

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
M.Sc. Mari Chikvaidze,
born in Tbilisi, Georgia.
Oral examination: March, 2013

UNDERSTANDING STRUCTURE, DYNAMICS AND FUNCTION
OF *ESCHERICHIA COLI* ALKALINE PHOSPHATASE BY
MOLECULAR MODELLING AND SIMULATIONS

Referees: PD. Dr. Karsten Rippe
Prof. Dr. Rebecca Wade

Dedicated to George and Mia Valentina.

ABSTRACT

The metalloenzyme *Escherichia coli* (*E. coli*) Alkaline Phosphatase (AP) has a homo-dimeric quaternary structure, which is essential for the enzyme to achieve its catalytic activity, the acceleration of phosphoester hydrolysis. Employing computational approaches, at different levels of complexity and sophistication, this thesis aims at understanding the interplay of structure, dynamics and function of this important enzyme, that is responsible for the supply of vital inorganic phosphate.

At low resolution, coarse grained models of the enzyme are constructed, using an Elastic Network Model (ENM), to explain the influence of the global dynamics, defined by the quaternary structure of the enzyme, on its functionality. Comparative analysis of the collective motions, of individual subunits within a homo-dimer of apo and holo enzymes, allows us to interpret the experimental proposal of negative cooperativity. The intrinsic asymmetry of the subunits within the dimer, is already encoded in the enzymes three-dimensional structure, as becomes apparent from the normal mode analysis of the low resolution ENMs, with alternate opening and closing motions present in only one subunit at a time.

These results, in conjunction with the analysis of Molecular Dynamics (MD) simulations of the atomistic models, at the nanosecond time scale, suggest a mechanical coupling between the correlated motions of individual subunits and their active sites, via the dimer interface. The negative cooperativity of the subunits is further explained by the analysis of multiple, independent MD simulations that reveals subtle differences in the hydrogen bonding network of the subunits and dynamics of their active site residues, demonstrating a distinct asymmetric behaviour, with increased flexibility of one subunit versus rigidity of a second one. Information about structural changes is transmitted between the subunits via the hydrogen bonding network across the interface. At least two such communication pathways can be proposed, based on the analysis of MD simulations, both via the interfacial α -helix containing residues T559 and T555 that are hydrogen bonded to the residue N416, which is in turn, connected to the residues in the active site.

Analysis of the MD simulations confirms that, while the active site of one subunit retains the inorganic phosphate product, as demonstrated by the small changes in the distances between the active site residues, another subunit is "letting the product go" in order to make the active site available for another turnover of substrate binding.

Analyses of the MD simulations further suggest that correlated motions of the monomeric subunits of the dimer, as well as the dynamics, and hence the architecture of the active sites, play an important role in the functionality of Alkaline Phosphatase. Although each subunit of the enzyme is equipped with its own catalytic sites, a monomeric AP does not exist in nature and the engineered mutants have significantly reduced activity. Due to the absence of a crystal structure, a model of a monomeric form of the enzyme is constructed based on the crystal structure of the apo dimer. In addition, a T59R mutant is built, where interface residues T59 and T559 are substituted by a bulky and charged Arginine residues. Experimental evidence suggests that such a substitution destabilizes the dimeric interface, resulting in a separation into isolated monomers, with reduced structural stability and catalytic activity. Our MD simulation results confirm that the overall dynamic behaviour of the monomer is different from that of the corresponding dimer and resembles more that of the T59R mutant. Furthermore, comparative analysis based on MD simulations of dimeric and monomeric forms of AP reveals important structural and dynamic features enabling the native dimer to be catalytically functional. The stabilisation provided by the interface of the two subunits in the dimeric form of AP is found to be essential for a catalytically competent structure of the active sites. Breaking of the hydrogen bond between residues Y402 and D330, that are located near the active site, as observed in the MD simulations of the monomer, results in the incorrect positioning of the catalytically important, divalent, zinc ion.

Understanding the nature of the correlated motions of the subunits within the dimer, and their connection to the enzyme's activity is an important step in completing our knowledge on structure-dynamics-function relationship of *E. coli* Alkaline Phosphatase and related enzymes. Our findings confirm that the structural stability of dimeric AP, provided by the hydrogen bonding network across the interface, is essential for the enzymatic activity. By a combination of different computational approaches, we gained an in-depth understanding of the relationship between the enzymes' dimeric quaternary structure and its functionality.

ZUSAMMENFASSUNG

Das Enzym *Escherichia coli* (*E. coli*) Alkaline Phosphatase (AP) hat eine für die katalytische Aktivität des Enzyms, die Beschleunigung von Phosphoester-Hydrolyse Reaktionen, essentielle homo-dimere Quarternärstruktur.

In dieser Arbeit werden computergestützte Methoden verschiedener Komplexität mit dem Ziel verwendet, das Zusammenspiel von Struktur, Funktion und Dynamik dieses wichtigen Enzyms zu verstehen, das für die lebenswichtige Versorgung mit anorganischem Phosphat verantwortlich ist.

Auf der Ebene mit niedrigster Auflösung wurde ein "coarse-grained"-Modell, ein elastisches Netzwerk-Modell (ENM), des Enzyms konstruiert, um den Einfluss der globalen Dynamik des Enzyms auf seine Funktionalität zu untersuchen. Diese globale Dynamik ist wiederum durch seine Quarternärstruktur vorgegeben. Ein Vergleich der kollektiven Bewegungen der individuellen Untereinheiten im Homodimer des Apo- und Holoenzym ermöglicht uns, die experimentell motivierte Idee der negativen Kooperativität zu interpretieren. Die intrinsische Asymmetrie der Untereinheiten des Dimers ist bereits in der dreidimensionalen Struktur des Enzyms enthalten. Dies wird durch die Normalmodenanalyse des ENMs deutlich: Bewegungen, die als Öffnen und Schließen aufgefasst werden können, zeigen sich immer nur abwechselnd in den beiden Untereinheiten. Dieses Ergebnis, zusammen mit der Hauptkomponentenanalyse atomistischer molekulardynamischer Simulationen auf der Nanosekundenzeitskala legt nahe, dass eine mechanische Kopplung zwischen den korrelierten Bewegungen der einzelnen Untereinheiten und deren aktiver Zentren über die Dimer-Grenzfläche existiert.

Die negative Kooperativität der Untereinheiten kann darüberhinaus mittels der Analyse mehrerer unabhängiger MD-Simulationen erklärt werden. Die Simulationen weisen feine Unterschiede im Wasserstoffbrückenbindungsnetzwerk der beiden Untereinheiten und in der Dynamik der Residuen in den aktiven Zentren auf. Diese zeigen eine deutliche Asymmetrie in welcher in einer Unterheit die Flexibilität dieser Residuen gegenüber der anderen deutlich erhöht ist, und in der anderen erniedrigt. Die Analyse der MD-Simulationen bestätigt, dass, während die eine Untereinheit das Phosphatprodukt zurückhält, erkennbar an den geringen Änderungen der Abstände der Residuen im aktiven Zentrum, die andere Untereinheit das Produkt loslässt, um deren aktives Zentrum für die Bindung eines neuen Substrats verfügbar zu machen.

Desweiteren legt die Analyse der MD-Simulationen nahe, dass die korrelierten Bewegungen der monomeren Untereinheiten sowie die Dynamik, und damit die Architektur der aktiven Zentren, eine wichtige Rolle in der katalytischen Funktion der Alkaline Phosphatase spielen. Obwohl jede der beiden Untereinheiten des Enzyms mit ihrem eigenen katalytischem Zentrum ausgestattet ist, existiert AP in der Natur nicht als Monomer, und auch die künstlich erzeugten Mutanten zeigen keine nennenswerte Aktivität. Mangels einer vorhandenen Kristallstruktur wurde in dieser Arbeit ein Modell für monomerische Alkaline Phosphatase konstruiert, das auf der Kristallstruktur des AP-dimers basiert. Zusätzlich wurde ein T59R-Mutant modelliert, in dem die Grenzflächenresiduen T59 und T559 durch sterisch anspruchsvolle und positiv geladene Arg-Residuen ersetzt wurden. Experimentellen Hinweisen zufolge destabilisiert solch eine Substitution die dimerische Schnittstelle und führt zu einer Trennung in isolierte Monomere mit reduzierter struktureller Stabilität und herabgesetzter katalytischer Aktivität. Die in dieser Arbeit durchgeführten MD-Simulationen bestätigen, dass die Gesamt-Dynamik des Monomers verschieden von der des entsprechenden Dimers ist und mehr der des T59R-Mutanten ähnelt. Darüberhinaus deckt die auf den MD-Simulationen basierende vergleichende Analyse der dimeren und monomeren Form der AP wichtige strukturelle und dynamische Aspekte auf, die dem Dimer seine katalytische Funktionalität verleihen. Die Stabilisierung, welche durch die Kopplung der beiden Untereinheiten in der dimeren Form der AP erreicht wird, ist essentiell für eine katalytisch kompetente Struktur der aktiven Zentren. In der MD-Simulation des Monomers wird ein Bruch der Wasserstoffbrückenbindung zwischen den Residuen Y402 und D330, welche sich nahe dem aktiven Zentrum befinden, beobachtet. Die daraus folgende inkorrekte Positionierung der katalytisch wichtigen, divalenten, Zink-ionen erklärt die reduzierte Aktivität in monomerer AP.

Das Verständnis der korrelierten Bewegungen der beiden Untereinheiten des Dimers, und deren Zusammenhang mit der enzymatischen Aktivität ist ein wichtiger Schritt hin zur Vervollständigung unseres Wissens über die Verknüpfung von Structure, Dynamik und Funktion von *E. coli* Alkaline Phosphatase und verwandter Enzyme. So bestätigen die in dieser Arbeit erzielten Resultate, dass die strukturelle Stabilität dimerer AP, bewirkt durch das Wasserstoffbrückenbindungsnetzwerk, essentiell für die enzymatische Aktivität ist. Durch die Kombination verschiedener computergestützter Verfahren konnte ein tiefgehender Einblick in den Zusammenhang zwischen der Quaternärstruktur des Enzyms und seiner Funktionalität gewonnen werden.

PUBLICATIONS

In connection with work performed during the thesis, the following papers have been prepared for submission into peer refereed journals:

- Mari Chikvaidze and Petra Imhof. A Comparative Molecular Dynamics Study of Mono- and Dimeric Forms of the Enzyme *Escherichia coli* Alkaline Phosphatase. Submitted.
- Mari Chikvaidze and Petra Imhof. Collective Dynamics and Importance of Subunit Interactions and Cooperativity in *Escherichia coli* Alkaline Phosphatase. In preparation.

“Tell your son to stop trying to fill your head with science

— for to fill your heart with love is enough.”

Richard Feynman¹

ACKNOWLEDGMENTS

Firstly, I would like to express the most sincere gratitude to my supervisor, Dr. Petra Imhof for her support and guidance, during the course of my PhD studies.

I am very grateful to PD. Dr. Karsten Rippe and Prof. Dr. Rebecca Wade, Prof. Dr. Oliver Gruss and Prof. Dr. Ursula Kummer, for their interest in being members of my PhD thesis defence committee.

Many thanks to all my friends and colleagues in Heidelberg for their readiness to help and share the bittersweet moments of my life, as a PhD student.

My special thanks to Dr. Mai Zahran for getting me started with NAMD, to Dr. Manuel Montenegro for his help with zinc parameters and to Dr. Héctor Mauricio Castañeda Cortés for interesting scientific discussions and most importantly, for his emotional support and friendship. I would like to say thank you to my dear friend, Dr. Anna Feldman-Salit for her friendship and support, especially during the week before submission of this thesis. I'd also like to thank Dr. George Giorgidze for his suggestions on programming and for proof reading parts of this thesis.

The simulations presented in this thesis were run on the Heidelberg Linux Cluster System (Helics) at the Interdisciplinary Centre for Scientific Computing (IWR) of the University of Heidelberg and clusters of the North-German Supercomputing Alliance (HLRN).

I acknowledge the International Max Planck Research School for Quantum Dynamics in Physics, Chemistry and Biology (IMPRS QD) for the doctoral scholarship and the generous financial support letting me attend number of useful scientific events.

Finally, I would like to thank my parents and relatives back in Georgia, in Canada and in the USA, and my dear friends in lovely little town of Tübingen, for their constant love and encouragement.

My biggest thank you to my loving little family, George and Mia, for always being there for me, and for keeping my spirits up.

¹ No Ordinary Genius : The Illustrated Richard Feynman (1996), by Christopher Simon Sykes, p. 161

CONTENTS

I	INTRODUCTION	1
1	INTRODUCTION	3
1.1	Motivation	3
1.2	Outline	6
II	BACKGROUND	7
2	BACKGROUND	9
2.1	Structure, dynamics and function of enzymes	9
2.2	Alkaline Phosphatase	12
2.3	Biomolecular Modelling and Simulations	18
2.4	Elastic Network Models	19
2.4.1	Gaussian Network Model	21
2.4.2	Anisotropic Network Models	22
2.4.3	Normal Mode Analysis	22
2.5	Classical Molecular Dynamics	24
III	MATERIALS AND METHODS	29
3	MATERIALS AND METHODS	31
3.1	Materials	31
3.2	Methods	31
3.3	Structural models	32
3.3.1	Coarse-grained models	32
3.3.2	Atomistic models	32
3.4	Molecular Dynamics Simulations	37
3.5	Analysis of MD Simulations	38
IV	RESULTS	39
4	RESULTS	41
4.1	Global Dynamics	41
4.1.1	Gaussian Network Model	41
4.1.2	Anisotropic Network Model	41
4.2	Local Dynamics	43
4.2.1	Holo and Apo Enzymes	43
4.2.2	The Wild Type (WT) Dimer, Monomer and T59R Mutant of the Apo Enzyme	61
V	DISCUSSION	91
5	DISCUSSION	93
5.1	The Holo and Apo Forms of the Enzyme	93
5.2	The WT Dimer, Monomer and T59R Mutant of the Apo Enzyme	95

VI	CONCLUSIONS	99
6	CONCLUSIONS	101
VII	FUTURE WORK	103
7	FUTURE WORK	105
VIII	APPENDIX	107
A	APPENDICES	109
	BIBLIOGRAPHY	119

LIST OF FIGURES

- Figure 1 Crystal structure of a wild type Alkaline Phosphatase (1ed9.pdb) coloured by chains and in new cartoon representation showing the secondary structure elements in the foreground and in surface representation, showing the overall shape of the enzyme in the background. Metal ions in the active site are not shown for clarity and the black star placed at the axes of symmetry schematically depicts location of the mutated residues. 4
- Figure 2 Diagrams to show the induced fit hypothesis of enzyme activity. 10
- Figure 3 Crystal structure of Alkaline Phosphatase in ribbon representation (1ed8.pdb), coloured by secondary structure elements, with a closeup showing inorganic phosphate and metal ions in licorice representation and coloured by the atom type. 12
- Figure 4 Schematic representation of the reaction mechanism of Alkaline Phosphatase. 13
- Figure 5 Crystal structure of AP (1ed9.pdb) in ribbon representation and coloured by chains, with disulphide bridges in space filling representation and coloured by atom type. 14
- Figure 6 Chain A of a wild type dimer shown in a secondary structure representation and coloured in grey (a.). Close up of the active site with metal ions and their ligands, and the key amino acid residues shown in licorice and coloured by residue type: green - polar, red - acidic and blue - basic. Water molecules are shown as blue balls (b.). Dotted lines correspond to the distances between metal ions in Angstroms (c.). 15
- Figure 7 Structure of Alkaline Phosphatase coloured by chains and in new cartoon representation showing the secondary structure elements and the interface in space filling representation. (a.) The side view; (b.) The top view with the interface in foreground. 17

- Figure 8 Schematic representation of nodes in elastic network model. Every node is connected to its spatial neighbours by uniform springs. Distance vector between two nodes, i and j , shown by an arrow and labeled r_{ij} . 20
- Figure 9 Schematic representation of nodes in elastic network of GNM. Every node is connected to its spatial neighbours by uniform springs. Distance vector between two nodes, i and j , shown by an arrow and labeled R_{ij} (left). Equilibrium positions of the i th and j th nodes, DR_i^0 and DR_j^0 , are shown in xyz coordinate system. R_{ij}^0 is the equilibrium distance between nodes i and j . Instantaneous fluctuation vectors, DR_i and DR_j , and instantaneous distance vector, R_{ij} , shown by the dashed arrows (right). 21
- Figure 10 Schematic representation of two coupled oscillators with equal mass connected by three springs of force constant with the edge points fixed. 22
- Figure 11 Simplified description of molecular dynamics algorithm. The simulation proceeds iteratively by alternatively calculating forces and solving the equations of motion based on the accelerations obtained from the new forces. In practise, almost all MD codes use much more complicated versions of the algorithm, including two steps (predictor and corrector) in solving the equations of motion and many additional steps for e.g. temperature and pressure control, analysis and output. 25
- Figure 12 Sketch of periodic boundary in two dimensions: The position of the particles in the central box are copied and placed in neighbouring boxes (left). When a molecule tries to leave the box during an MD simulation, it reappears at the opposite end of the box, so that the number of particles in the central box stays constant (right). 26
- Figure 13 Crystal structure of Alkaline Phosphatase as dimer (a.) and model of an isolated monomer (b.), based on the same crystal structure in tube representation and coloured by chains with the interface shown in space filling representation and coloured by the atom types (1ed9.pdb). 33

- Figure 14 Amino acid sequence of the AP along with the secondary structure elements. Residues comprising the interface are coloured in pink and the active site residues - in cyan. 34
- Figure 15 Last frame of the 100ns MD simulation of the monomer represented in tube and coloured by secondary structure elements with interface shown in cartoon representation and labeled as alpha helix (H) or a beta sheet (Sh). The flexible secondary structure elements comprising the “wing” region (solid circle) and the “crown” domain (dotted circle) are shown in new cartoon representation and labeled by the residue numbers. Terminal residues - K449 and T1 are shown in VDW representation and coloured by the residues type (blue for basic and green for polar residues respectively), while the N-terminal loop is shown in gray. Active site is depicted by the presence of the Zn1, Zn2 and Mg metal ions. 36
- Figure 16 First six normal modes of the apo dimer from GNM. Colouring is based on the mobility along the mode directions, red being most mobile and blue - least. The picture for the holo enzyme is essentially the same, and therefore not shown here. 42
- Figure 17 Cross-Correlation maps of the first six normal modes of the apo (a.) and the holo (b.) enzymes, as calculated from GNM. 42
- Figure 18 Square-fluctuations of the first six normal modes of the apo (a.) and the holo (b.) enzymes calculated from ANM. 43
- Figure 19 Cross-Correlation maps of the first 6 normal modes of the apo (a.) and the holo (b.) enzymes, calculated from ANM. 44
- Figure 20 First six normal modes of the holo dimer based on the ANM. Arrows indicate direction of the modes while colouring is based on the mobility along the mode directions, red being most mobile and blue least. Apo enzyme describes similar motions with small differences in the fifth and the sixth modes (not shown). 44

- Figure 21 Crystallographic B-factors per residue of (a.) the side and (b.) the main chain atoms of subunits A and B of the holo AP (1ed8.pdb) (left) and the apo AP (1ed9.pdb) (right). Comparison of the B-factor values of the subunits A (left) and B (right) of the apo enzyme with the respective subunits of the holo enzyme (c.). 46
- Figure 22 RMSF of the CA atoms of the (a.) subunits A and (b.) subunits B of the holo and apo enzymes, and (c.) subunit A of the apo enzyme vs subunit B of the holo enzyme, and (d.) visa versa. 48
- Figure 23 RMSF of the CA atoms of subunits A and B as calculated from the last 80ns of 100ns long MD simulations of the holo (a.) and apo (b.) enzymes; Comparison of the B-factor values calculated from MD simulations (red) versus those obtained from the experiment (blue) for subunits A (b.) and B (c.) of the holo (left) and apo (right) enzymes. 49
- Figure 24 Time evolution of the carbon alpha RMSD with respect to the starting structure of the holo (left) and apo (right) enzymes (a.) and their Rgyr (b.) as calculated for subunits A and B and the entire dimer (c.). Only the last 80ns of the 100ns long MD simulations is considered. 50
- Figure 25 Proposed pathway for the transfer of information on structural changes between subunits A and B, (coloured in blue and red respectively) via the N416->D555 hydrogen bond. R559 is located on the same interfacial helix as D555 and in the apo enzyme, forming another N416->R559 highly occupied hydrogen bond, which has a decreased occupancy in the holo enzyme. Hydrogen bonding network across the interface, may serve as means of communication between the subunits, transferring the information of product binding and release. 52
- Figure 26 Distance distribution plots of the key residues in the active site of the holo AP, averaged over three independent MD runs. PO₄(P)-Zn1 (a.), PO₄(P)-Zn2 (b.), PO₄(O₃)-Zn2 (c.), PO₄(O₃)-H370 (d.), and PO₄(O₃)-S102 (e.). 55

- Figure 27 Coordination structure of the Zn1 (a.), Zn2 (b.) and Mg (c.) metal ions as observed in the last frame of the 100ns MD simulation. Subunit A and subunit B are shown in then left and right hand-sides respectively. 57
- Figure 28 Time evolution of the key distances in the active sites of the subunits A and B of the holo enzyme, as calculated from the last 80ns of MD simulations. R166(NH2)-PI2(O11) (a.), R166(NH1)-PI2(O14) (b.), H331(NE2)-PI2(O12) (c.), H412(NE2)-PI2(O12) (d.), H370(NE2)-PI2(O13) (e.), S102(OG)-PI2(O13) (f.). 58
- Figure 29 Time evolution of the distances between (a.) Zn1 and (b.) Zn2 metal ions and the phosphorus atom of the inorganic phosphate. 59
- Figure 30 Time evolution of the distances between Zn1 and Zn2 (a.), Zn1 and Mg (b.) and Zn2 and Mg ions (c.) and the distance between Zn1 (d.) and Zn2 (e.) metal ions and the phosphorus atom of the inorganic phosphate. 59
- Figure 31 Distance distribution plots between the metal ions of the individual subunits of holo AP and averaged over three independent MD runs. Zn2-Mg (a.), Zn1-MG (b.), and Zn1-ZN2 (c.). 60
- Figure 32 Distance distribution plots of the intra-metallic distances of the holo and apo AP in the two subunits and averaged over three independent MD runs. Left: subunit A, and right: subunit B. Zn2-Mg (a.), ZN1-Mg (b.), and ZN1-ZN2 (c.). 62
- Figure 33 Time evolution of the CA atoms' RMSD with respect to the starting structure as calculated for: the entire WT dimer, the T59R mutant and the monomer (a.), residues at the interface of a single subunit (chain A) (b.), residues of the active site (c.). Dashed line indicates the fraction of the trajectories, not considered in the further analysis. 63
- Figure 34 Time evolution of the Rgyr as calculated for: the entire dimer and the mutant (a.), the single subunit (chain A) of the dimer and the mutant and the monomer (b.). Dashed line indicates the fraction of the trajectories, not considered in the further analysis. 64

