should contain a  $\beta$ -strand and that the E-helix should not start with residue number 21 instead of 17, like it does in the case of *engrailed* which is naturally the protein with highest sequence homology to P3W, because that is where most of its sequence originates from. In addition, the part of the loop that is not supposed to form a  $\beta$ -strand and might then have been assumed to be more trustworthy has the lowest local quality score of the whole sequence. This means that all results should be viewed with caution as it is known that the central loop area of the protein has problems stemming from the homology model.

Simulations were done using MDAnalysis, which is a script that is meant to discover conformational changes of proteins, that had been developed and used in a previous project. (Ernits, Bolotakis et al. 2016)

Along the way, problems were discovered in the approach of MDAnalysis and a number of simulations were run, discovering new issues with the approach and finding solutions to work around them. In particular, the order of temperatures was changed so that higher temperature simulations are run before lower temperature ones. In addition, an error was initially made in the setup of the distance restraints used to direct the ion, that was supposed to bind to the protein, to ligate the correct residues that had been previously determined using NMR analysis. (Joel T. Welch 2003)

Another lesson that was learned is that it is necessary to make sure that the distance restraints used are sufficiently strong. This helps make sure that the system can cross potential barriers quickly enough.

Analysis of results revealed that sometimes the binding energy of the ion was negative, meaning that it was energetically unfavorable for the ion to stay bound. This was initially assumed to be due to an unnatural configuration enforced by artificial distance restraints. However, results of MDAnalysis simulations starting from configurations resulting from simulations where strong constraints had been used still exhibited this phenomenon. It turned out that performing repeated energy minimization experiments would cause these apparent negative binding energies to turn positive. This phenomenon was spatially manifested as ligands being situated closer to the ion than natural.

This unfortunately means that all MDAnalysis simulations may suffer from inadequate energy minimization as the same energy minimization method is used extensively throughout the process. It would be a good idea to check whether or not using a more thorough would change the results of

## Results and discussion

an MDAnalysis, and if yes, then to what extent. Unfortunately, time available for this project ran out at this point and it was not possible to do this check.

For the results that were obtained, it appears that in most cases the large scale structure obtained using strong restraints is relatively close to what remains after restraints are removed, however not all cases are like this and therefore it is still valuable to check this.

To act as a FRET sensor, a large change of distance of the first and last residue of the protein is desirable, for that purpose, the most efficient ion would appear to be  $\text{Ca}^{2+}$  as that is the ion which produced the most stretched out configuration. This is assuming that the termini of the protein are close to each other like the homology models suggest. In the opposite case,  $\text{Fe}^{3+}$  gave the best result in the simulations.

The issue of the homology model being inaccurate is an important one and warrants attempts to search for alternative templates in order to try and produce higher a higher quality model.

As always, the best way to create a model would be to solve the structure from experimental data.

## Conclusion

P3W was modeled using three different methods of which the one created using YASARA was chosen. The chosen model still has problems, it received a relatively low quality score from QMEAN, and is simply in disagreement with experimentally determined secondary structure. Simulations were performed with and without a number of binding ions. Various techniques were tried to find the lowest energy configuration of the protein with and without ions. MDAnalysis, the method developed to try and accomplish this task, was improved from the form that had been previously used in another project. A further problem with the MDAnalysis procedure, insufficient energy minimization, was discovered at the end of the project that adds doubt to the accuracy of all the results of this project. Based on the best results achieved in this project, the order of usefulness of the ions simulated is, from best to worst, Ca<sup>2+</sup>, Eu<sup>3+</sup>, Mg<sup>2+</sup>, Fe<sup>2+</sup>, Fe<sup>3+</sup>. The last two ions were calculated to produce an efficiency of FRET of 12.6% and 17% respectively. These numbers are likely to get even worse if actual fluorescent protein models were to be attached to the termini of P3W due to the angles of the helices, this is especially the case with Fe<sup>3+</sup>.

## Outlook

This chapter lists some possible ways to improve and continue on the work of this project.

First, it would be a good idea to determine the effect of the weak energy minimization effect on the results of MDAnalysis, and to repeat simulations with a more thorough minimization process in case the effect is significant.

It might also be possible to create a more realistic model by searching for homologues that have a  $\beta$ -strand in the correct location. Perhaps creating a hybrid model might improve the quality.

FRET efficiency calculations might be improved by taking into account the orientations of the helices and even better would be to add the fluorescent proteins into the simulation, simulate, and then measure the distance.

It might give a better overview of the stability of a configuration to align a number of lowest energy snapshots and analyze the similarities.

Doing repeated and longer simulations might give more accurate results.

Furthermore, it might be interesting to see how the system would behave when the ion was removed from a bound state.

Finally, building a model from experimental data and comparison of experimental and computational results would allow for better understanding of the system, perhaps eventually even rational design of the protein for a FRET sensor.

## Bibliography

Albertazzi, L., D. Arosio, L. Marchetti, F. Ricci and F. Beltram (2009). "Quantitative FRET analysis with the EGFP-mCherry fluorescent protein pair." Photochem Photobiol **85**(1): 287-297.

Arnold, K., L. Bordoli, J. Kopp and T. Schwede (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics(22): 195-201.

Benkert, P., M. Biasini and T. Schwede (2011). "Toward the estimation of the absolute quality of individual protein structure models." Bioinformatics(27): 343-350.

Benkert, P., M. Künzli and T. Schwede (2009). "QMEAN server for protein model quality estimation." Nucleic Acids Research **37**(suppl\_2).