- Figure 35 Time evolution of the CA atoms RMSD with respect to the starting structure as calculated from the two independent MD simulations of: the entire dimer and the monomer (a.), residues located at the interface of a single subunit (dimer A) of a dimer and the monomer (b.), the residues located at the active site of the dimer and the monomer (c.). Dashed line indicates the fraction of the trajectories not considered in the further analysis. 66
- Figure 36 RMSD time-trajectory of N-Terminal loop with respect to the equilibrated structure as calculated from two independent MD simulations. 67
- Figure 37 Time evolution of the Rgyr as calculated from the two independent MD simulations for the single subunit (dimer A) of the dimer and the monomer. Dashed line indicates the fraction of the trajectories not considered in the further analysis. 67
- Figure 38 Secondary structure elements of the most flexible residues, located at the interface, as a function of time, as calculated from the last 80ns of the dimer (a.), the mutant (b.) and the monomer (c.) simulations. H3 alpha helix corresponds to the residues: 425-430 and Sh3 beta sheet - residues: 431-434. Colour-coding is as follows: alpha helix - pink, beta sheet - yellow, coil - white, turn - green and 3-10 helix - blue. 68
- Figure 39 RMSF values of the WT dimer and the monomer, with labels at the interface of the monomeric subunit. Alpha helices are labeled as H1 (29-35), H2 (54-66), H3 (425-430) and H4 (435-447) while beta sheets are labeled as Sh1 (80-85), Sh2 (417-422) and Sh3 (431-434); Residues comprising the "wing" region and the "crown" domain are marked with solid and dashed circles respectively. Labels are coloured to match the Three-dimensional (3D) representation of the residues at the interface as in Figure 15. 69
- Figure 40 Secondary structure elements of the WT dimer as a function of time, as calculated from the last 80ns of the dimer simulation (chain A). Colour-coding is as follows: alpha helix - pink, beta sheet - yellow, coil - white, turn - green and 3-10 helix - blue. 70

- Figure 41 Secondary structure elements of the T59R mutant as a function of time, as calculated from the last 80ns of the mutant simulation (chain A). Colour-coding is as follows: alpha helix - pink, beta sheet - yellow, coil - white, turn - green and 3-10 helix - blue. 71
- Figure 42 Secondary structure elements of the residues, located at the interface, as a function of time, as calculated from the last 80ns of the monomer simulation. Colour-coding is as follows: alpha helix - pink, beta sheet - yellow, coil - white, turn - green and 3-10 helix - blue. 71
- Figure 43 RMSF values of the dimer, mutant and monomer, with labels at the interface of the monomeric sub-unit. Alpha helices are labeled as H1, H2, H3 and H4 while beta sheets are labeled as Sh1, Sh2 and Sh3. 72
- Figure 44 Number of hydrogen bonds as a function of time in the dimer, the mutant and the monomer calculated from 100ns MD simulations. 72
- Figure 45 Residue-residue interactions across the interface coloured by residue type: Light blue - positive (H,K,R), red - negative (D,E), green (S,T,N,Q) - neutral, gray (A,V,L,I,M) - aliphatic, purple (F,Y,W) - aromatic, orange (P,G) - Proline and Glycine, yellow C - cysteine. The joining lines are coloured light blue for hydrogen bonds and orange for non-bonded contacts. Diagram is generated only for the residues at the interface of each chain, contributing to the hydrogen bonding network, as defined by the PDBSum [44]. 74
- Figure 46 Occupancies of hydrogen bonds of the residues in the active sites (circled) of the dimer (red) and the monomer (green), as well as Hbonds at their (former) interfaces. Arrows point from hydrogen donors to hydrogen acceptors. Residues interacting with the backbone atoms are marked with "bb", while the side chain atoms are given in parentheses. Averages and standard deviations are calculated from 2 independent MD runs. 75

- Figure 47 Single subunit of the dimer with the active site and the Y402->D330 hydrogen bonds highlighted. Sh2 interfacial beta sheet is thicker and coloured in red while the “crown domain loop” is shown in green. Zn1 and the residues in its coordination sphere are shown in licorice representation and coloured by atom type. 76
- Figure 48 Distance distributions of the T402-D330 hydrogen bond calculated from the distances between the hydrogen atom of Y402 and its hydrogen bonding partner, Oxygen atom of D330. 76
- Figure 49 Coordination structure of (a.) Zn1, (b.) Zn2, and (c.) Mg metal ions from a typical MD simulation snapshot of the dimer (left) and the monomer (right). The ligands of the metal ions are shown in licorice representation and coloured by the atom type, whereas the oxygen atoms of the water molecules coordinating the metal ions are shown as red balls. The residues that are present within the coordination sphere of one MD run and absent in another run, or present in one snapshot and absent in others - are shown in blue. 79
- Figure 50 Intra-metallic distance distributions between Zn1 and Zn2 (a.), Zn1 and Mg (b.) and Zn2 and Mg ions (c.) and averaged over two independent MD runs. 80
- Figure 51 Distance distributions calculated from the distances between Zn1 and Zn2 ions and the corresponding atoms of their coordination shell residues: Zn1-D327(OD1) (a.), Zn1-H412(NE2) (b.), Zn1-H331(NE2) (c.), Zn2-H370(NE2) (d.), Zn2-D369(OD2) (e.), Zn2-D369(OD1) (f.), Zn2-S102(OG) (g.), Zn2-D51(OD1) (h.). 81
- Figure 52 Distance distributions calculated from the distances between Mg ion and the corresponding atoms of its coordination shell residues: Mg-D51(OD2) (a.), Mg-T155(OG1) (b.), Mg-E322(OE2) (c.). 83
- Figure 53 Time-dependent variation of the salt bridge between oxygen and nitrogen atoms of the interfacial residues D76 and R418 (a.), D434 and K114 (b.) and E66 and K73 (c.). 85

- Figure 54 Time-evolution of the inter-metallic distances as a function of simulation time (right) and corresponding distance distributions, calculated from 80ns trajectories of MD simulations. Zn1-Zn2 (a.), Mg-Zn1 (b.), Mg-Zn2 (c.). 86
- Figure 55 Time-trajectories of the centre of mass distances between subunit A and the residue S102 (a.), subunit A and residue R166 (b.) and the residue R166 and residue S102 (c.); residues R166 and D101(d.); Time-trajectories of the Chi1 (e.) and Chi2 (f.) dihedral angles of the residue R166. 88
- Figure 56 Distance distributions calculated from the distances between Zn1, Zn2 and Mg ions and the corresponding atoms of their coordination shell residues of the T59R mutant: Zn1-D327(OD1) (a.), Zn1-D327(OD2) (b.), Zn1-412(NE2) (c.), Zn1-331(NE2) (d.), Zn2-H370(NE2) (e.), Zn2-369(OD2) (f.). 89
- Figure 57 Zn1 metal ion coordination sphere in the last frame of the monomer (left) and the mutant (right) shown in licorice representation and coloured by the atom type. Amino acids that are in the proximity of the metal ions, but not directly coordinated to them are shown in blue. 90
- Figure 58 Cooperative binding mechanism of AP. 94
- Figure 59 Square-fluctuations of the first 6 normal modes of holo and apo enzymes calculated from GNM. 110
- Figure 60 Shapes of the first 6 normal modes of the holo and apo enzymes calculated from GNM. 110
- Figure 61 The mode array of the first 6 normal modes of the apo dimer based on the ANM. 111
- Figure 62 The mode array of the first 6 normal modes of the holo dimer based on the ANM. 112
- Figure 63 Rgyr as a function of time in the dimer (a.), the mutant (b.) and the monomer (c.) calculated from four individual, 20ns long chunks of the last 80ns MD simulations. 113
- Figure 64 RMSD as a function of time in the dimer (a.), the mutant (b.) and the monomer (c.) calculated from the four individual, 20ns long chunks of the last 80ns MD simulations. 114

- Figure 65 Time-evolution of the intra-metallic distances of the dimer (left) and the monomer (right) as a function of simulation time, calculated from the last 80ns of two independent MD trajectories. Zn1-Zn2 (a.), Mg-Zn1 (b.) and Mg-Zn2 (c.). 115
- Figure 66 Time-evolution of the distances between Zn1 ion and the corresponding atoms of its coordination shell residues, as calculated from the last 80ns MD trajectories of the dimer (left) and the monomer (right). Zn1-D327(OD1) (a.), Zn1-412(NE2) (b.) and Zn1-331(NE2) (c.). 116
- Figure 67 Time-evolution of the distances between Zn2 ion and the corresponding atoms of its coordination shell residues, as calculated from the last 80ns MD trajectories of the dimer (left) and the monomer (right). Zn2-369(OD1) (a.), Zn2-D369(OD2) (b.), Zn2-51(OD1) (c.), Zn2-S102(OG) (d.), Zn2-370(NE2) (e.). 117
- Figure 68 Time-evolution of the distances between Mg ion and the corresponding atoms of its coordination shell residues, as calculated from the last 80ns MD trajectories of the dimer (left) and the monomer (right). Mg-D51(OD1) (a.), Mg-T155(OG1) (b.) and Mg-E322(OE2) (c.). 118

LIST OF TABLES

Table 1	pKa values of protonizable residues. of the apo and holo structures of AP. 33	
Table 2	Summary of MD simulations. All simulations consist of an equilibration period of at least 0.5ns (during which the protein was restrained) followed by a production run of 100ns in which all atoms were free to move. 37	
Table 3	RMSD of the CA atoms along with the heavy atoms (in parenthesis), as calculated from the superimposition of the crystal structures of apo and holo enzymes and their individual subunits. 45	
Table 4	Hydrogen bond occupancies of the residues across the interface of the apo and holo dimers, as computed from the last 80ns of 100ns long MD simulations. Hydrogen bonds with highest differences in the occupancies are shown in bold. 51	
Table 5	Hydrogen bond occupancies of the residues in the active site of the subunits of the apo and holo dimers, as computed from the last 80ns of 100ns long MD simulations. Hbonds with highest differences in the occupancies are shown in bold. 53	
Table 6	Averages and standard deviations of RMSD and Rgyr calculated over all frames of 100ns MD simulations and RMSF and number of Hbonds calculated from the last 80ns of the dimer and the monomer. 61	
Table 7	Averages and standard deviations calculated over all frames of 100ns MD simulations of the dimer, the mutant and the monomer. 65	
Table 8	Hydrogen bond occupancies of the residues at and across the interface, computed over the simulation time of last 80ns for the dimer, the mutant and the monomer. 78	
Table 9	Ensemble average of the distances between metal ions and the coordinating amino acid residues along with respective standard deviations in Å. In all cases the simulation length was 100ns and only the last 80ns were used to calculate the averages. PDB refers to the coordinates of the crystal structure with ID: 1ED9. 84	

Table 10	Ensemble average of the distances between metal ions and the coordinating amino acid residues along with respective standard deviations in Å. In all cases the simulation length was 100ns and only the last 80ns were used to calculate the averages.	85
----------	--	----

ABBREVIATIONS

AP	Alkaline Phosphatase
PDB	Protein Data Bank
VMD	Visual Molecular Dynamics
MD	Molecular Dynamics
EDA	Essential Dynamics Analysis
PCA	Principle Component Analysis
NMA	Normal Mode Analysis
ENM	Elastic Network Model
GNM	Gaussian Network Model
ANM	Anisotropic Network Model
RMSD	Root Mean Square Deviation
3D	Three-dimensional
PBC	Periodic Boundary Conditions
PME	Particle Mesh Ewald
RMSF	Root Mean Square Fluctuation
Rgyr	Radius of Gyration
Hbond	Hydrogen Bond
NMR	Nuclear Magnetic Resonance
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
WT	Wild Type
SD	Steepest Descent
CG	Conjugate Gradient

Part I

INTRODUCTION

INTRODUCTION

1.1 MOTIVATION

Many physiologically important enzymes have a dimeric quaternary structure. In some dimeric enzymes, monomeric domains fold independently, before dimerisation, which is probably reflecting the evolutionary origins of enzymes as smaller, simpler, independent functional proteins. Monomeric counterparts of the existing enzymes, might have been prevalent in earlier organisms and only later, merged together into dimers [13]. There are many reasons explaining the selective advantage favouring the evolution of monomeric species into larger dimers; among them are functional requirements such as having multiple binding sites, small solvent exposed surface area and high stability against denaturation [30]. Remarkably, dimeric proteins tend to be symmetric, but the driving force behind symmetric association of the proteins, and its connection to their function, is not well understood [13, 30, 37].

E. coli Alkaline phosphatase, presented in this thesis, is a dimeric metalloenzyme, where a 2-fold axis of symmetry relates two subunits that are 32Å apart (Figure 1) [23]. AP is a hydrolase enzyme with a broad range of substrates and ability to hydrolyse a wide variety of phosphate containing compounds [56], at alkaline pH. Due to its catalytic efficiency, AP is widely used in molecular biology labs, especially for the removal of 5' phosphates from DNA and RNA prior to labelling with radioactive phosphate and for the DNA cloning experiments [81]. The catalytic functionality of AP is important under conditions of inorganic phosphate starvation, as this enzyme is responsible for the supply of vital inorganic phosphate, in biological systems [76].

Symmetry is often associated with cooperativity; i.e., stability and catalytic properties of one subunit are often controlled by the conformation of another. Interactions stabilising the dimeric enzymes include salt-bridges, hydrogen bonds and non-polar interactions across the interface [37]. Binding of a ligand to one subunit impairs subsequent binding of a ligand to another subunit [58, 36], resulting in negative cooperativity between the subunits of Alkaline Phosphatase [36]. It has been demonstrated experimentally, that in the course of the catalytic cycle the conformation of the subunits of AP alternates between two states (free enzyme and enzyme-substrate) in order to enable "time-regulated" substrate binding and product release [58]. However, the exact mechanism allowing two remote binding sites

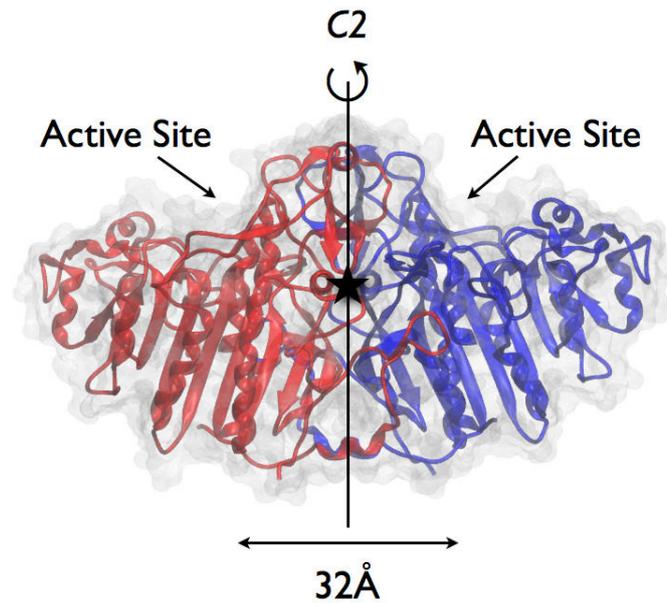


Figure 1: Crystal structure of a wild type Alkaline Phosphatase (1ed9.pdb) coloured by chains and in new cartoon representation showing the secondary structure elements in the foreground and in surface representation, showing the overall shape of the enzyme in the background. Metal ions in the active site are not shown for clarity and the black star placed at the axes of symmetry schematically depicts location of the mutated residues.

to communicate the information is not clear and the atomistic understanding of the residues participating in this information transfer remains to be completed.

Due to the fact that the structure and mechanism of *E. coli* AP is well studied and many wild type and mutant crystal structures are available, it is a good model for investigating subunit interactions in symmetric enzymes. Many homo-dimeric enzymes have been studied both experimentally and computationally during the last decades, however the debate on the conformational selection versus induced fit playing the main role in subunit cooperativity remains unanswered [21]. In any case, atomic level precision is required to understand allosteric changes upon ligand binding and the nature of the information transfer between the often remote binding sites of symmetric proteins. Fluctuations of the individual atoms and correlation between groups of atoms in dimeric enzymes, can be understood in light of the so called “functional allostery” encoded solely by the 3D structure of the protein, thereby linking structure dynamics and function [24]. In order to understand the mechanism of control and regulation of AP achieved by the subunits’ interactions, we constructed coarse grained and atomistic models and conducted a comparative local and global dynamics analysis of the individual subunits of the apo and holo enzyme with inorganic phosphate bound and not bound in the active

site, respectively. Our results show the influence of collective motions on the dynamic variation of key residues in the active site and explain the relationship between mechanical properties attributed to the dimeric architecture of the enzyme and its functionality.

Analyses of the MD simulations further suggest that correlated motions of the monomeric subunits of the dimer, as well as the local dynamics of key residues in the active sites, play an important role in the functionality of Alkaline Phosphatase. Although each subunit of the enzyme is equipped with its own catalytic sites, a monomeric AP does not exist in nature and the engineered mutants have significantly reduced activity. No crystal structure of the monomeric AP is available, and it is very difficult to characterise it experimentally, under optimal conditions, due to the fact that it dimerises spontaneously [52]. No previous computational studies exist, up to now, explaining structural and dynamic differences between monomer and dimer with atomic level precision. To understand and explain the dynamic properties of Alkaline Phosphatase, MD simulations are performed, on the WT dimer and the monomer that has been built using the crystal structure of the dimer. By analysing the structural variation of the key residues in the active site, and at the dimer interface, of Alkaline Phosphatase, we aim to gain insights of the driving force behind protein evolution and by doing so, illustrate the power and importance of dimerisation in Alkaline Phosphatase and related enzymes.

In addition, a T59R mutant is built, where interface residues T59 and T559 are substituted by bulky and charged Arg residues (Figure 1). Experimental evidence suggests that such a substitution destabilises the dimeric interface, resulting in a separation into isolated monomers, with reduced structural stability and catalytic activity. Our MD simulation results confirm that the overall dynamic behaviour of the monomer is different from that of the corresponding dimer and resembles more that of the T59R mutant. Analysis of MD simulations with a focus on structural stability, dynamics of the active site, and interfacial residues, helps to explain the reduced stability and activity of the mutant with respect to the native state dimer and makes the prediction of the alternative mutations, destabilising interface, possible. Such predictions open the door into the *in silico* design of novel dissociative inhibitors targeting the dimer interface, rather than the active site of an enzyme.

1.2 OUTLINE

The work described in this thesis is organised in seven parts.

The first part is an introduction to the thesis and the motivation for conducting it.

The second part provides background information about enzymes in general and a detailed literature overview of *E. coli* Alkaline Phosphatase and the work related to this thesis. The same part describes the field of biomolecular modelling and simulations and provides a theoretical background of computational methods used in connection to this thesis.

The third part is on materials and methods, describing the structural models, used in this thesis, and the detailed simulation protocol.

The results obtained using those computational tools, and their analysis are shown in the fourth part. It is further divided into two chapters. In the first chapter the holo and apo forms of Alkaline Phosphatase are compared. The second chapter presents the analysis of the three models of the apo enzyme: the wild type dimer, a model of monomer and the T59R mutant dimer.

The fifth part discusses the results obtained from the analysis of the apo and holo enzymes. The changes in the local dynamics associated with the binding of the product are illustrated and explained in this part. Furthermore, the structural and dynamic properties of the different models of the apo enzyme are discussed in the light of the importance of the dimeric quaternary structure for the enzyme's functionality.

The sixth part of the thesis concludes the results.

The seventh part discusses future directions of the work presented in the thesis.

Part II
BACKGROUND

BACKGROUND

2.1 STRUCTURE, DYNAMICS AND FUNCTION OF ENZYMES

Enzymes are not rigid molecules, instead, they have “dynamic personalities” expressed by constant changes in their configuration and the ability to perform a number of important functions inside the cell of all living organisms [34]. Enzymes catalyse (accelerate) range of biochemical reactions allowing the cell to grow and reproduce [28]. Enzymes, like other proteins, have unique primary structure defined by the order of amino acids, which in turn, is determined by the genes in the cells. The primary structure of enzymes, to a large extent, determines their 3D shape, as defined by the atomic coordinates, known as the tertiary structure. In addition to the tertiary structure of the single subunit, enzymes with multiple subunits arrange themselves into a quaternary structure [28]. Such a self-association of the small monomeric units of the enzymes, into larger oligomers, with often multiple active sites, is advantageous for the structural and thermal stabilities of the enzyme, as well as control and regulation of catalytic activity [13, 30, 75].

Regardless the fact that enzymes are typically large molecules, with hundreds of amino acids, only a small part of them, called the active site, participates directly in the catalysis of biochemical reactions. The amino acid side chains that form the active site are arranged to ensure the steric fit with the substrate and to interact specifically with the substrate in an attractive manner. Hence, there is a geometric and chemical complementarity between the active site of the enzyme and the biological substrate to be transformed. Specificity arises from these complementarity. as most enzymes are highly specific, acting on only a small number of substrates, while others, such as Alkaline Phosphatase presented in this thesis, are less specific in their actions [28].

Often, residues in the active site of a dimeric enzyme are located at the interface between the subunits, thereby ensuring that the enzyme is catalytically functional. A well studied example is the homodimeric enzyme, immunodeficiency virus type 1 (HIV-1) protease (PR) with two identical subunits, each contributing one catalytic Aspartic acid to define the active site, that simply does not exist in a single monomer [47]. MD study of HIV-1 PR dimer and monomer, showed the effect of dimerisation on the stability of the enzyme [47] and suggested no coupling between monomer folding and dimerisation [48]. In the homo-dimeric enzyme, *E. coli* Alkaline Phos-

phatase, the scenario is different: catalytically important residues are not shared between the subunits; yet the active monomer of this enzyme does not exist in nature, and the engineered mutants, lack structural stability and activity [15]. The widely accepted opinion about dimeric enzymes is that the overall structural stability and enzymatic activity is induced by dimerisation [15]; although exceptions contradicting this opinion exist, such as the isolated monomer of Glutamate Racemase (RacE), that is structurally more flexible and catalytically more active, than the native dimer [54].

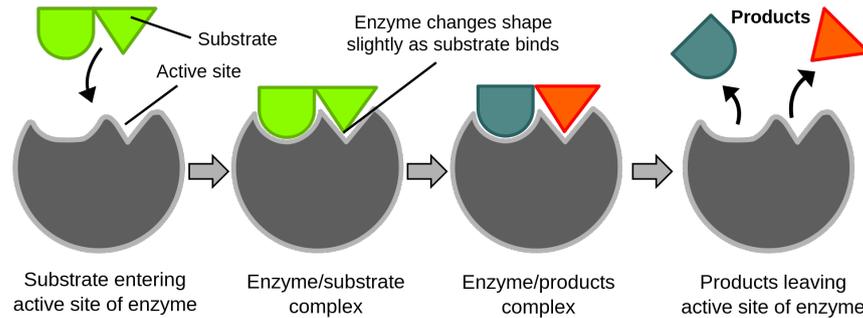


Figure 2: Diagrams to show the induced fit hypothesis of enzyme activity.

A quantitative theory of enzyme kinetics, originally proposed by Victor Henri in 1902 and further developed by Leonor Michaelis and Maud Menten, explains the enzymatic mechanism of binding to substrates and turning them into products via intermediate step of forming an enzyme-substrate (also called Michaelis) complex. Activity of many enzymes is assessed by the rate of enzymatic reactions, explained with the simple Michaelis-Menten equation, that relates reaction rate to the concentration of a substrate. However, the actual activity of enzymes is far more complicated and involves structural dynamics [28].

In 1894 a "lock and key" model was proposed by Emil Fischer to explain enzyme specificity in terms of enzyme substrate complementarity [25]. However, it has been debated that this model fails to explain the transition state stabilisation, in contrast to a more accurate, so called "hand in a glove" model suggested by Daniel Koshland in 1958 [43]. This is when the "dynamic personalities" of the enzymes emerged first, and the induced fit model proposed by Koshland soon became a key to understanding mechanism of the conformational change mediating "allosteric" interactions between different parts of the enzyme (Figure 2)¹ [21]. However the debate over the validity of the induced fit model continued for almost five decades. A number of experimental and computational studies favour an alternative

¹ Source: Created by Tim Vickers, public domain, via Wikimedia Commons, http://upload.wikimedia.org/wikipedia/commons/2/24/Induced_fit_diagram.svg

model, that assumes conformational selection of enzymes' preexisting conformations, upon binding to the ligand [21].

Number of dimeric enzymes with multiple binding sites, have been studied, both, experimentally and computationally. Binding to a ligand in one site can increase or decrease affinity of binding to the other sites, resulting in phenomena called positive or negative cooperativity, respectively. A well studied example of subunit cooperativity is a dimeric Lysyl-tRNA synthetase (LusU). Interestingly, MD simulations in conjunction to Principle Component Analysis (PCA) revealed functional asymmetry of LusU. This prediction was further tested and confirmed by a set of biochemical experiments [39, 38]. Flexibility of the LusU protein is manifested by domain rotation as a rigid body upon binding to the ligand, resulting in a subsequent breakage of the C₂ symmetry [39]. The Catabolite Activator Protein (CAP) is another example of a dimeric protein, with allostery leading to negative cooperativity. Essential Dynamics Analysis (EDA) were carried out for apo, singly liganded, and doubly liganded CAP, both, in the presence and absence of Deoxyribonucleic Acid (DNA) [50] resulting in an observation that "the system experienced a switch in motion" upon binding of a ligand [24, 50].

Binding to molecules can both decrease and increase the activity of enzymes. There are two types of such molecules, called effectors: inhibitors, that decrease enzyme activity and activators, that increase enzyme activity. Binding to effectors is often associated to the conformational changes called allosteric regulation [28].

Enzymatic reaction rates are millions of times faster than those of a comparable reaction in solution [63]. Activity of enzymes is affected by the environmental factors, such as temperature, pressure, pH, and the concentration of substrate. Enzymes typically speed up only handful of chemical reactions, and use a thermodynamically favourable reaction to "drive" a thermodynamically unfavourable ones without altering the chemical equilibrium itself and by catalysing the forward and backward reactions equally. The mechanism, that enzymes use for achieving such a spectacular rate enhancement is via lowering the activation energy, which is the energy required for a transformation of reactant into product [28]. In order to understand and quantify the rate of chemical reactions it is common to refer to transition state theory. For lowering the activation energy, enzymes are providing an environment in which the transition state is stabilised by electrostatic effects, that simply do not exist in the comparable uncatalysed reaction in solution [77]. Often, enzymes pursue a different reaction mechanism than the reaction in solution, which can be understood as providing a different (electrostatic) environment. However, for specific enzymes, so called "promoting vibrations" or "coupled promoting motions", that refer to equilibrium, thermally averaged conformational fluctuations, along the collective reaction co-

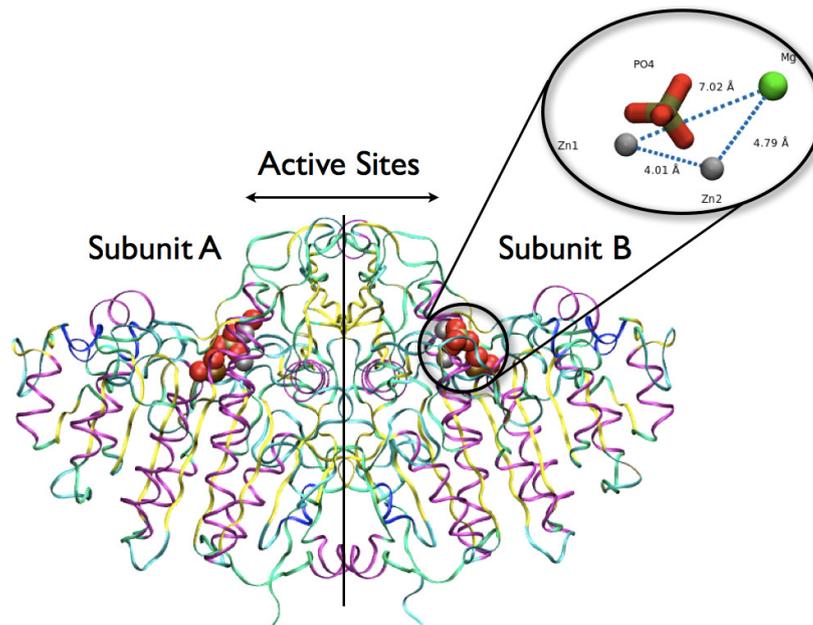


Figure 3: Crystal structure of Alkaline Phosphatase in ribbon representation (1ed8.pdb), coloured by secondary structure elements, with a close-up showing inorganic phosphate and metal ions in licorice representation and coloured by the atom type.

ordinate, that bring two reaction partners close together (near-attack conformation), play central role [2, 11].

2.2 ALKALINE PHOSPHATASE

The importance of phosphate transfer reactions in biological systems is reflected by the ubiquity of phosphate compounds. Phosphate containing compounds fulfil biologically important multiple roles, such as being essential intermediary metabolites, genetic materials, energy sources and reservoirs of biochemical energy [79]. Due to the low reaction rates of phosphate ester hydrolyses the use of catalysts is required. The super-family of enzymes called phosphatase has a biochemical machinery enabling it to hydrolyse phosphoesters rapidly under mild cellular conditions.

Alkaline Phosphatase catalysis hydrolysis of phosphate esters and transfer of free phosphate groups to supply biological systems with inorganic phosphate (Figure 4). It catalyses the cleavage of a phosphate groups from a variety of compounds, including Ribonucleic Acid (RNA) and DNA and the artificial substrates [81]. AP also catalysis transphosphorilation of phosphate esters, which is another biologically important function of AP together with hydrolysis. In the presence of a phosphate acceptors, enzyme AP transfers phosphate to alcohol to form another phosphate monoester, thereby recycling inorganic phosphate within the living cells. Free inorganic phosphate is

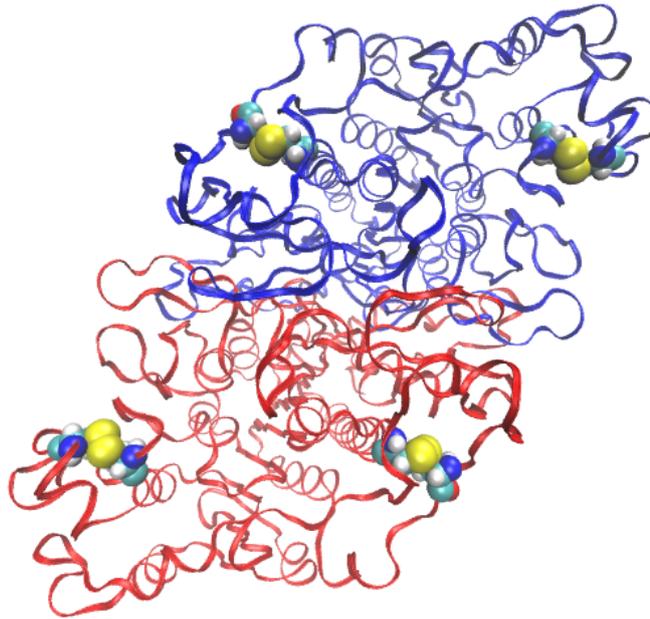


Figure 5: Crystal structure of AP (1ed9.pdb) in ribbon representation and coloured by chains, with disulphide bridges in space filling representation and coloured by atom type.

triangle with sides 3.9\AA (Zn1-Zn2), 4.9\AA (Zn2-Mg), and 7.1\AA (Zn1-Mg) lined by two active site amino acid residues: R166 and S102 [35, 23]. Metal ions play important role in the catalysis; Monomeric AP has reduced metal binding properties, and it was suggested that significant conformational changes occur in monomers upon dimerisation allowing binding of metals, thereby making active sites functional [15]. The loss of divalent ions leads to a monomerisation process in Alkaline Phosphatase [14].

Monomeric AP doesn't exist in nature. A number of experimental, mutational studies were conducted during the past decades, in order to isolate stable monomers by disrupting the interactions along the dimer interface [15, 52]. The interface of AP is comprised of a secondary structure of four α -helices and three β -strands [72]. Two subunits in the dimer are held together by salt bridges, hydrogen bonds and hydrophobic interactions (Figure 13) [42]. Substitution of a single amino acid at the interface can disrupt the dimeric interface, and also affect the stability of the monomer fold [52]. For example, mutation of T59 located at the interface, to a bulky and charged R59, produced a stable enough monomer to allow its characterisation under physiological conditions. Thermal stability, as well as catalytic activity of the mutant, was reduced; most importantly, there were almost undetectable amounts of tightly bound metal ions in the isolated monomer, supporting the idea, that metals are bound only to dimeric APs [52].

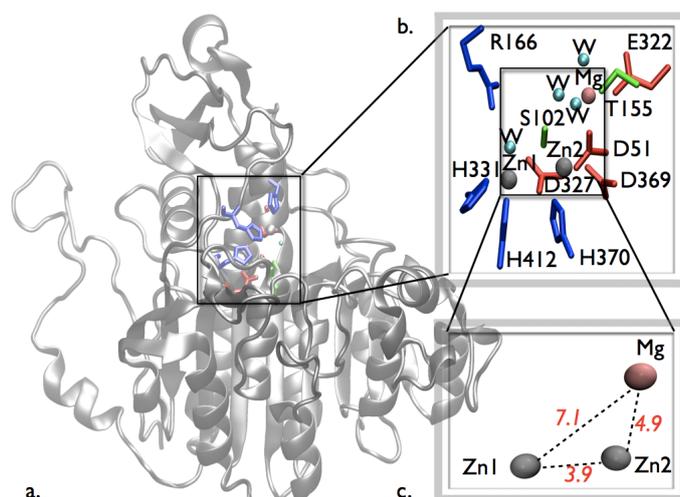


Figure 6: Chain A of a wild type dimer shown in a secondary structure representation and coloured in grey (a.). Close up of the active site with metal ions and their ligands, and the key amino acid residues shown in licorice and coloured by residue type: green - polar, red - acidic and blue - basic. Water molecules are shown as blue balls (b.). Dotted lines correspond to the distances between metal ions in Angstroms (c.).

The reaction mechanism of phosphate ester hydrolysis in AP, originally proposed by Kim *et. al* explains the metalloenzyme nature of this enzyme. In 1991, based on the crystal structure of AP, Kim and Wyck-off proposed a mechanism of phosphate monoester hydrolysis. The proposed reaction mechanism, involves two amino acid residues of an enzyme - R166 and S102, two zinc metal ions and one out of three water molecules in an active site [42]. Hydrogen bonds formed by the guanidinium group of R166 and two phosphate oxygens are maintained throughout the whole catalytic cycle. Two other oxygens of phosphate monoester coordinate Zn1 and Zn2 metal ions by forming a phosphate bridge between them. A phosphate ester is bound to the active site of S102 which leads to the phosphorylation of this residue and formation of a covalent phosphoseryl intermediate. Hence, in the first step of hydrolysis Serine performs a nucleophilic attack on the phosphorus atom. Subsequently, S102 is dephosphorylated to form noncovalent complex of the enzyme and inorganic phosphate. This is possible due to the presence of a water molecule, which is activated in the second step on the reaction and acts as a nucleophile. The final step is dissociation of phosphate which is the slowest, and therefore rate-determining step [42].

In 2000 a revised mechanism of hydrolysis was proposed by Stec *et al*. This mechanism was based on a high resolution crystal structures and claimed to explain the role of Mg ion, as well as mechanism of S102 deprotonation for nucleophylic attack on the phosphorus atom [73].

In the free enzyme, the phosphate-binding site is filled with 3 water molecules. A hydrogen bond is formed between the hydroxyl group of S102 and the Mg-coordinated hydroxide ion. Upon formation of the enzyme-substrate complex Mg-coordinated hydroxide ion acts as a general base and deprotonates S102 oxygen, enabling it for the nucleophilic attack on the phosphorus atom. In the second step, a nucleophilic hydroxide ion that is coordinated to Zn1, attacks the covalent phosphoserine intermediate. This results in formation of the non-covalent enzyme-phosphate complex. In the final step the water molecule coordinated to Mg acts as a general acid, which donates a proton to oxygen of S102. Hence, Mg first serves as a general base to generate S102 nucleophile, and later as a general acid that regenerates S102 hydroxyl group [42].

Three metal ions of AP form 3 different binding sites. If we zoom into those metal binding sites we can see the metal ligands, which are the amino acid residues that coordinate those metal ions. Zn1 binding site consists of H331 and H412 and an D327 residue. Both carboxyl oxygens of D327 are coordinated to Zn1 along with imidazole nitrogen atoms of histidines. In holo form of AP, Zn1 is also coordinated to one of the oxygens of phosphoester, thereby stabilising the leaving group. Zn2 binding site consists of D369 and D51, one of the carboxyl oxygens of those residues coordinate Zn2, whereas second carboxyl oxygen of D51 is coordinated to Mg ion. H370 is also Zn2 ligand and its nitrogen atom is coordinated to Zn2. Mg binding site contains D51, E322 and T155 to coordinate one of the carboxyl oxygens of each [23].

Each of the three metal ions are in one or another way involved in catalysis. The presence of the zinc ion in the active site is beneficial for the enzyme. It is usually coordinated by nitrogen, oxygen or sulphur atoms of the substrate or the surrounding molecules and acts as a good Lewis acid, thereby playing role of a useful catalytic agent in enzymatic reactions. Due to a flexible coordination geometry, of zinc metal ions, the enzyme is allowed to shift its conformation rapidly which enables it to perform biological reactions. In the active site of AP Zn is surrounded by four neighbouring atoms in a tetrahedral coordination geometry. The occupancy of the Zn1 binding site is an essential requirement for catalysis, whereas catalysis is greatly enhanced by the occupancy of Zn2 binding site [42]. It has been also shown that the presence of Mg in a third metal binding site plays an important role in catalysis. In particular Mg ion forms 6 coordinated bonds with neighbouring atoms; 3 out of 6 are the bonds with active site water molecules, and remaining 3 bonds are formed with amino acid ligands. Hence, Mg participates in creating octahedral geometry, thereby positioning the water molecule to take part in S102 protonation/deprotonation required for catalysis [73]. It was initially thought

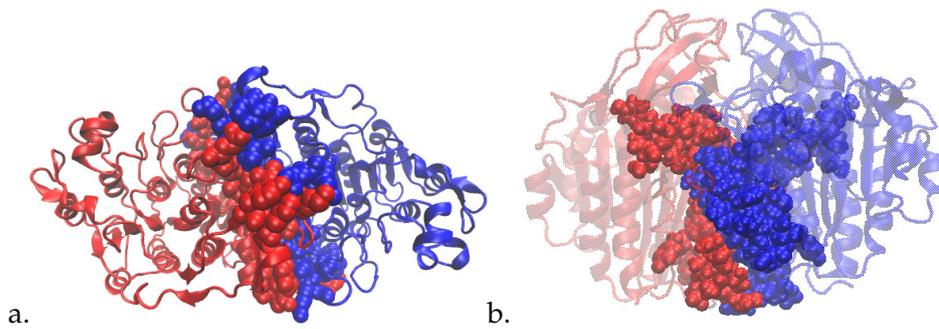


Figure 7: Structure of Alkaline Phosphatase coloured by chains and in new cartoon representation showing the secondary structure elements and the interface in space filling representation. (a.) The side view; (b.) The top view with the interface in foreground.

that Mg ions only contributes to catalysis by creating the electrostatic potential around the active centre [23].

A number of mutant APs were created to study the role of metal ions and the active site amino acid residues. Those studies are helpful in the revealing contribution of single residues in a catalytic mechanism and in design of alternative catalysts based on the structure of mutants. Mutation of R166 decreases substrate binding, thereby indicating the role of R166 in initial binding of substrate and in release of inorganic phosphate from non-covalent phosphate-enzyme complex. Based on those studies, active site nucleophile S102 is not required to increase the rate of hydrolysis. S102G and S102A mutants are still able to have catalytic activity, even-though no phosphate-enzyme intermediate is formed. This suggests the importance of metal ions, and the electrostatics provided by them for catalysis [23].

The geometry of active site metal ions is crucial not only for catalytic activity of the enzyme but also for the communication between the subunits. It has been proposed that binding of Mg increases the rate of conformational change and increases enzymes' catalytic activity by "a mechanism that involves interactions between the two subunits" [58]. In fact, communication between the subunits results in negative cooperativity, i.e., binding of a ligand to one subunit impairs subsequent binding of ligand to another subunit [58, 36]. The strong negative cooperativity of AP for the binding of inorganic phosphate and catalysis can play important regulatory function which is mediated by changes in the concentration of divalent Mg ions [36].

The "alternate-site" mechanism proposed by Boyer et al. [70] provides a general explanation of negative cooperativity; The binding of substrate to one subunit is coupled with more efficient catalysis of substrate or a product release in another subunit. Such a reaction mechanism, enables the enzymes to have much more "effective metabolic control with respect to time" [36]. This basic idea of ligand binding

to one subunit affecting the subsequent binding to the other subunit is going back to the induced-fit theory by Koshland et al. [43].

A more specific mechanism of subunit cooperation called the "flip-flop" model (or "half-of-the-sites" reactivity) has been proposed by Lazdunski et al. and applied to explain the Michaelis-Menten kinetics in AP [22]. In the course of the catalytic cycle the conformation of the sub-units of AP alternates between two states (free enzyme and enzyme-substrate) in order to enable time-regulated substrate binding and product release [58].

Number of dimeric structures have been scrutinised using the computational approaches at a different level of complexity and sophistication [24]. It has been demonstrated numerous times that the functionally important global motions are very well described by normal mode analysis of a few low-frequency modes [24] of an elastic network model, while analysis of MD simulations of an atomic description of a system is required to find the link between the local dynamics of the active site and enzymes' overall functionality [8].

2.3 BIOMOLECULAR MODELLING AND SIMULATIONS

The field of modelling and simulations of the biomolecules is a vibrant but rather young, roughly 40 years old [67]. Despite the young age, it has a global recognition as over the years, it proved itself useful in explaining number of experiments and even in predicting important results prior to experiments. The predictive power of modelling and simulations grows with technological advances and opens the perspective of understanding the ever growing number of biological molecules in ever tiny details, previously overlooked by the limitations of the experimental techniques. Cheap, fast and highly distributed computer hardware, in conjunction with open-source, robust and automated modelling and simulations software toolkits, has an immense potential to contribute to the advance of biology, medicine, and technology [67].

Molecular modelling and simulations is a field of science that aims to understand or predict the structure of the molecules and by doing so, simulate their behaviour *in silico* and explain their biological function using different computational techniques. There are number of computational approaches available for modelling and simulation of biomolecules. As famously noted by English statistician George E.P. Box in 1987:

"All models are wrong, but some are useful."²

Hence, depending on the size of the system of interest and the computational resources available, one can come up with a computational

² Empirical Model-Building and Response Surfaces (1987), co-authored with Norman R. Draper, p. 424, ISBN 0471810339

model, that is simple enough to be simulated with approximate computational methods in a reasonable time, yet robust enough to reproduce correctly experimental results without violating universal laws of physics. As notoriously written by Max Planck in his letter to R.W. Wood in 1931 [78]:

“It was an act of desperation. For six years I had struggled with the blackbody theory. I knew the problem was fundamental, and I knew the answer. I had to find a theoretical explanation at any cost, except for the inviolability of the two laws of thermodynamics.”

A number of computational techniques have been utilised successfully during the past decades to understand the fundamental relationship between structure, dynamics and function of biomolecules as large as enzymes. There is a tradeoff between increased accuracy and time of computation. Modelling and simulation of enzymes, with size ranging from tens to several thousand residues [16], can be computationally challenging. Thus, it is often useful to reduce the size of the system of interest by introducing a simplified, coarse-grained model. In such a model several atoms or residues are often lumped together as a single particle and the enzyme is represented as an interaction network of such particles. Despite the apparent simplicity, a coarse-grained model can perform remarkably well, reproducing the protein fluctuations with an accuracy comparable to the experiment and/or all-atom MD simulations [31]. This is the simplicity of coarse-grained models that results in a significant speed up of calculations of collective motions of the protein domains or even entire subunits [66]. Coarse-grained models are particularly useful for investigating the longer time- and length-scale dynamics that is often the key to understanding biological function of enzymes [8]. When the local dynamics of the protein is of interest, more accurate, atomistic classical molecular dynamics simulation [64, 74, 53] technique is used. In the nanosecond to microsecond time scale it is possible to observe side chain rotations and the motions of the loops, whereas smaller motions, such as bond vibration between two atoms, occurs at even shorter time scales (femto second to pico second), and hence quantum mechanical methods need to be used to take into account the electronic structure of the atoms [46].