Benkert, P., S. C. E. Tosatto and D. Schomburg (2008). "QMEAN: A comprehensive scoring function for model quality assessment." Proteins: Structure, Function, and Bioinformatics **71**(1): 261-277.

Biasini, M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli and T. Schwede (2014). "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information." Nucleic Acids Research(42).

Biozentrum, U. o. B. (2017). "QMEAN Server - Quick Help." Retrieved 12.06.2017, from <https://swissmodel.expasy.org/qmean/help>.

Bohunicky, B. and S. A. Mousa (2010). "Biosensors: the new wave in cancer diagnosis." Nanotechnology, Science and Applications **Volume 4**: 1-10.

Brennan, R. G. and B. W. Matthews (1989). "The helix-turn-helix DNA binding motif." The Journal of biological chemistry **264**(4): 1903-1906.

Duan, Y., C. Wu, S. Chowdhury, M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo and T. Lee (2003). "A point-charge force field for molecular mechanics simulations of proteins." J.Comput.Chem.(24): 1999-2012.

Ernits, M., A. Bolotakis, E. Deimantavičiūtė and L. P. Taberner (2016). Protein engineering the substrate specificity of a fungal lipase, Aalborg University.

Essman, U., L. Perera, M. Berkowitz, T. Darden, H. Lee and L. Pedersen (1995). "A smooth particle mesh Ewald method." J.Chem.Phys.B(103): 8577-8593.

Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling (2006). "Comparison of multiple AMBER force field and development of improved protein backbone parameters." Proteins(65): 712-725.

Hussain, S. A. (2012). "An Introduction to Fluorescence Resonance Energy Transfer (FRET)." Science Journal of Physics **2012**: 4.

Ikura, M., M. Osawa and J. B. Ames (2002). "The role of calcium-binding proteins in the control of transcription: structure to function." BioEssays **24**(7): 625-636.

- Jian, P. and X. Jinbo (2011). "A multiple-template approach to protein threading." PROTEINS(79).
- Jian, P. and X. Jinbo (2011). "RaptorX: exploiting structure information for protein alignment by statistical inference." PROTEINS **79**(S10): 161-171.
- Jianzhu, M., W. Sheng, Z. Feng and X. Jinbo (2013). "Protein threading using context-specific alignment potential." Bioinformatics (Proceedings of ISMB 2013) **29**(13): i257-i265.
- Joel T. Welch, W. R. K., Sonya J. Franklin (2003). "Lanthanide-binding helix-turn-helix peptides: Solution structure of a designed metallonuclease." PNAS **100**(7): 5.
- Källberg, M., H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu and J. Xu (2012). "Template-based protein structure modeling using the RaptorX web server." Nature Protocols **7**(8): 1511-1522.
- Konagurthu, A. S., J. C. Whisstock, P. J. Stuckey and A. M. Lesk (2006). "MUSTANG: A multiple structural alignment algorithm." Proteins(64): 559-574.
- Kovacic, R. T., J. T. Welch and S. J. Franklin (2003). "Sequence-selective DNA cleavage by a chimeric metallopeptide." Journal of the American Chemical Society **125**(22): 6656-6662.
- Kretsinger, R. H., V. N. Uversky and E. A. Permyakov (2013). Encyclopedia of metalloproteins.
- Krieger, E. (2017). Your guide to YASARA Structure. YASARA 17.4.14, YASARA Biosciences GmbH.
- Krieger, E., K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker and K. Karplus (2009). "Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8." Proteins **77** Suppl(9): 114-122.
- Krieger, E., J. Nielsen, C. Spronk and G. Vriend (2006). "Fast empirical pKa prediction by Ewald summation." J.Mol.Graph.Model.(25): 481-486.
- Krieger, E. and G. Vriend (2014). "YASARA View - molecular graphics for all devices - from smartphones to workstations." Bioinformatics(30): 2981-2982.
- Krieger, E. and G. Vriend (2015). "New ways to boost molecular dynamics simulations." J.Comput.Chem. **36**: 996-1007.
- Kuchelmeister, H. Y. and C. Schmuck (2013). "Designing Receptors for the Next Generation of Biosensors." springer **12**: 53-65.
- Li, P., B. Roberts, D. Chakravorty and K. M. Jr (2013). "Rational design of particle mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent." J.Chem.Theory Comput.(9): 2733-2748.
- Li, P., L. Song and K. M. Jr (2015). "Parameterization of highly charged metal ions using the 12-6-4 LJ type nonbonded model in explicit water." J.Phys.Chem.B(119): 883-895.

S.C. Lovell, I. W. D., W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson (2002). "Structure validation by C geometry: / and C deviation." Proteins: Structure, Function & Genetics(50): 437-450.

Schwede, T. (2003). "SWISS-MODEL: an automated protein homology-modeling server." Nucleic Acids Research **31**(13): 3381-3385.

Shaner, N. C., G. H. Patterson and M. W. Davidson (2007). "Advances in fluorescent protein technology." J Cell Sci **120**(Pt 24): 4247-4260.

Sippl, M. J. (1995). "Knowledge-based potentials for proteins." Current Opinion in Structural Biology(0959-440X): 229-235.

Zheng, H., M. Chruszcz, P. Lasota, L. Lebioda and W. Minor (2008). "Data mining of metal ion environments present in protein structures." J Inorg Biochem **102**(9): 1765-1776.