2.4 ELASTIC NETWORK MODELS

Due to the fact that the functionally important motions of the enzymes occur on the long time scales, in order to capture such long scale motions, of the large parts of the protein, such as entire domains, representation of the protein is often coarse-grained. At the expense of the structural resolution, significant speed up of calcula-

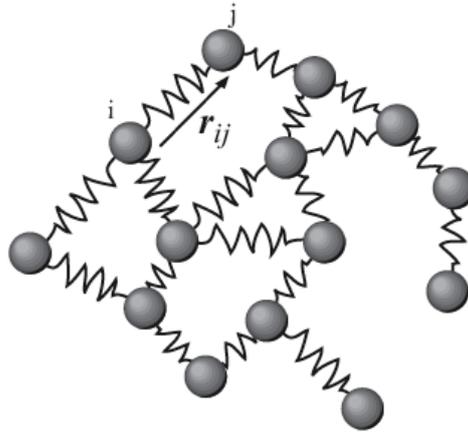


Figure 8: Schematic representation of nodes in elastic network model. Every node is connected to its spatial neighbours by uniform springs. Distance vector between two nodes, i and j , shown by an arrow and labeled r_{ij} .

tions is possible by the use of computational techniques such as normal mode analysis in connection to elastic network models [7, 75].

In elastic network models [4, 5] a protein structure is represented as a network of nodes and springs and the dynamics of the protein is described in terms of a collection of independent harmonic oscillators (Figure 8)³ [75]. This coarse-grained representation of a protein is based on the harmonic approximation of the potential energy and has the advantage that no prior minimisation is required and the initial structure (often crystal or NMR ensemble of structures) is assumed to be in minimum [75, 7].

Two most widely used ENMs: Gaussian Network Model (GNM) and Anisotropic Network Model (ANM) are discussed in the following subsections. In both models, the node positions are identified by the coordinates of $C\alpha$ and each residue pair with $C\alpha$ located within a specified cutoff distance is connected by a spring of uniform force constant. The difference between the two models is the description of the potential; the GNM potential includes contributions from both distance and orientation changes, whereas the ANM potential is exclusively based on distance changes [55]. In both ENM models types of interactions between nodes are not as important as their existence or absence defined by the network topology; those nodes that are close to each other are connected by uniform force constant, also called stiffness; it is the distribution of interactions that defines the collective motions and hence the structural dynamics [8].

³ Source: Created by Soumya Luthra, public domain, via Wikimedia Commons, http://upload.wikimedia.org/wikipedia/commons/c/c8/Elastic_network_model.png

2.4.1 Gaussian Network Model

The GNM [32] is a “mass-and-spring” elastic network representation of a biomolecule, that is using simplified harmonic potential (V_{GNM}) approximation to understand its long-scale dynamics (Figure 9)⁴.

$$V_{GNM} = \frac{\gamma}{2} \left[\sum_{i,j}^N (\Delta R_j - \Delta R_i)^2 \right] = \frac{\gamma}{2} \left[\sum_{i,j}^N \Delta R_i \Gamma_{ij} \Delta R_j \right]$$

where γ is a uniform force constant and $\Delta R_j, \Delta R_i$ are instantaneous fluctuation vectors and Γ is the Kirchhoff or connectivity matrix of inter-residue contacts. For r_c being a cutoff distance for spatial interactions between the residues the ij th element of the Kirchhoff matrix is defined by

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0, & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{j,j \neq i}^N \Gamma_{ij}, & \text{if } i = j \end{cases}$$

As the name suggests in the GNM, the probability distribution of all fluctuations is isotropic and Gaussian. Spatial interactions between the nodes of the network are defined by a cutoff distance, which is the only predetermined parameter in the model. Due to the fact that the GNM potential takes into account orientational deformations, as well as distance changes, expected residue fluctuations obtained from GNM is in very good agreement with the experimental equilibrium fluctuations (such as X-ray crystallography β -factor, for instance) [6].

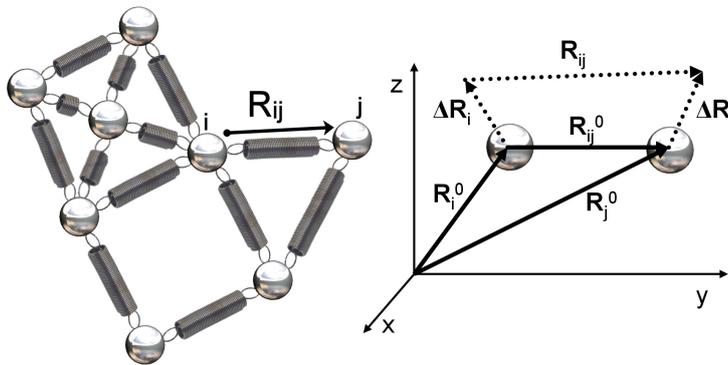


Figure 9: Schematic representation of nodes in elastic network of GNM. Every node is connected to its spatial neighbours by uniform springs. Distance vector between two nodes, i and j , shown by an arrow and labeled R_{ij} (left). Equilibrium positions of the i th and j th nodes, ΔR_i^0 and ΔR_j^0 , are shown in xyz coordinate system. R_{ij}^0 is the equilibrium distance between nodes i and j . Instantaneous fluctuation vectors, ΔR_i and ΔR_j , and instantaneous distance vector, R_{ij} , shown by the dashed arrows (right).

⁴ Source: Created by Ahmet Bakan, public domain, via Wikipedia, <http://upload.wikimedia.org/wikipedia/en/e/e7/MassSpringNetwork.jpg>

2.4.2 Anisotropic Network Models

ANM [4] is an extension of the Gaussian Network Model that accounts for directionality. In the context of proteins, ANM nodes often represent x,y,z coordinates of the $C\alpha$ atoms while the springs connecting the nodes represents interaction between two nodes, hence only one degree of freedom is needed to describe the internal motions of two atoms separated by a certain cutoff distance.

For any two nodes i and j , with equilibrium distance between the atoms represented by S_{ij}^0 and with the instantaneous distance S_{ij} , the harmonic potential is given by:

$$V_{ij(ANM)} = \frac{\gamma}{2}(S_{ij}-S_{ij}^0)^2,$$

where γ is a spring force constant. Diagonalization of the Hessian, which is $3N \times 3N$ matrix of second derivatives of the energy of a system with N number of atoms, gives anisotropic information regarding the orientation of nodes. Dynamics of the system can be understood in terms of a collection of independent oscillators, with low-frequency oscillators representing collective motions corresponding to large amplitude [75]. In theory, correlated motions of the groups of atoms are studied by constructing the covariance matrices to describe the mean-square deviations in atomic coordinates from their mean position (diagonal elements) or the correlations between their pairwise fluctuations (off-diagonal elements) [10].

2.4.3 Normal Mode Analysis

Typically, Normal Mode Analysis (NMA) [17, 19, 20] follows the time consuming procedure of the energy minimisation that can sometimes disrupt the structure of the protein [75], however in the context of ENM, no prior minimisation is needed as the starting structure is assumed to be in energetic minimum. Elastic network models are employed to characterise mechanical aspects of the protein's global dynamics. Large scale motions of the ENMs are easily accessible from 3D structure of a protein in its native state.

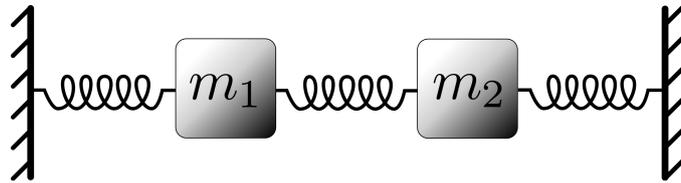


Figure 10: Schematic representation of two coupled oscillators with equal mass connected by three springs of force constant with the edge points fixed.

If we ignore the gravity, forces acting on the oscillators can be described by Hooke's law (Figure 10)⁵:

$$\vec{F}_1 = -k\vec{x}_1 + k(\vec{x}_2 - \vec{x}_1) = -2k\vec{x}_1 + k\vec{x}_2 \quad (1)$$

$$\vec{F}_2 = -k\vec{x}_2 + k(\vec{x}_1 - \vec{x}_2) = -2k\vec{x}_2 + k\vec{x}_1 \quad (2)$$

Horizontal displacement of the oscillators is described by:

$$\vec{x}(t) = Ae^{i\omega t} \quad (3)$$

Substituting (3) into (1) and (2) and $\vec{F} = m\frac{d^2\vec{x}}{dt^2}$, written in the form of the equations of motions we get the following:

$$-\omega^2 mA_1 e^{i\omega t} = -2kA_1 e^{i\omega t} + kA_2 e^{i\omega t} \quad (4)$$

$$-\omega^2 mA_2 e^{i\omega t} = -2kA_2 e^{i\omega t} + kA_1 e^{i\omega t} \quad (5)$$

By omitting common exponential factor we get the following matrix representation:

$$\begin{bmatrix} \omega^2 m - 2k & k \\ k & \omega^2 m - 2k \end{bmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 0 \quad (6)$$

For the equation to have non-trivial solution the determinant of the matrix must be equal to 0, so:

$$(\omega^2 m - 2k)^2 - k^2 = 0 \quad (7)$$

by solving (7) we get two solutions for the frequencies: $w_1 = \sqrt{\frac{k}{m}}$ and $w_2 = \sqrt{\frac{3k}{m}}$ these frequencies are the eigenvalues and (A_1, A_2) are the eigenvectors. If we substitute w_1 and w_2 to (6) we get (1,1) and (1,-1). First normal mode (8) corresponds to the masses moving in the same direction at the same time and second normal mode (9) - opposite, while centre of mass remains stationary:

$$\vec{r}_1 = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cos(\omega_1 t + \varphi_1) \quad (8)$$

$$\vec{r}_2 = c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} \cos(\omega_2 t + \varphi_2) \quad (9)$$

The general solution is a superposition of the normal modes where the coefficients C_1, C_2 , and the phases φ_1, φ_2 , are determined by the

⁵ Source: Created by Jim Belk, CC-BY-SA-3.0, via Wikimedia Commons, http://upload.wikimedia.org/wikipedia/commons/3/3f/Coupled_Harmonic_Oscillator.svg

initial conditions. General procedure of NMA in the context of GNM and ANM is summarised in the following steps:

- Set equilibrium positions (3D coordinates of a protein)
- Calculate the potential/forces for all atoms (Kirchhoff matrix for GNM and Hessian matrix for ANM)
- Diagonalise matrix to get eigenvectors/eigenvalues (as described above)
- Scale contribution of each mode with the inverse of that modes frequency
- Obtain expected values of fluctuations and cross-correlations from linear combinations along these normal modes

Diagonalization of the Kirchhoff and Hessian matrix decomposes the normal modes of collective motions of a biomolecule. The expected values of fluctuations and cross-correlations are obtained from linear combinations of fluctuations along these normal modes. Due to the scaling with the inverse of the mode frequency slow modes correspond to the low frequency and contribute most to the expected fluctuations. ENMs can yield a large number of modes, particularly $(n-1)$ for GNM and $(3n-6)$ for ANM, where n is the number of residues for a coarse-grained protein [82]. However, not all modes need to be analysed in detail, protein motions only along the few slowest modes, are shown to be collective and potentially relevant to functionality of the proteins [8]. These cooperative, low frequency motions are intrinsically accessible to protein through its 3D shape [7].

Despite their simplicity, coarse-grained elastic network models such as GNM and ANM have proven to be remarkably robust in predicting of global dynamics of biomolecules [6, 8, 7, 9]. The limitation of NMA is that it provides insight on harmonic motions only [31].

2.5 CLASSICAL MOLECULAR DYNAMICS

Classical molecular dynamics is a widely used simulations technique that provides the assessment of protein flexibility and dynamics on the nanosecond to microsecond time scale [33] and beyond, with special-purpose machine [68]. With the advance in development of biomolecular force field, computer technology and algorithms, MD has become a rapid and accurate theoretical framework to complement available experiment data [67] and even to spur the design of new experiments [29].

Essentially, Molecular dynamics is a computer simulation of physical movements of atoms, according to Newton's equations of motion, or as it has often been termed - "MD is statistical mechanics by numbers, on modern supercomputers" [67].

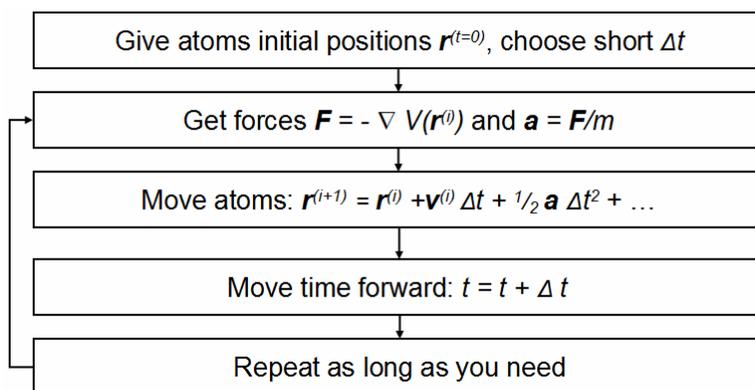


Figure 11: Simplified description of molecular dynamics algorithm. The simulation proceeds iteratively by alternatively calculating forces and solving the equations of motion based on the accelerations obtained from the new forces. In practise, almost all MD codes use much more complicated versions of the algorithm, including two steps (predictor and corrector) in solving the equations of motion and many additional steps for e.g. temperature and pressure control, analysis and output.

In MD, the potential energy function (which is a function of a particle coordinate) is often called a force field, as it represents the sum of the forces acting on the atoms within the system. A number of different force fields exist, where generally the potential energy is defined as the sum of bonded and non-bonded terms for describing: bond stretching, angle bending and bond rotations and Van der Waals and electrostatic interactions respectively [46].

For large systems such as proteins the time evolution of the coordinates are integrated by simplistic methods such as Leap-Frog and Verlet algorithms, and the time series of positions and velocities is called a trajectory [46]. The initial set of particle coordinates of a biomolecule are often available from X-ray crystallography or NMR spectroscopy, while initial velocities are often drawn from a Maxwell-Boltzmann distribution at a physiological temperature. Microscopic parameters are predicted deterministically at any time point in the future or the past (Figure 11)⁶. The laws of statistical thermodynamics are employed to convert microscopic parameters into macroscopic observables, based on the ergodic hypothesis, according to which: “if one allows the system to evolve in time indefinitely, that system will eventually pass through all possible states” [33, 46, 65]. That is why MD simulation has to be long enough to provide sufficient ensemble sampling.

Microcanonical, or NVE ensemble, where number of particles (N), volume (V) and energy (E) are constant, corresponds to an adiabatic process with no heat exchange. Canonical ensemble, or NVT, where

⁶ Source: Created by Kai Nordlund, public domain, via Wikimedia Commons, <http://upload.wikimedia.org/wikipedia/commons/b/be/Mdalgorithm.PNG>

number of particles (N), volume (V) and temperature (T) are conserved the energy of endothermic and exothermic processes is exchanged with a thermostat. A variety of thermostat methods are available to control temperature including velocity rescaling, the Nosé-Hoover thermostat [27] and Langevin dynamics [46]. Isothermal–isobaric ensemble, or NPT, where number of particles (N), pressure (P) and temperature (T) are conserved, in addition to a thermostat, a barostat is also needed [46].

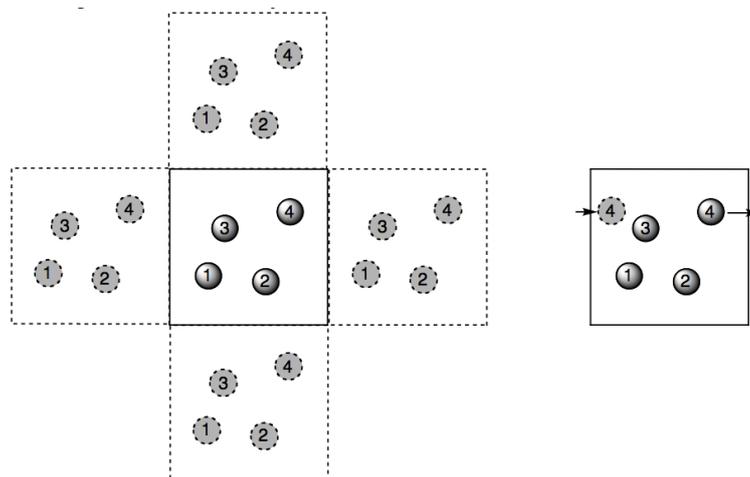


Figure 12: Sketch of periodic boundary in two dimensions: The position of the particles in the central box are copied and placed in neighbouring boxes (left). When a molecule tries to leave the box during an MD simulation, it reappears at the opposite end of the box, so that the number of particles in the central box stays constant (right).

Number of approximations and clever algorithms are often employed to match the time scale of interesting biological process, with MD simulations time and available computational resources. Choosing the integration time-step that is “smaller than the fastest vibrational frequency in the system”, is possible by using SHAKE algorithm, which fixes the vibrations of the hydrogen atoms and allows time step in the order of 2fs [65]. A fast approach to taking the solvent effects into account is an implicit water model that makes use of an extra potential on the solvent accessible surface. Slower, but more realistic is an explicit water model, such as TIP3P, however, boundary conditions must be imposed to limit the number of solvent molecules being calculated [41]. The most time consuming part of MD simulations is the evaluation of the non-bonded interactions and to overcome this problem, interactions between atoms further apart than a selected cut-off are neglected and the Particle Mesh Ewald (PME) summation is often used to account for long range electrostatic interactions when using Periodic Boundary Conditions (PBC) [60]. PBC are chosen in a way that the central unit cell containing the protein in a water box, is surrounded by neighbouring cells so that one side of the simulation

loops back to the opposite side (Figure 12)⁷. PBC minimise surface effects and preserve a constant number of atoms in each unit cell thereby mimicking a bulk phase [41, 46, 65].

Before performing MD simulations it is often recommended to optimise experimental or model derived structures using different minimisation algorithms such as Steepest Descent (SD), during the first steps of minimisation, followed by more accurate, Conjugate Gradient (CG) algorithm that in contrast to SD effectively reaches desired local minima. The energy minimised structure is solvated, heated, to obtain desired temperature, equilibrated, and eventually subjected to a long MD run in a production dynamics phase, to generate trajectories that are reflecting conformational changes over time [33, 65, 46]. Quality of Simulations can be significantly affected by two main factors: inaccuracy in the models, and insufficient sampling; Performing, multiple, independent simulations, can help in overcoming these limitations [31].

⁷ Source: Created by Jan Jensen [41], <http://www.flickr.com/photos/molmodbasics/4152853877/lightbox/>

Part III

MATERIALS AND METHODS

MATERIALS AND METHODS

3.1 MATERIALS

This chapter lists the materials, in the form of the files containing nuclear coordinates of Alkaline Phosphatase, that have been downloaded from the Protein Data Bank (PDB) [12], and used as initial structures in the simulations. Also among the materials is the file containing the sequence of *E. coli* Alkaline Phosphatase in so called “fasta” format. Two crystal structures of AP, at 1.75Å resolution each, are selected for the study presented in this thesis:

1. An apo structure without the inorganic phosphate bound to the active site - PDB ID: 1ED9
2. A holo structure with the inorganic phosphate bound to the active site - PDB ID: 1ED8

3.2 METHODS

The following web servers have been used in the study:

- PDBSum - a database that provides an overview of some properties obtained from the analysis of 3D coordinates of the crystal structures deposited in the PDB [44]. The server is used to identify residues at the interface between the subunits that form the hydrogen bonds or salt-bridges across the interface.
- POLYVIEW - a protein structure visualisation server [62]. The server is used to visualise primary structure of AP and to show the amino acid sequence, as well as secondary structure elements, comprising the interface and the active sites.
- PROPKA - a method for prediction and rationalisation of pKa values [49]. The server PROPKA3.1 is used to predict pKa values of the protonizable residues.

The following modelling, simulation and analysis software packages have been used in the study:

- ProDy - a Python package for analysis and modelling of protein structural dynamics [10]. The program is used to perform the coarse-grained modelling using ENMs.

- CHARMM - a molecular simulation program [18]. An empirical CHARMM22 force field with customised topology and parameter files is used for building and initial energy minimisation of the atomistic models.
- VMD - a Visual Molecular Dynamics (VMD), a molecular graphics software [40]. The program is used for the atomistic modelling, the analysis of MD simulations, and for rendering high resolution graphical images.
- NAMD - a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems [61]. All classical MD simulations presented in this thesis are performed using NAMD2.8.

3.3 STRUCTURAL MODELS

3.3.1 *Coarse-grained models*

Coarse-grained models of AP are built using two different elastic network models: GNM and ANM, both as implemented in the ProDy package [10]. The α -carbon atoms of AP are chosen as the nodes of the network; these nodes are connected by springs of uniform force constant, if they are located within a cutoff distance 10\AA and 15\AA for GNM and ANM respectively. NMA is performed and total of six slowest normal modes, further examined, for the apo and holo enzyme and compared with each other. Molecular graphics showing fluctuations along the selected modes is produced using VMD plugin NMWiz [10].

3.3.2 *Atomistic models*

Coordinates for the starting structure of a wild type AP, free (1ed9.pdb) and phosphate bound (1ed8.pdb) forms, both crystallised at pH 7.5, are used for building the structural models using VMD [40]. There are two chains, chain A and chain B, one magnesium and two zinc ions per chain, and two sulphate ions that are approximately 20 to 25\AA away from the active site. Holo AP also has one inorganic phosphate bound into each active site. There are a total of 616 crystal water oxygen atoms in the apo enzyme; this is 12 oxygen atoms more than the number found in the holo form (604), possibly, due to the fact that in the case of a holo enzyme, inorganic phosphate occupies the place previously occupied by water molecules. In the atomistic models, zinc and magnesium ions, inorganic phosphate and all crystal water molecules are kept, while inorganic sulphate is removed.

Residues with multiple protonation states are edited individually. pKa values of the amino acids are determined using the server PROPKA; the effect of ligands in the pKa prediction is also included [49], which

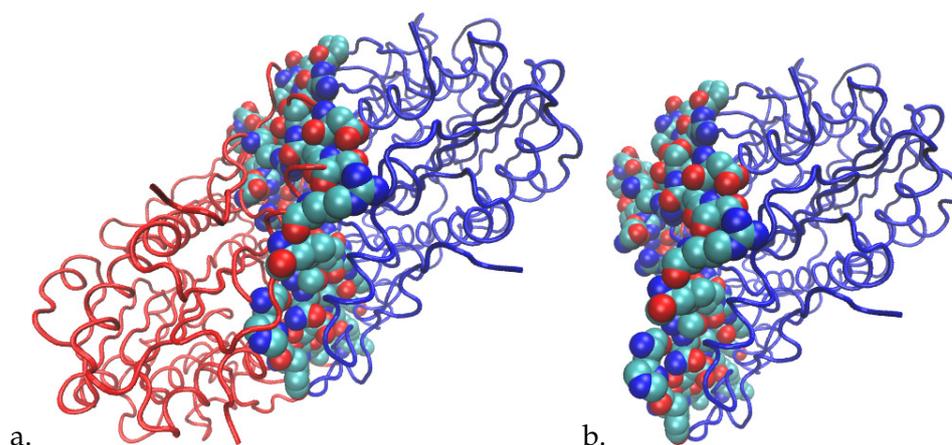


Figure 13: Crystal structure of Alkaline Phosphatase as dimer (a.) and model of an isolated monomer (b.), based on the same crystal structure in tube representation and coloured by chains with the interface shown in space filling representation and coloured by the atom types (1ed9.pdb).

Table 1: pKa values of protonizable residues. of the apo and holo structures of AP.

Amino Acid	Residue Number	Chain	Apo	Holo
			pKa	
H	162	A	8.24	8.24
D	330	A	11.85	12.02
E	341	A	8.36	8.38
H	662	B	7.85	7.85
D	830	B	11.60	11.83
E	841	B	8.32	8.33

is reflected by a slightly different pKa values for the apo and holo enzymes. Based on the prediction at experimentally consistent pH 7.5, all Lysine side chains are positively charged; pKa values of Histidine, Aspartic and Glutamic Acids, that are greater than pH 7.5 are given in Table 1.

There are a total of 20 histidine residues in dimer AP; The protonation states of all of them are observed individually, based on the distance measurements of surrounding residues. For instance, to decide the protonation state of H162 pKa=8.24 the distance is measured between the Histidine imidazole ring nitrogen atoms and all atoms of their immediate surrounding within a cut-off of 3Å. There is a negatively charged E134 residue near H162, which suggests that it is likely for one of the carboxyl oxygens of the Glutamate side chain to form hydrogen bond with the protonated Histidine. Hence, there is an electron donor near the NE2 atom of H162, while there is no

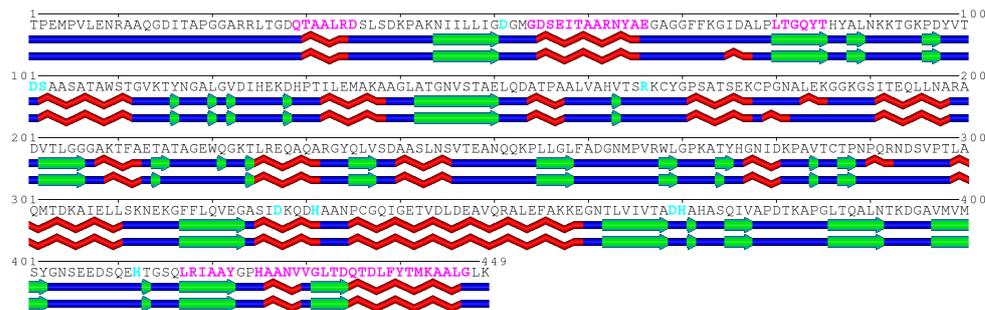


Figure 14: Amino acid sequence of the AP along with the secondary structure elements. Residues comprising the interface are coloured in pink and the active site residues - in cyan.

electron donor near the second nitrogen atom of H162 (ND₁). The same is true for H662 in chain B. We assume that these Histidine residues are singly protonated. There is not a single Histidine that is assumed to be (doubly) protonated. It is likely for the Zn-bound Histidine residues located in the active sites of chains A and B, H331(A), H370(A), H412(A) and H831(B), H870(B), H912(B) to be neutral with the proton located at the nitrogen atom that points away from the metal ion.

Patches are applied to the terminal residues and to the following disulphide bonds C168(A) - C178(A), C286(A) - C336(A) and C668(B) - C678(B), C786(B) - C836(B). Customised topology and parameter files are used for the protein residues, while dummy atoms are introduced, in order to keep a tetrahedral coordination of the zinc ions [59]. The initial geometry of the dummy atoms is built manually for the pdb files of each model. The zinc ions, with their dummy atoms, are kept fixed during the initial minimisation of a solvated and ionised system; this resulted in the energy minimised system with tetrahedral configuration of zinc ions along with dummy atoms facing towards the atoms of the coordination sphere. In a subsequent minimisation and short equilibration using PBC, the zinc ions and their dummy atoms are no longer fixed.

A model of a wild type dimer is referred to as “dimer”, whereas individual chains of a dimer are referred to as “dimer A” and “dimer B”. When the dimer is bound to the product, it is called “holo” and when it is free and has no product bound to its active site, it is called “apo”. A model of a monomeric AP is called “monomer”; A model of the T59R mutant is called “mutant”. In this thesis, different models are compared with each other; in particular apo dimer (often only its chain A) with the monomer and the mutant (and/or chain A of the mutant); holo dimer is compared with the apo dimer (both chains).

A model of the T59R mutant is built based on the structure of the wild type apo dimer, where the Threonine residues located at the monomer-monomer interface are mutated to Arginine residues. To find an optimal conformation of the bulky residues, R59 and R559,

they are kept fixed during the minimisation, heating and equilibration stages; and are free to move only during the subsequent 100ns production dynamics.

Monomeric AP is constructed using a definition of the interface based on the comparison of the accessibility surface area of the dimer with that of the separate monomers according to Sowadski et al. [72]. The residues comprising the interface are identified based on their numbers and are referred to as "H" if they comprise a helix and as "Sh", if they comprise a beta-sheet (Figure 15). There are in total four helices and three beta-sheets at the interface comprising the following residues (Figure 14):

HELICES:

H1: Q29, T30, A31, A32, L33, R34, D35

H2: G54, D55, S56, E57, I58, T9, A60, A61, R62, N63, Y64, A65, E66

H3: H425, A426, A427, N428, V429, V430

H4: Q435, T436, D437, L438, F439, Y440, T441, M442, K443, A444, A445, L446, G447

SHEETS:

SH1: L80, T81, G82, Q83, Y84, T85

SH2: L417, R418, I419, A420, A421, Y422

SH3: G431, L432, T433, D434

During the model building of a monomer all amino acid residues of one subunit (chain A of 1ed9.pdb) are kept intact, while all residues of the second subunit (chain B of 1ed9.pdb) are deleted (Figure 13). Residues at the (former) interface of the monomer are "kept frozen" during minimisation, heating and equilibration runs. A force constant corresponding to $1.00\text{kcal}/(\text{mol}\text{\AA}^2)$ is used to restrain the interface atoms during the first 2ns of the production run, and decreased to $0.5\text{kcal}/(\text{mol}\text{\AA}^2)$ during the following 2ns simulation time, and decreased further to $0.01\text{kcal}/(\text{mol}\text{\AA}^2)$ for another 2ns of production run. Only after 6ns of harmonically restrained simulation time of a monomer, restraints are released and simulations extended up to 100ns. Independent MD simulations of the same equilibrated structure with randomly assigned initial velocities, are launched for the dimer and the monomer; four independent MD runs are used for the comparison of the dynamic properties of the dimer and the monomer, reaching the total of 400ns simulation time. Only one MD run of a mutant is used for the comparative analysis of the dimer, the mutant and the monomer. Finally, three independent MD runs, of 100ns each are used for the local dynamics analysis of the apo and holo forms of the enzyme, reaching 600ns simulation time. The total simulation time of the results presented in this thesis is $1.1\mu\text{s}$.

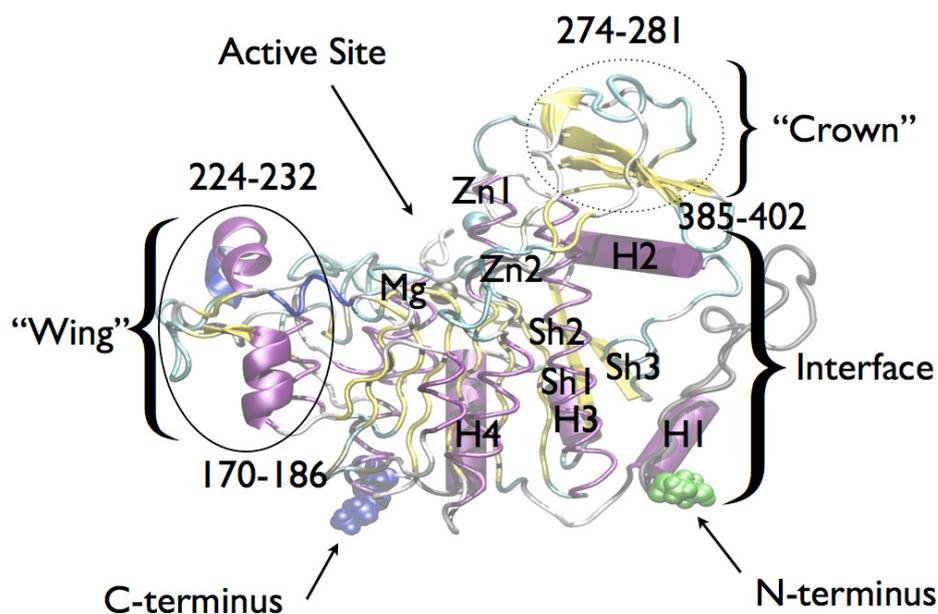


Figure 15: Last frame of the 100ns MD simulation of the monomer represented in tube and coloured by secondary structure elements with interface shown in cartoon representation and labeled as alpha helix (H) or a beta sheet (Sh). The flexible secondary structure elements comprising the "wing" region (solid circle) and the "crown" domain (dotted circle) are shown in new cartoon representation and labeled by the residue numbers. Terminal residues - K449 and T1 are shown in VDW representation and coloured by the residues type (blue for basic and green for polar residues respectively), while the N-terminal loop is shown in gray. Active site is depicted by the presence of the Zn1, Zn2 and Mg metal ions.

Table 2: Summary of MD simulations. All simulations consist of an equilibration period of at least 0.5ns (during which the protein was restrained) followed by a production run of 100ns in which all atoms were free to move.

Simulation	TIP3P	Total	Box (\AA^3)	Counter ions
Apo Dimer	58640	71891	{76 x 79 x 118}	65(Na^+) and 55(Cl^-)
Monomer	37695	44335	{81 x 74 x 71}	40(Na^+) and 35(Cl^-)
Mutant	57080	70347	{76 x 83 x 108}	62(Na^+) and 54(Cl^-)
Holo Dimer	58762	71893	{78 x 80 x 113}	67(Na^+) and 55(Cl^-)

3.4 MOLECULAR DYNAMICS SIMULATIONS

All simulations are performed using programs CHARMM [18] and NAMD2.8 [61] with an empirical CHARMM22 force field. Initial structures are prepared by the “autopsf” extension of VMD [40] to create a valid psf files for each model. By using “solvate” and “autoionize” plugins of VMD, solvated and neutralised systems are built and pre-minimized within VMD for 1000 steps of conjugate gradient minimisation. Each system is neutralised by adding Cl^- and Na^+ counter ions, corresponding to a physiological concentration of 150mM sodium chloride (NaCl). Each system is solvated in an orthorhombic water box filled with TIP3P water molecules; The distance from solute to the edge of the solvent box is 10\AA . At first, the solvated system is minimised with the protein harmonically restrained for 5000 steps of conjugate gradient minimisation, followed by 1000 steps of unconstrained minimisation and gradual heating for 30ps, from 0K to 298K with 1K temperature step.

Each system is equilibrated in three steps of 25ps each, resulting in 75ps of equilibration run with the numbers of particles, pressure (1bar) and temperature kept constant (NPT ensemble). The harmonic restraints are gradually lifted (to 0.5, 0.25 and $0.05\text{kcal}/(\text{mol}\text{\AA}^2)$) in order to introduce slow equilibration. In the first 25ps, velocities are rescaled every 0.1ps, and in the second 25ps, Langevin dynamics is used to maintain constant temperature. Pressure control is introduced in the third 25ps and in the production run using the Nosé–Hoover Langevin piston with a decay period of 500fs. NPT production run is performed for 100ns with integration time step of 2fs and coordinates are saved with a sampling interval of 2ps. All simulations are performed using PBC and PME algorithm for the long-range electrostatic interactions [26]. Non-bonded cut-off of 16\AA is applied and the short range electrostatics and van der Waals interactions are truncated at 12\AA using a switch function starting at 8\AA . The SHAKE algorithm is applied to all bonds involving hydrogen atoms [27] to allow a time step of 2fs.

3.5 ANALYSIS OF MD SIMULATIONS

Analysis of trajectories is performed for coordinates saved every 50ps. The convergence is monitored by plotting Radius of Gyration (Rgyr) and Root Mean Square Deviation (RMSD) as a function of the entire simulation time. The first 20ns of production run, is assumed to be the equilibration stage and the following analyses is carried out on the last 80ns using Hydrogen Bond (Hbond)s (to calculate the hydrogen bonds) and Timeline (to calculate and display a two-dimensional graph of secondary structure elements as a function of simulation time), plugins of VMD, as well as a number of homemade bash and tcl scripts. Criteria for formation of hydrogen bonds is that the distance between hydrogen donor (D) and hydrogen acceptor (A) atoms is less than the cut-off distance of 3.2\AA and the angle D-H-A is less than the cut-off angle of 42° . The average occupancy of the hydrogen bonds is calculated as a ratio between the number of the trajectory frames with the Hbond of interest formed and the total number of frames. Distance distribution plots are made by first finding the maximum and minimum values of the distance for each trajectory, then stepping over the range of distances using a bin size of 0.01\AA . Mean values and the sample standard deviations are calculated over the independent MD runs. VMD is used for rendering images of 3D structural models [40] and Gnuplot for plotting the data files [80]. NMA is performed using ProDy [10] and the corresponding molecular graphics is produced using VMD plugin, NMWiz [10].

Part IV
RESULTS

RESULTS

4.1 GLOBAL DYNAMICS

4.1.1 *Gaussian Network Model*

To understand the intrinsic dynamics of AP as dictated by its 3D structure, we employed GNM, to build a coarse-grained representation of holo and apo enzymes using their respective X-ray structures. A cut-off of 10Å was used to define the distance separation between pairs of nodes (CA atoms) in a Gaussian Network, interacting with each other with a uniform force constant. These two parameters were used to construct the Kirchhoff matrix and analyse individual normal modes as described in chapter 3.

GNM shows a very high overlap score for the first six modes of the apo and holo enzymes, suggesting nearly identical global dynamics of these two forms, dictated solely by the shape and architecture of the enzyme and independent from the binding to the product (Figure 16).

Large values of the square-fluctuations are observed in the terminal residues, and the residues of the so called “wing” region and the “crown” domain (Figure 59). The pairwise cross-correlations of the fluctuations of residues in the holo and apo dimer are almost identical and both show that the flexible regions are also highly correlated. For instance, residues at the wing region (170 to 230) are positively correlated in the subunits A and B of both, apo and holo enzymes. Similarly, motions of interfacial residues are positively correlated with the crown domain and negatively correlated with the wing region (Figure 17).

4.1.2 *Anisotropic Network Model*

The ANM representation of a holo and apo enzymes is performed using a cut-off of 15Å is used; this cut-off defines the distance separation between pairs of nodes (CA atoms) interacting with each other with a uniform force constant. Using these two parameters the Hessian matrix is constructed and the first six normal modes analysed (see details in chapter 3).

Interestingly, the first four normal modes of the apo and holo enzyme are highly correlated, the fifth mode shows a smaller correla-

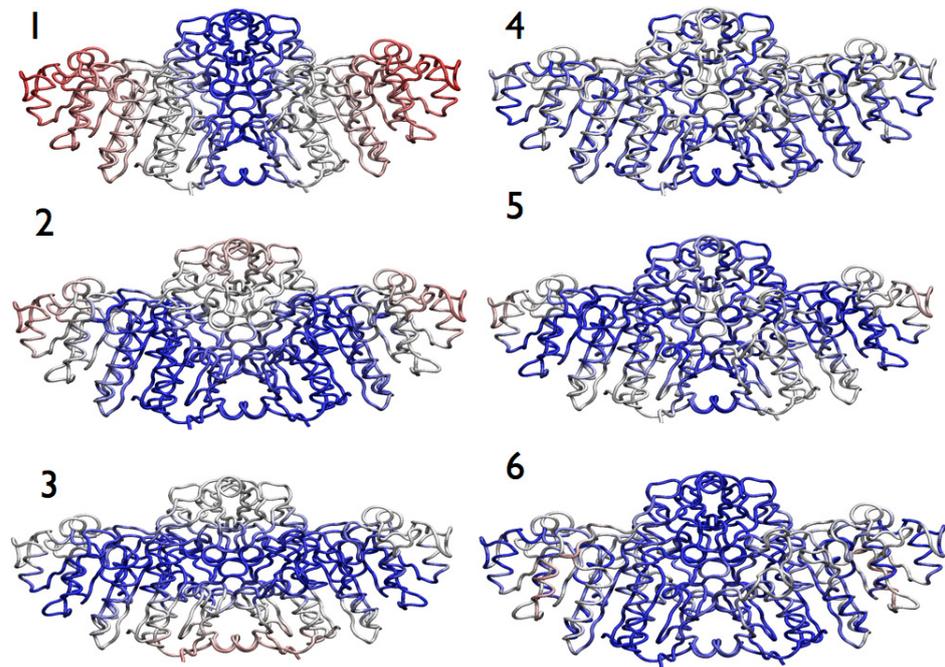


Figure 16: First six normal modes of the apo dimer from GNM. Colouring is based on the mobility along the mode directions, red being most mobile and blue - least. The picture for the holo enzyme is essentially the same, and therefore not shown here.

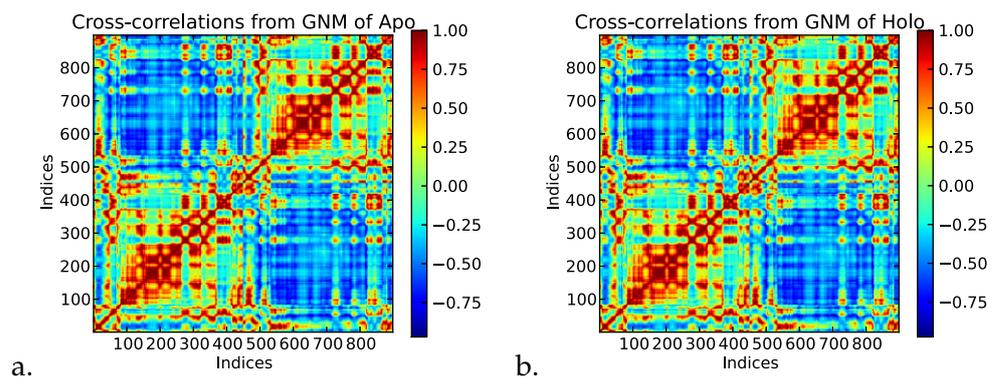


Figure 17: Cross-Correlation maps of the first six normal modes of the apo (a.) and the holo (b.) enzymes, as calculated from GNM.

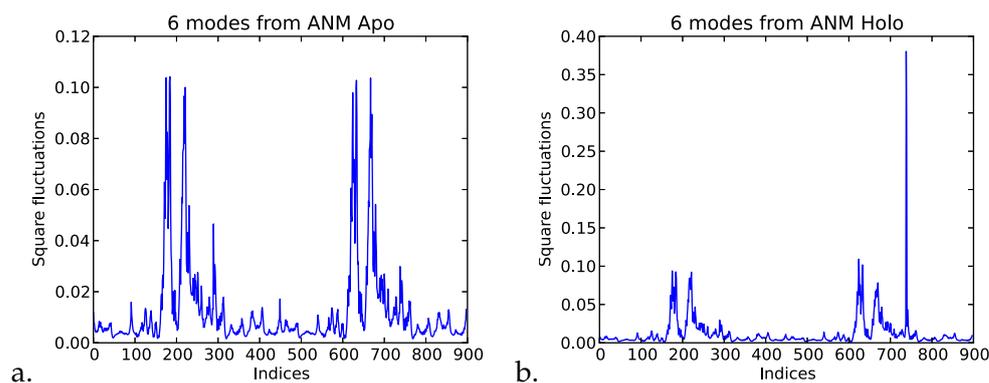


Figure 18: Square-fluctuations of the first six normal modes of the apo (a.) and the holo (b.) enzymes calculated from ANM.

tion between the apo and holo forms and the sixth mode of the apo enzyme overlaps most with the seventh mode of the holo form.

Visual inspection of the first few modes confirms that the global motions of the apo and holo enzymes described by ANM, are very similar (Figure 20). Pairwise cross-correlations of the residue fluctuations in the first six modes of the apo and holo enzymes are also similar; in both systems, the wing region is the most flexible (Figure 19). Moreover, the apo enzyme shows a clearly defined symmetry in the flexibility of the residues at the wing region of the two subunits, whereas the holo enzyme shows asymmetric dynamics with increased flexibility of one subunit compared to the other (Figure 18).

Simple elastic network models (both, GNM and ANM) show that the first two lowest eigenvector of the correlation matrix (first and second modes) corresponds to the global, in and out of the plane motions of the subunits with respect to the interface; third and fourth most correlated motion correspond to the opening and closing of the active site. Normalised square-fluctuations of the wing region of the holo and apo enzymes, show very high degree of flexibility (Figure 18), suggesting that this region of the enzyme plays role in the enzymes' functionality.

4.2 LOCAL DYNAMICS

4.2.1 Holo and Apo Enzymes

In this section the structural and dynamical properties of the apo enzyme are compared to those of the holo enzyme. Comparative

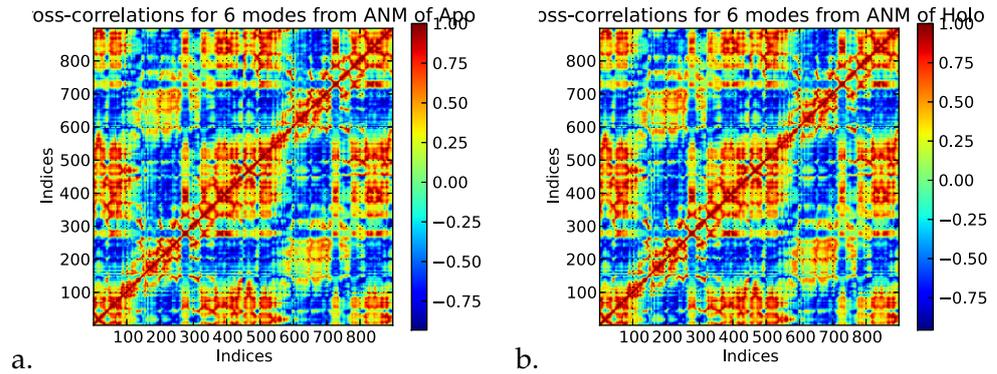


Figure 19: Cross-Correlation maps of the first 6 normal modes of the apo (a.) and the holo (b.) enzymes, calculated from ANM.

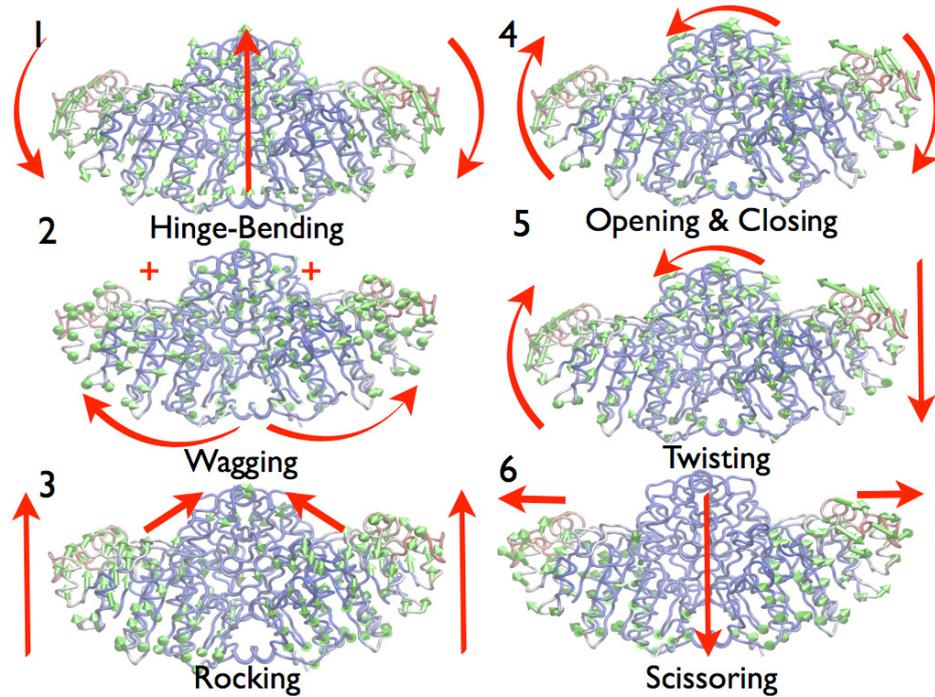


Figure 20: First six normal modes of the holo dimer based on the ANM. Arrows indicate direction of the modes while colouring is based on the mobility along the mode directions, red being most mobile and blue least. Apo enzyme describes similar motions with small differences in the fifth and the sixth modes (not shown).

Table 3: RMSD of the CA atoms along with the heavy atoms (in parenthesis), as calculated from the superimposition of the crystal structures of apo and holo enzymes and their individual subunits.

	Apo A	Apo B	Holo A	Holo B
Apo A	0	0.26(0.28)	0.17(0.19)	0.26(0.64)
Apo B	0.26(0.28)	0	0.28(0.64)	0.19(0.24)
Holo A	0.17(0.19)	0.28(0.64)	0	0.25(0.61)
Holo B	0.26(0.64)	0.19(0.24)	0.25(0.61)	0

analysis of MD simulations performed on individual subunits of each form of the enzyme explains:

1. The structural and dynamic changes in the enzyme upon binding to the inhibitor
2. Asymmetric local dynamics of the individual subunits of the holo enzyme and its relation to the negative cooperativity

4.2.1.1 *Crystal Structures*

Comparative structural analysis of individual subunits within the crystal structure of the holo dimer, shows that structurally subunits are remarkably similar, with RMSD of the CA atoms, with respect to each other, equal to 0.25Å (0.61Å for the heavy atoms). RMSD of the individual chains of the holo and apo enzymes is 0.17Å and 0.19Å for the CA and 0.24Å and 0.27Å for the heavy atoms chains A and B respectively.

Structure variance analysis based on the crystallographic B-factor suggest that the interface region of the enzyme is remarkably rigid in both subunits, of apo and holo enzymes, as opposed to the “wing” and “crown” regions, where B-factor values are higher and the subunit B is more flexible than subunit A (Figure 21).

4.2.1.2 *Overall Dynamics*

Root Mean Square Fluctuation (RMSF) values calculated from the last 80ns of 100ns long MD simulations are in a good agreement with experimental B-factors when scaled accordingly. If we convert RMSF values obtained from the last 80ns of 100ns long MD simulation into the B-factor using the following formula: $B = 8 * PI^2 * RMSF^2 / 3$ we can see agreement between experimental B-factor and RMSF with large values in the loop and terminal regions, as well as the crown domain and the wing region (Figure 23 (b.) and (c.)). Asymmetric behaviour of the individual subunits is preserved during the course

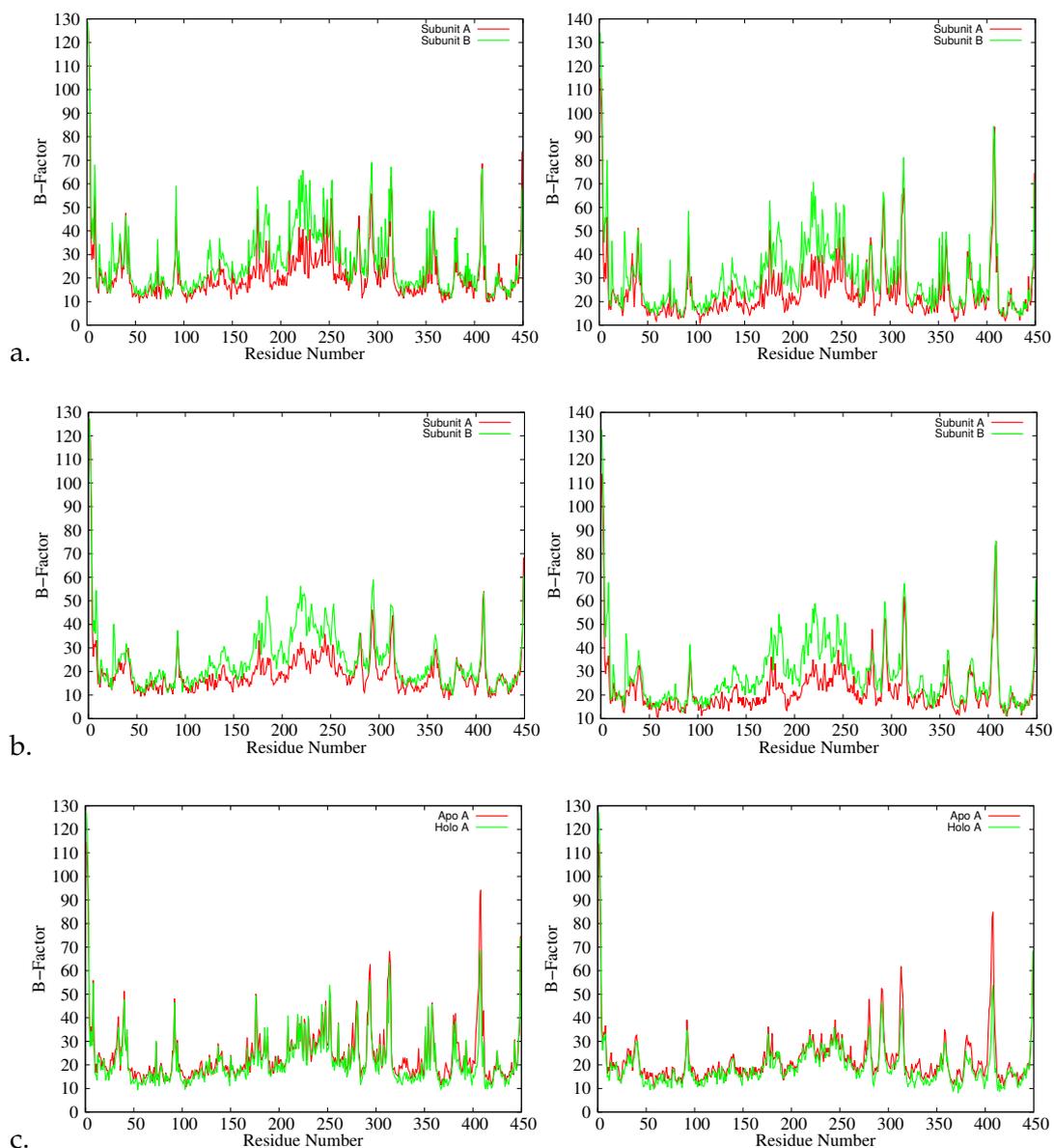


Figure 21: Crystallographic B-factors per residue of (a.) the side and (b.) the main chain atoms of subunits A and B of the holo AP (1ed8.pdb) (left) and the apo AP (1ed9.pdb) (right). Comparison of the B-factor values of the subunits A (left) and B (right) of the apo enzyme with the respective subunits of the holo enzyme (c.).

of the simulations with the higher flexibility of the residues in one subunit as opposed to the other. Flexibility of the residues at the crown domain and the wing region is also very well replicated by the simulations indicating functional importance of those residues (Figure 23). The residues of the apo and holo enzymes have very similar flexibilities, however the B-factor values per residue of the apo enzymes are slightly higher, indicating the relative stability of the holo enzyme with respect to the apo one, possibly, achieved upon binding to the product. Notably, RMSF values of the subunit A in the apo dimer are more similar to those of the subunit B in the holo dimer, than those of the corresponding subunit A (Figure 22).

This results might be an indication of the asymmetry in the holo dimer; subunit A of the apo dimer that was more rigid in the crystal structure becomes flexible to a degree similar to that of the subunit B of the holo dimer. This data suggest that the apo dimer can exhibit different flexibilities in its subunits, whereas the holo dimer is more restricted, possibly due to the fact that the structure of the holo enzyme is already stabilised by the product. The interpretation of the RMSF values goes hand in hand with the idea of the selection of the preexisting conformations by the substrate/product as opposed to the induced fit model where the conformational changes are assumed to take place upon binding. Indeed, already in the 100ns simulation time, we see significantly different degrees of flexibility in the two subunits, indicating an intrinsic asymmetry of the enzyme rather than asymmetry induced upon binding. Such an asymmetry of the subunits, is profound near the so called “wing” region (especially the residue numbers 210 to 240) and the crown domain (residue numbers 270 to 290 and 390 to 410) (Figure 23).

Stability of the MD simulations is monitored by root mean square deviation traces showing that in the holo dimer subunit A is significantly less flexible, than the subunit B, during the course of MD simulations (Figure 24 (a.)). In contrast, the overall shape and size of the subunits of the apo enzyme, as demonstrated by the R_{gyr} , is comparable, but not as similar as one would expect from the structural symmetry (Figure 24 (b.)). Subunits of the holo enzyme seem more asymmetric; R_{gyr} of the subunits indicates that the subunit B is larger in overall size and shape, than the subunit A. This might be an indication that during the 100ns simulations time, the conformation of the holo dimer is sampled, with one active site being opened while another one closed. This makes sense, as the holo enzyme has to favour both, open and closed conformations, to facilitate product binding and release. The entire apo dimer has larger average R_{gyr} and lower average RMSD, than the holo dimer, during the 100ns simulation time (Figure 24 (c.)). This is because apo form is more open and ready to bind the substrate, while the holo dimer is inhibited by the product and its subunits exhibit different dynamic behaviour.

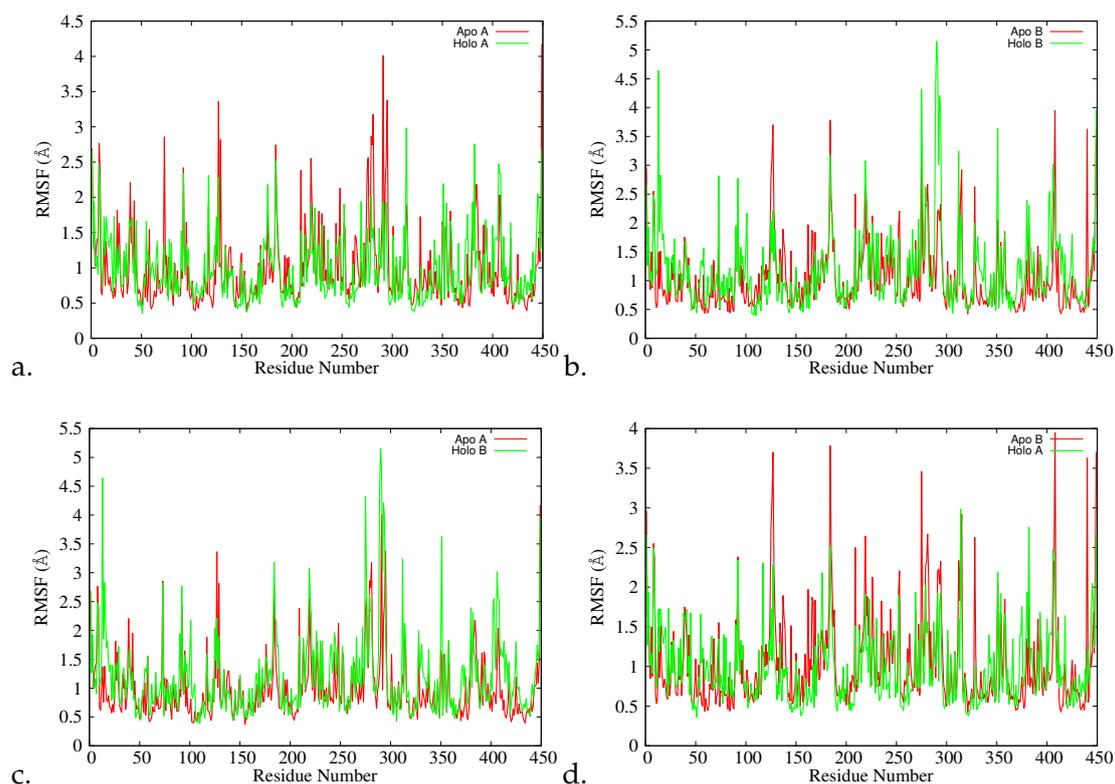


Figure 22: RMSF of the CA atoms of the (a.) subunits A and (b.) subunits B of the holo and apo enzymes, and (c.) subunit A of the apo enzyme vs subunit B of the holo enzyme, and (d.) visa versa.

4.2.1.3 Hydrogen Bonds

Analysis of the hydrogen bonding network of the individual subunits during the course of the MD simulations reveals very different occupancies of the important hydrogen bonds. Hydrogen bonds between the residues in the active site of the subunit A of holo dimer are highly occupied, whereas the corresponding residues in the active site of the counter subunit have relatively low occupancies (Table 5). This results can explain the overall stability of subunit A compared to subunit B. Particularly interesting is the high occupancy of the hydrogen bond between the active site residue pairs, S102->D369 (the ligand of Zn₂); the occupancy of this bond is almost 100% for the stable subunit A and near 0% for the flexible subunit B. It can be speculated that the high occupancy of this active site hydrogen bond plays a role in positioning D369 near the Zn₂ ion to keep its tetrahedral coordination optimal for the subsequent strong interaction with the bound product. When this hydrogen bond is gone, the coordination structure of the Zn₂ ion is altered and D369 is no longer coordinated to

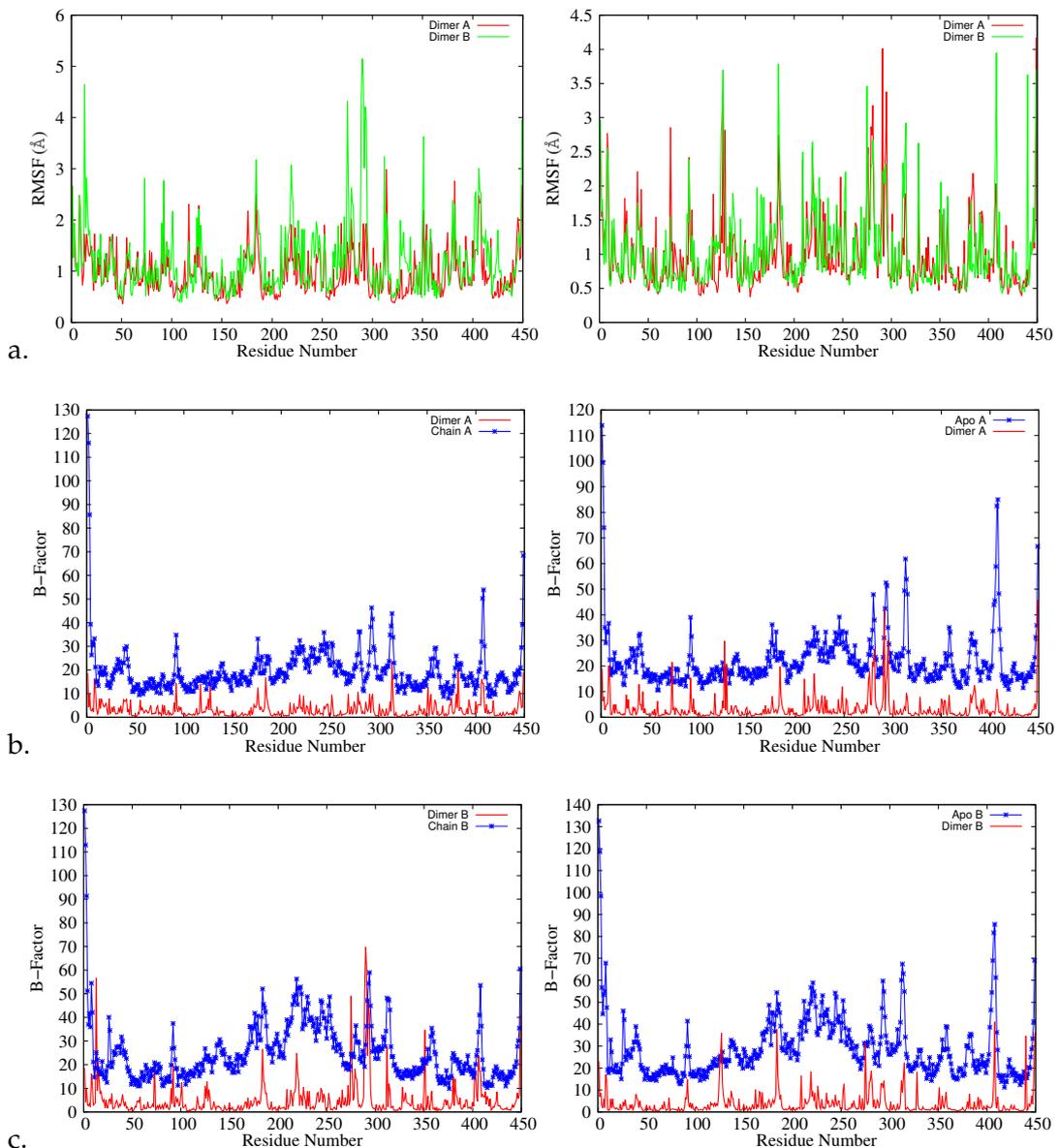


Figure 23: RMSF of the CA atoms of subunits A and B as calculated from the last 80ns of 100ns long MD simulations of the holo (a.) and apo (b.) enzymes; Comparison of the B-factor values calculated from MD simulations (red) versus those obtained from the experiment (blue) for subunits A (b.) and B (c.) of the holo (left) and apo (right) enzymes.

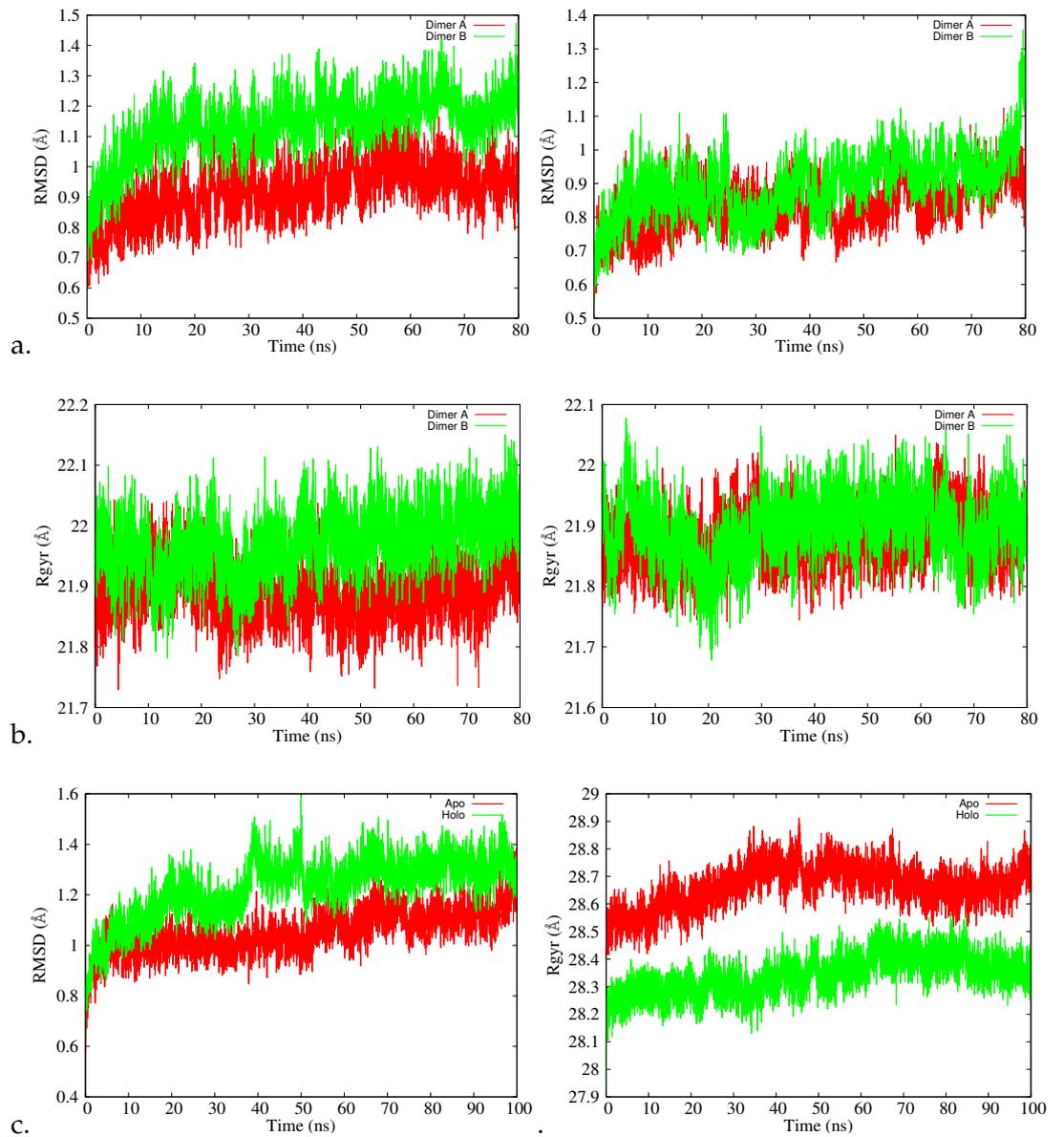


Figure 24: Time evolution of the carbon alpha RMSD with respect to the starting structure of the holo (left) and apo (right) enzymes (a.) and their Rgyr (b.) as calculated for subunits A and B and the entire dimer (c.). Only the last 80ns of the 100ns long MD simulations is considered.

Table 4: Hydrogen bond occupancies of the residues across the interface of the apo and holo dimers, as computed from the last 80ns of 100ns long MD simulations. Hydrogen bonds with highest differences in the occupancies are shown in bold.

Hydrogen Bond		Occupancy (%)	
Donor	Acceptor	apo	holo
ARG24-Main-N	ASP937-Side-OD1	31.89	95.94
THR392-Side-OG1	TYR598-Side-OH	86.70	84.96
ARG34-Side-NH1	ASP539-Side-OD1	47.00	77.74
ARG10-Side-NH1	VAL930-Main-O	88.81	19.50
THR892-Side-OG1	TYR98-Side-OH	91.36	78.86
ARG34-Side-NH2	ASP539-Side-OD1	57.26	23.00
ARG34-Side-NH2	ASP539-Side-OD2	42.75	79.19
ARG510-Side-NH1	LEU432-Main-O	91.80	0
ARG510-Side-NH1	VAL430-Main-O	84.19	0.60
LEU525-Main-N	ASP437-Side-OD1	82.62	73.12
ARG34-Side-NH1	ASP539-Side-OD1	47.00	77.74
ARG34-Side-NH1	ASP539-Side-OD2	56.88	21.57
ARG562-Side-NH1	GLN416-Side-OE1	87.61	94.05
ARG562-Side-NH2	GLN416-Side-OE1	86.11	66.45
GLN416-Side-NE2	ASP555-Side-OD2	85.65	50.80
GLN416-Main-N	THR559-Side-OG1	85.34	52.71

the zinc. Visual inspection of the individual subunits confirms that in subunit B amino acid residue D369 moves away from Zn2 and is replaced by a water molecule (Figure 27 (c.) and (d.)). This result is a possible explanation for the increased product binding affinity of one subunit as opposed to the other.

Hydrogen bonding network across the interface is a means of communication between the subunits. When the substrate/inhibitor is bound to the enzyme, conformational (such as structural stabilisation of a holo enzyme) and chemical (such as new interactions formed with the product and the active site residues) changes take place in the active site resulting in the changes at the interface as manifested by increase or decrease of the occupancies of certain hydrogen bonds (Figure 4). Interestingly, two Arginine residues located at the flexible N-terminal loop, that is “inserted” from one subunit to another, have very different Hbond occupancies in the apo and holo forms of the enzyme; occupancy of R24->D937 hydrogen bond is increased

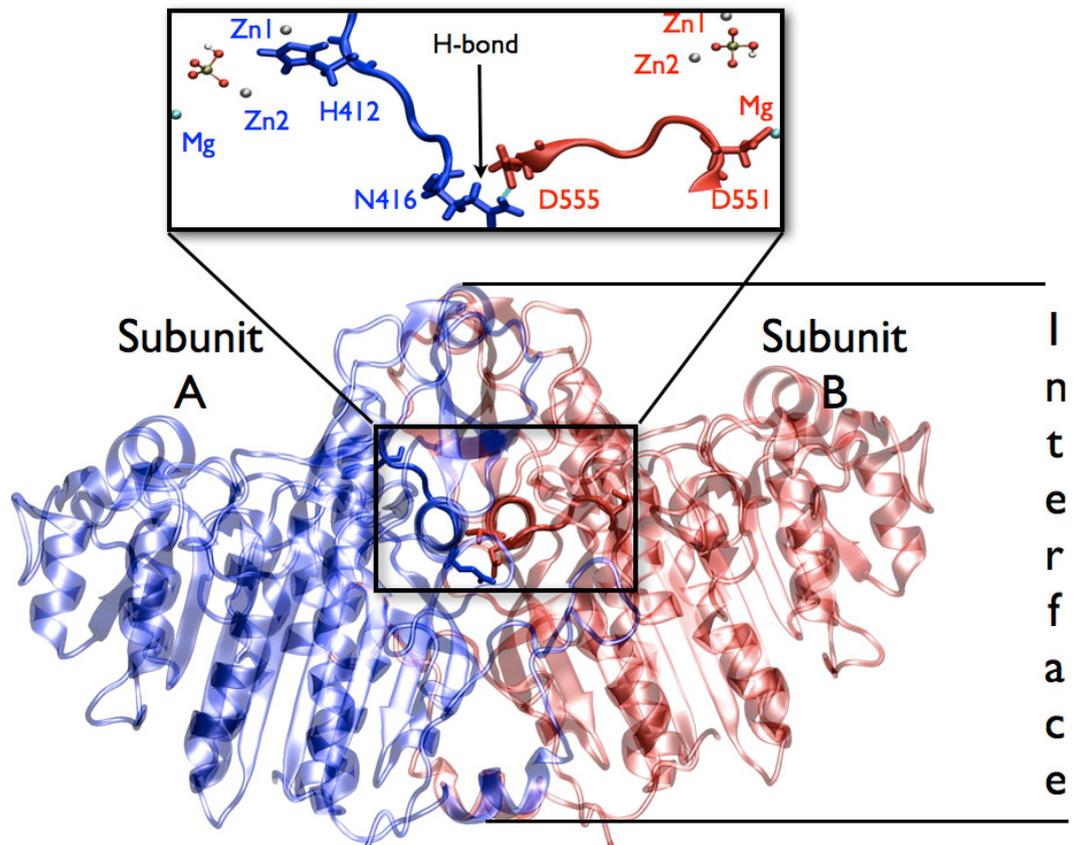


Figure 25: Proposed pathway for the transfer of information on structural changes between subunits A and B, (coloured in blue and red respectively) via the N416->D555 hydrogen bond. R559 is located on the same interfacial helix as D555 and in the apo enzyme, forming another N416->R559 highly occupied hydrogen bond, which has a decreased occupancy in the holo enzyme. Hydrogen bonding network across the interface, may serve as means of communication between the subunits, transferring the information of product binding and release.

Table 5: Hydrogen bond occupancies of the residues in the active site of the subunits of the apo and holo dimers, as computed from the last 80ns of 1000ns long MD simulations. Hbonds with highest differences in the occupancies are shown in bold.

Donor->Acceptor	Occupancy (%)		Donor->Acceptor	Occupancy (%)	
	Apo	Holo		Apo	Holo
Subunit A			Subunit B		
R166(NH1)->D101(OD1)	0	65	R666(NH1)-D601(OD1)	12	67
ARG166(NH1)->D101(OD2)	1	37	R666(NH1)->D601(OD2)	11	24
R166(N)->G118(O)	26	20	R666(N)->G618(O)	38	0
-	0	0	R666(NH1)->S602(OG)	0	32
S102(OG)->D369(OD1)	0	96	S602(OG)->D869(OD1)	0	3
S105(OG)->D369(O)	51	85	S605(OG)->D869(O)	31	95
G52(N)->D369(OD2)	14	52	G552(N)->D869(OD2)	22	82
S105(N)->D101(O)	28	86	S605(N)->D601(O)	7	41
A103(N)->D101(OD1)	0	59	A603(N)->D601(OD1)	0	0
D327(N)->D51(O)	22	61	D827(N)->D551(O)	40	1
D330(N)->I326(O)	49	49	D830(N)->I826(O)	65	38
A333(N)->D330(O)	62	78	A833(N)->D830(O)	78	62
K328(NZ)->D153(OD1)	4	44	K828(NZ)->D653(OD1)	0	1
K328(NZ)->D153(OD2)	23	46	K828(NZ)->D653(OD2)	0	1
Y402(OH)->D330(OD2)	98	0	Y902(OH)->D830(OD2)	1	27
H331(N)->D327(O)	60	70	H831(N)->D827(O)	34	75
H370(CE1)->H412(NE2)	0	7	H870(CE1)->H912(NE2)	0	0
H412(ND1)->T100(O)	97	0	H912(ND1)->T600(O)	63	42
H370(ND1)->H370(O)	0	44	H870(ND1)->H870(O)	0	0
H370(ND1)->D369(OD2)	0	0	H870(ND1)->D869(OD2)	0	83
D51(N)->G323(O)	44	68	D551(N)->G823(O)	68	44
E322(N)->V146(O)	72	72	E822(N)->V646(O)	67	59
S147(OG)->E322(OE1)	100	99	S647(OG)->E822(OE2)	100	100
Q320(NE2)->T155(O)	28	35	Q820(NE2)->T655(O)	43	57

upon binding to the ligand, whereas R10->V930 hydrogen bond occupancy decreases significantly (Table 4). The subtle interplay of the occupancies in the hydrogen bonding network in the active site and across the interface might be responsible for substrate binding and product release. Information about structural changes taking place upon binding to the substrate/inhibitor is propagated from one subunit to another, via the H-binding network across the interface. By tracing the residues forming hydrogen bonds from the active site of one subunit towards that of another, and taking into account changes in the occupancy of certain Hbonds upon binding to the inhibitor, it becomes possible to identify communication pathways, transmitting the information between the subunits. One such pathway lays right at the centre of the axis of symmetry and connects Zn1 ligand, H412 of subunit A with Mg ligand, D551 of subunit B via N416->D555 and N416->T559 hydrogen bonds, that are highly occupied in the apo enzyme and have reduced occupancy in the holo form (Figure 25). Chemical changes taking place in one subunit upon binding to the ligand are communicated to the counter subunit using the flexible loops connecting the active site with the interface via the network of intra-subunit hydrogen bonds.

4.2.1.4 *Active Site Dynamics*

Understanding dynamics of the active site residues directly coordinating the bound product, as well as the metal ions, is crucial for identifying key structural and dynamic features necessary for catalysis (Figure 26). Due to their proximity to inorganic phosphate in the crystal structure, we consider the following amino acid residues in the active site: S102, R166, D101, H331, H412, H370 and H372 (Figure 28). In the crystal structure of the holo AP, Hbond is formed between the guanidinium group of R166 and two phosphate oxygen atoms. Our simulation results show that this phosphate bridge is broken in one subunit (subunit A) and maintained in another (subunit B) (Figure 29). Visual inspection of MD trajectories, also confirms that the dynamics of Zn1 coordination structure is very different within the subunits of the holo dimer.

To identify the residues coordinating metal ions we observed the coordination structure of Zinc ions with 2.49Å radius from their centre in subunits A and B and of Mg ion with radius 2.95Å (Figure 27). The differences are striking: Zn1 of the subunit A is saturated by the negative charge of the surrounding Histidine and Aspartate residues, while in the subunit B the coordination of Zn1 ion is maintained by the help of the oxygen atoms of nearby water molecules and inorganic phosphate. Furthermore, tetrahedral coordination geometry of Zn1 is altered in subunit B and due to the flexibility of zinc ion, Zn1 becomes penta-coordinated. The fact that the active site of subunit B has more

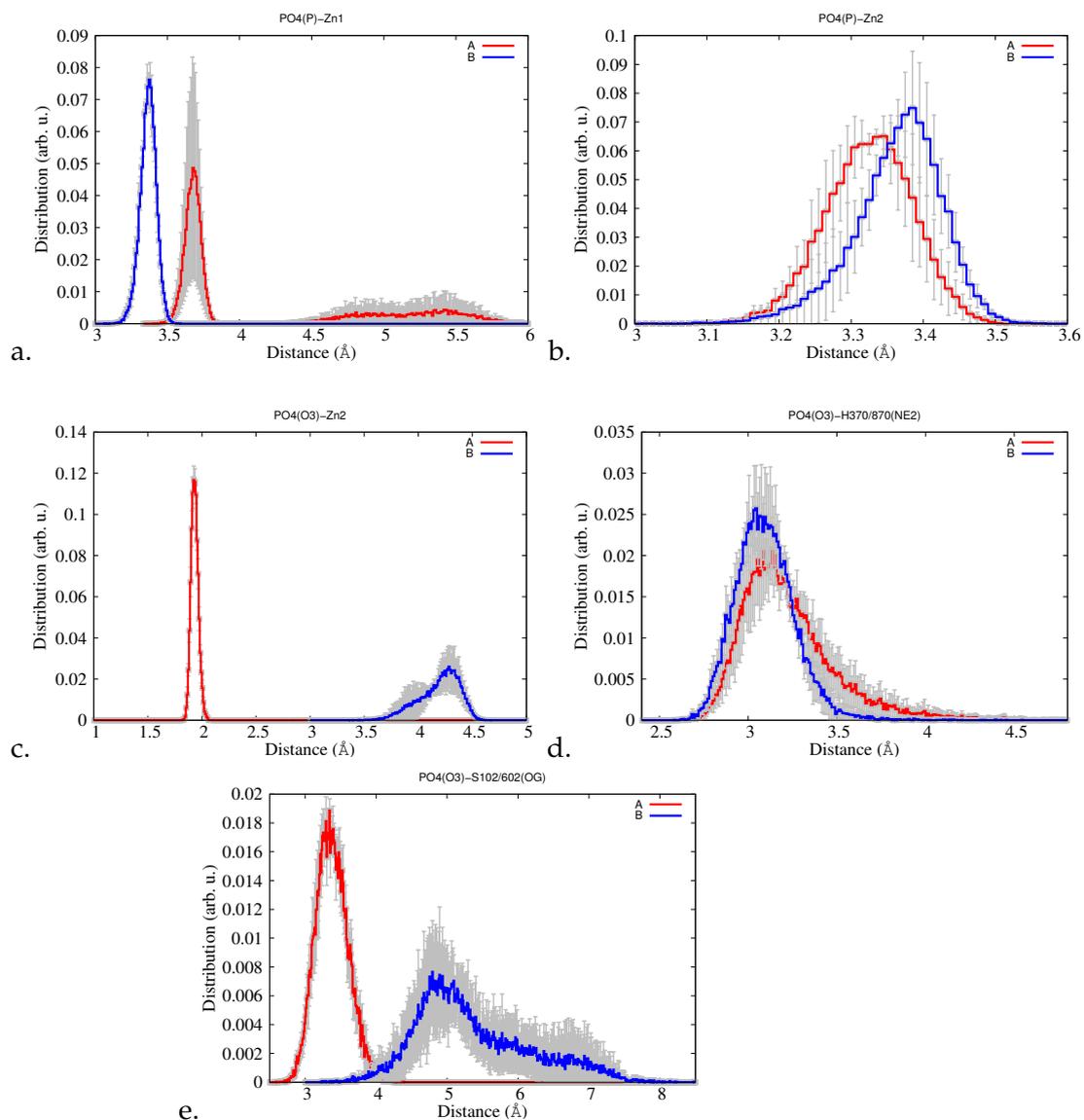


Figure 26: Distance distribution plots of the key residues in the active site of the holo AP, averaged over three independent MD runs. PO₄(P)-Zn1 (a.), PO₄(P)-Zn2 (b.), PO₄(O₃)-Zn2 (c.), PO₄(O₃)-H370 (d.), and PO₄(O₃)-S102 (e.).

water molecules might be explained by its increased flexibility resulting in the neighbouring amino acid residues drifting away from the Zinc ions, thereby making space for the water molecules available. In contrast, the Mg binding sites, which is further away from the phosphate, show similar coordination structures in both subunits (Figure 27 (e.) and (f.)).

Zinc ions that are coordinated by nitrogen and oxygen atoms of the substrate and the surrounding amino acids, act as useful catalytic agents in the hydrolysis of phosphate esters. Coordination dynamics of the metal ions in the subunit A is more stable than the one in the subunit B (Figure 28). Time evolution of the crucial distances illustrate flexible coordination geometry of zinc metal ions, that allows rapid shift of the enzymes' conformation enabling tight binding of the product into the active site of one subunit of a homo-dimer, as opposed to the looser binding to the other subunit. This is particularly true for $ZN_1-PO_4(O_2)$. In addition, $H_{412}(NE_2)-PO_4(O_2)$ distance fluctuates a lot more in subunit A than in subunit B, and $H_{370}(NE_2)-PO_4(O_3)$ distance fluctuates a bit more in subunit A than in subunit B (Figure 28). However, the average values are larger as one can see in the distance distribution plots (Figure 26).

The centre-to-centre distances of metal ions form a closely packed triangle with sides 4\AA (Zn_1-Zn_2), 4.8\AA (Zn_2-Mg_3), and 7\AA (Zn_1-Mg_3). These distances are almost identical in the two active sites of the crystal structure, but very different during the course of the MD simulation. When plotted as a function of simulation time, intra-metallic distances show distinct asymmetric behaviour with subunit A being more rigid than subunit B (Figure 30). Moreover, the centre-to-centre distance between Zn_2 and Mg in subunit B is almost twice as large as in subunit A, indicating a looser binding of the metal ions (most likely Zn_2) of the subunit B (Figure 30 (c.)).

Comparison of the intra-metallic distances of the apo and holo enzyme indicates significant increase in the average distance between Zn_2 and Mg in subunit B ions upon binding to the phosphate ion. In contrast the average Zn_2 -Mg distance slightly decreases in the subunit A of the holo enzyme (Figure 32 (a.)). Same is true for the average Zn_1 -Mg distance, that remains unchanged in the subunit A of both, the apo and holo enzymes, and increases in subunit B of the holo enzyme. The shape of the Zn_1 -Mg distance distribution plot in the apo enzyme, has two peaks, one resembling the shape of the distance distribution of the holo enzyme (Figure 32 (b.)). Similarly, the distance distribution plot corresponding to the average distances between two

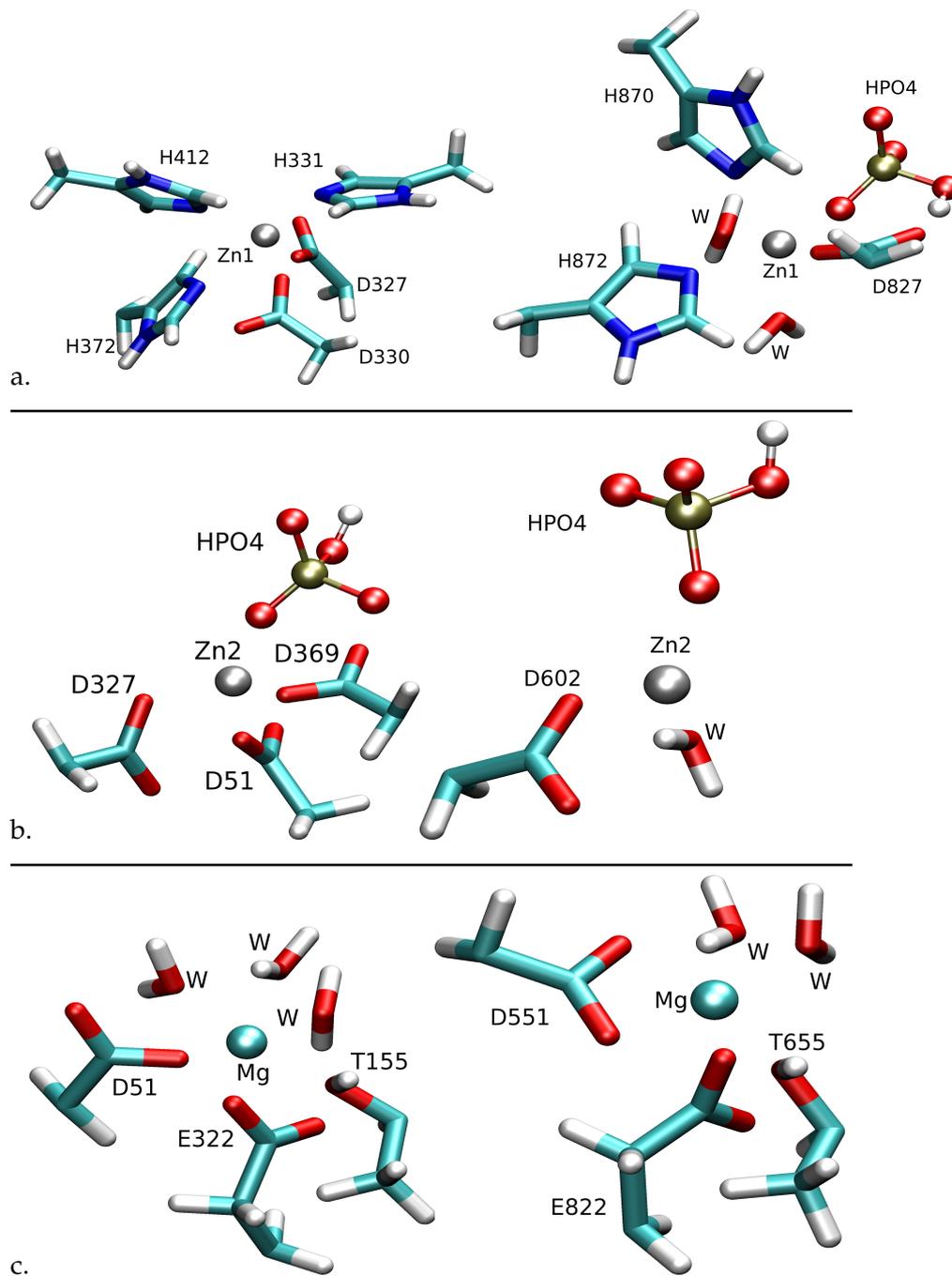


Figure 27: Coordination structure of the Zn1 (a.), Zn2 (b.) and Mg (c.) metal ions as observed in the last frame of the 100ns MD simulation. Subunit A and subunit B are shown in then left and right hand-sides respectively.

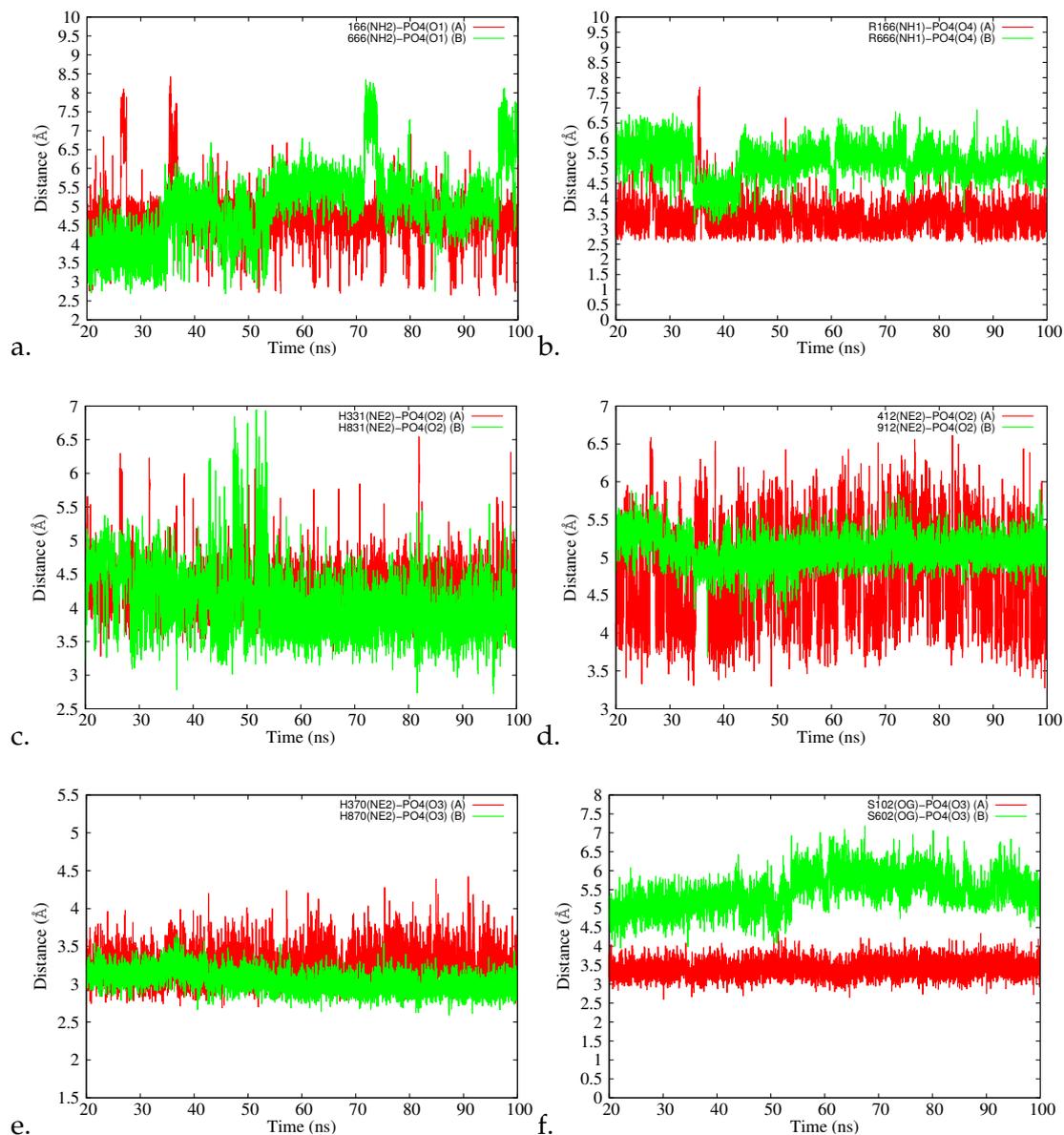


Figure 28: Time evolution of the key distances in the active sites of the subunits A and B of the holo enzyme, as calculated from the last 80ns of MD simulations. R166(NH2)-PI2(O11) (a.), R166(NH1)-PI2(O14) (b.), H331(NE2)-PI2(O12) (c.), H412(NE2)-PI2(O12) (d.), H370(NE2)-PI2(O13) (e.), S102(OG)-PI2(O13) (f.).

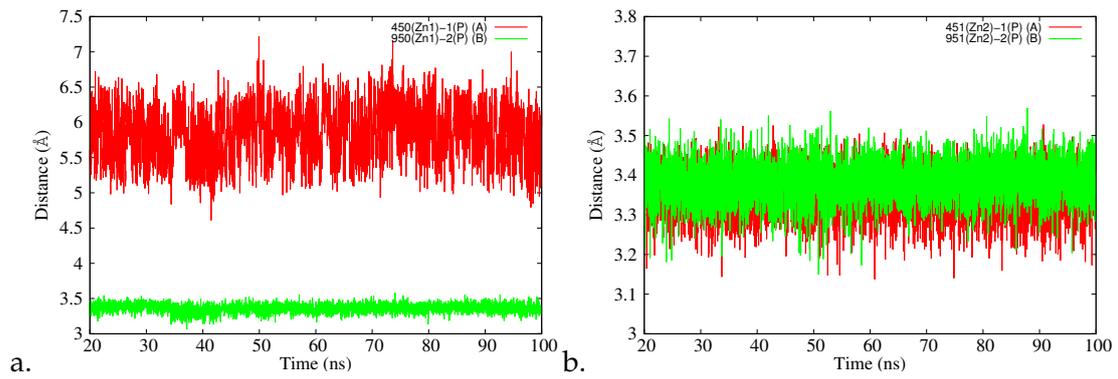


Figure 29: Time evolution of the distances between (a.) Zn1 and (b.) Zn2 metal ions and the phosphorus atom of the inorganic phosphate.

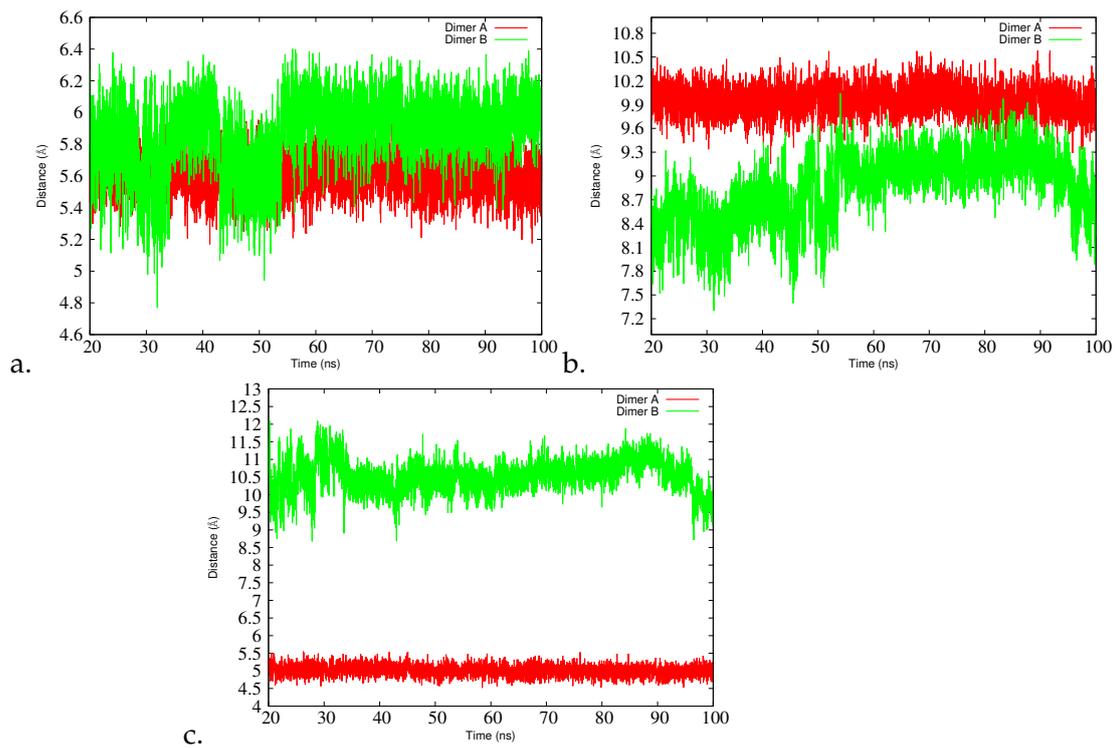


Figure 30: Time evolution of the distances between Zn1 and Zn2 (a.), Zn1 and Mg (b.) and Zn2 and Mg (c.) ions and the distance between Zn1 (d.) and Zn2 (e.) metal ions and the phosphorus atom of the inorganic phosphate.

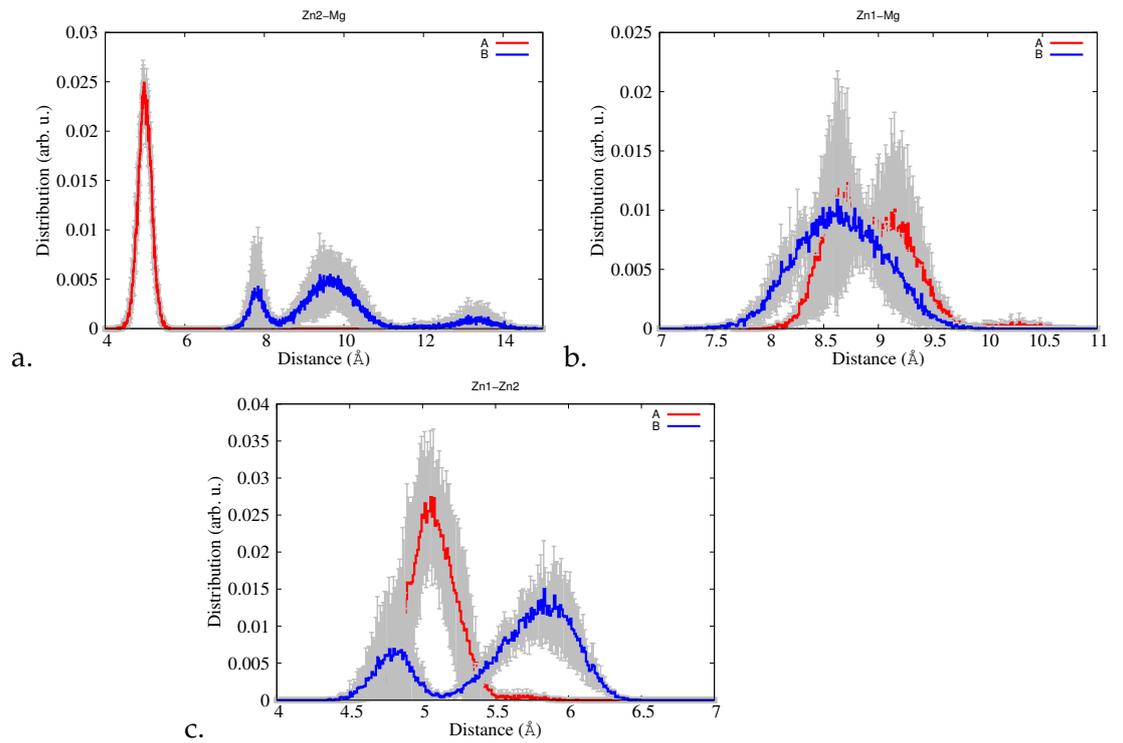


Figure 31: Distance distribution plots between the metal ions of the individual subunits of holo AP and averaged over three independent MD runs. Zn2-Mg (a.), Zn1-MG (b.), and Zn1-ZN2 (c.).

Table 6: Averages and standard deviations of RMSD and Rgyr calculated over all frames of 100ns MD simulations and RMSF and number of Hbonds calculated from the last 80ns of the dimer and the monomer.

	RMSD(Å)	Rgyr(Å)	RMSF(Å)	# of Hbonds
Dimer (MD1)	1.1±0.1	28.6±0.1	1.00±0.6	596±16
Dimer (MD2)	1.0±0.1	28.3±0.1	1.0±0.6	606±13
Monomer (MD1)	1.7±0.2	22.0±0.1	1.4±1.1	280±11
Monomer (MD2)	2.3±0.4	22.1±0.1	1.5±1.1	276±9

catalytic zinc ions, indicate asymmetric behaviour of the active sites, with the subunit B of the holo enzyme sampling larger distances than the corresponding subunit of the apo form (Figure 32 (c.)). The striking differences between the intra-metallic distances of the apo and holo enzyme are in favour of the proposed negative cooperativity scheme; increased flexibility of subunit B of the holo enzyme results in the overall enlargement of the binding pocket, making it easy for inorganic phosphate to be released, whereas more rigid subunit A provides optimal distances required to keep the product in the active site.

4.2.2 The WT Dimer, Monomer and T59R Mutant of the Apo Enzyme

The apo enzyme has been simulated in three different setups: as WT dimer, the T59R mutant of the dimer and a monomer, modelled as described in 3. In this chapter the results of these simulations are presented and the properties of the three models are compared.

4.2.2.1 Overall Dynamics

Stability of the MD simulations is monitored by RMSD and Rgyr. RMSD values of the CA atoms are calculated using the first frame of the production dynamics as a reference for each of the MD trajectories. These values, calculated for the entire protein (WT dimer, T59R mutant and monomer, respectively), the active site and the interface, provide a reliable indicator of structural variability and the relative contribution of the individual regions to the stability of the MD simulations. Two independent MD runs are performed for the WT dimer and the monomer, the differences within the individual runs are due to the randomly assigned initial velocities drawn from a Maxwell-Boltzmann distribution at 298K in each run [69].

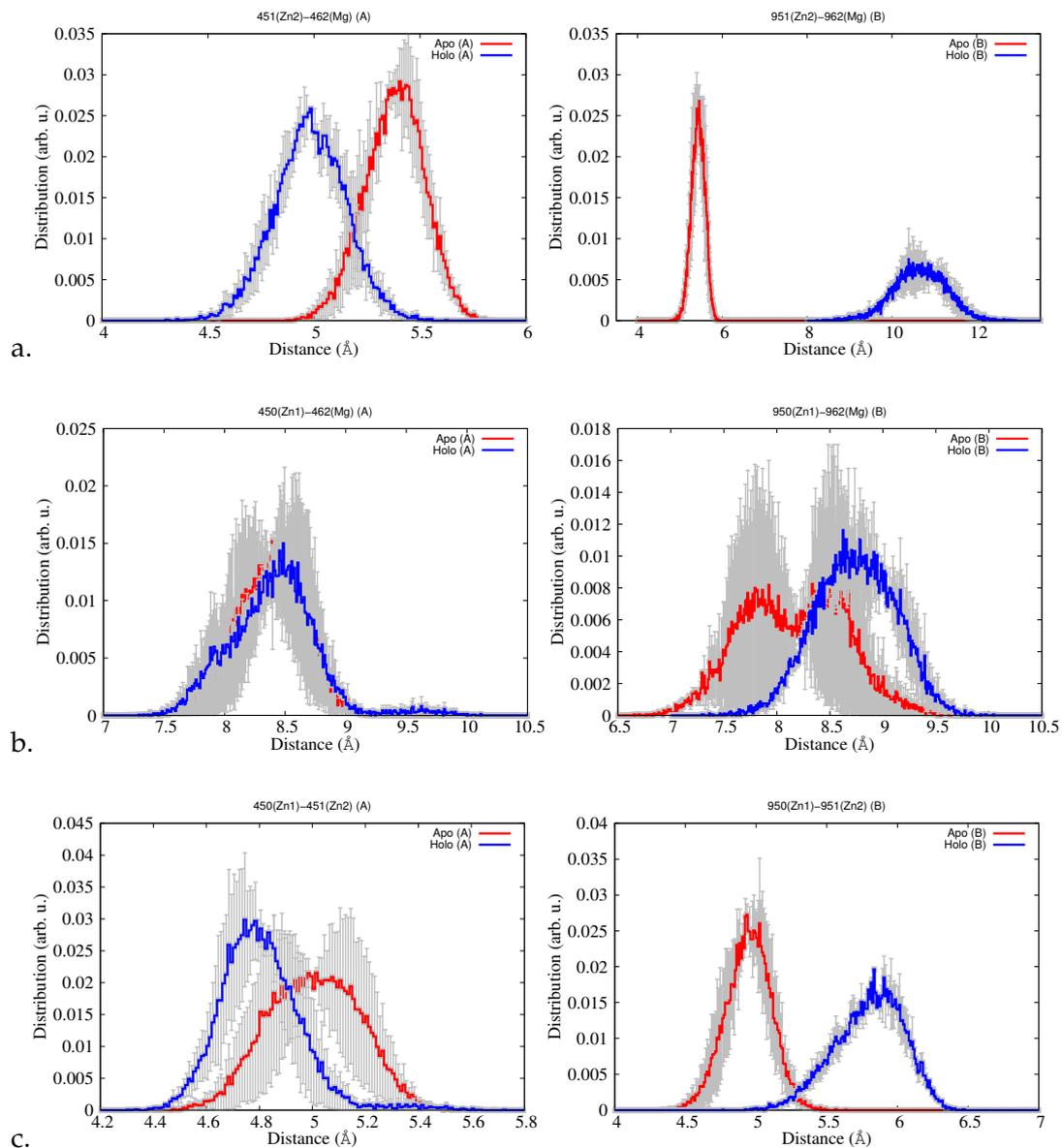


Figure 32: Distance distribution plots of the intra-metallic distances of the holo and apo AP in the two subunits and averaged over three independent MD runs. Left: subunit A, and right: subunit B. Zn2-Mg (a.), ZN1-Mg (b.), and ZN1-ZN2 (c.).

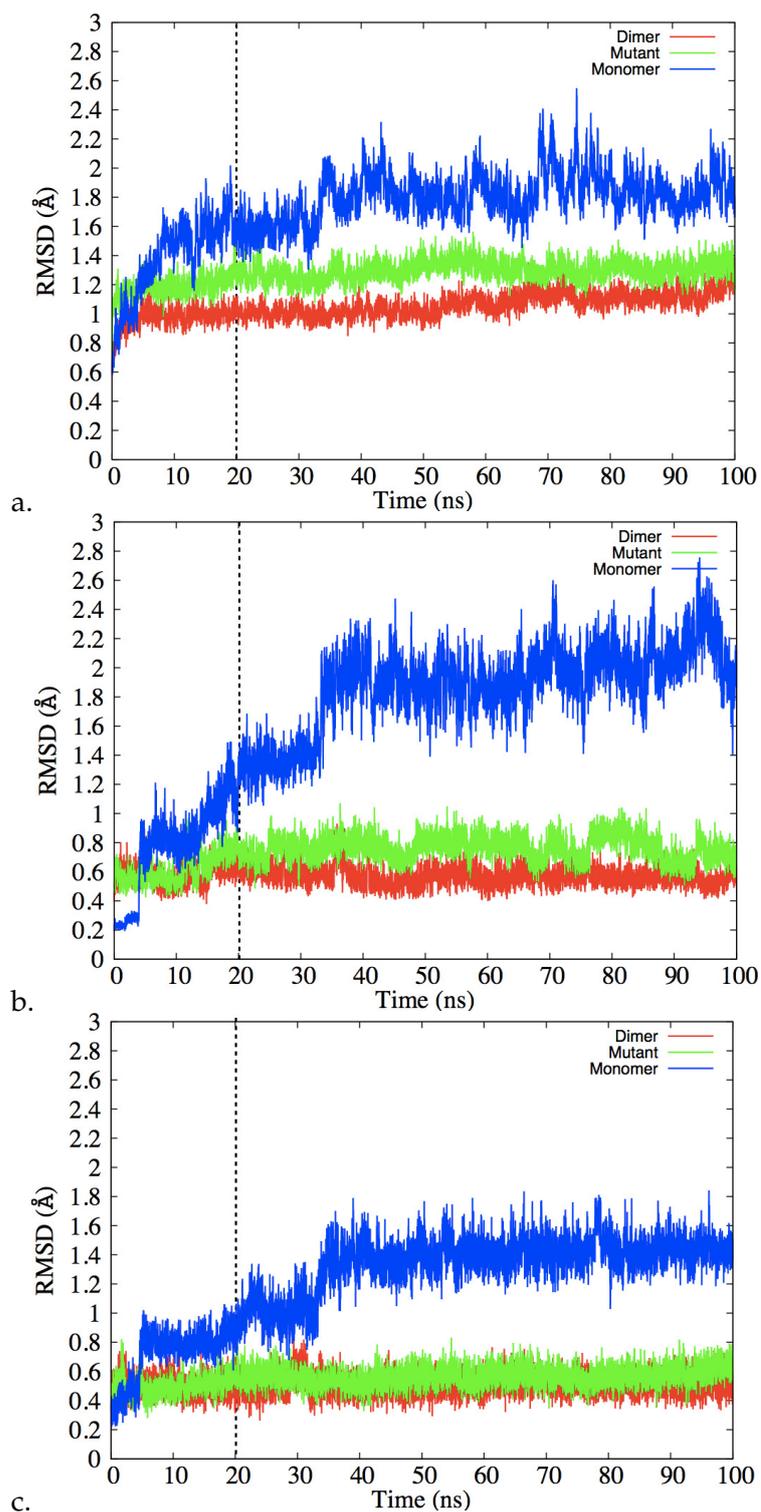


Figure 33: Time evolution of the CA atoms' RMSD with respect to the starting structure as calculated for: the entire WT dimer, the T59R mutant and the monomer (a.), residues at the interface of a single subunit (chain A) (b.), residues of the active site (c.). Dashed line indicates the fraction of the trajectories, not considered in the further analysis.

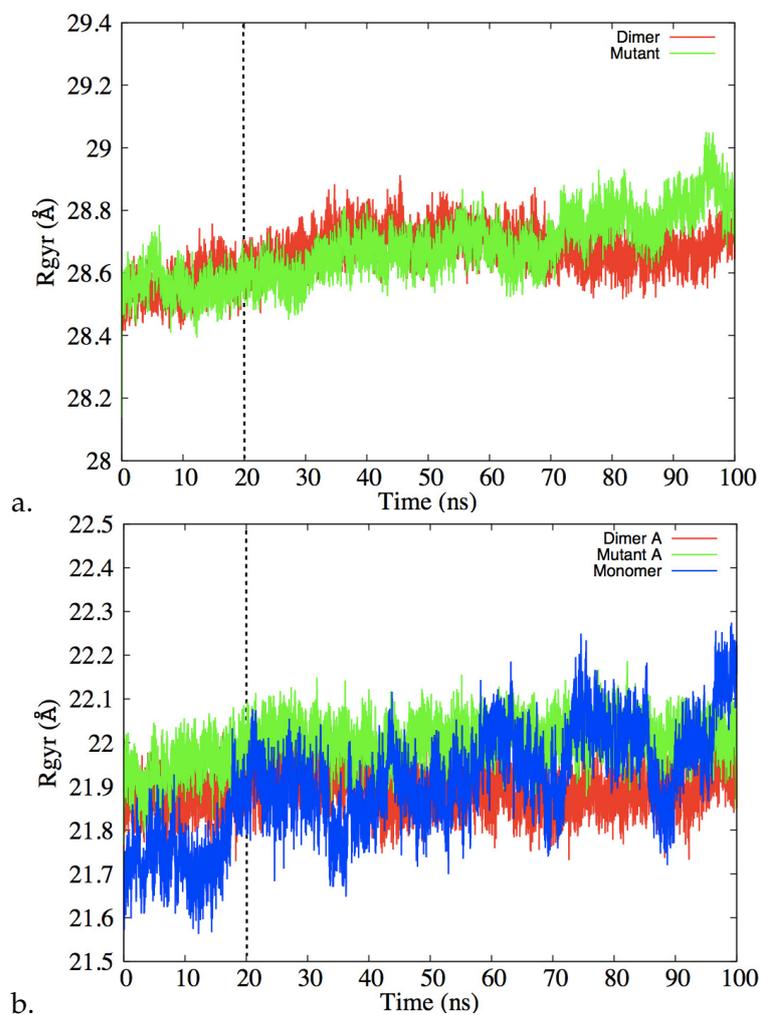


Figure 34: Time evolution of the Rgyr as calculated for: the entire dimer and the mutant (a.), the single subunit (chain A) of the dimer and the mutant and the monomer (b.). Dashed line indicates the fraction of the trajectories, not considered in the further analysis.

Table 7: Averages and standard deviations calculated over all frames of 100ns MD simulations of the dimer, the mutant and the monomer.

	RMSD	RMSF	Rgyr	# of Hbonds
Dimer	1.05 ± 0.1	1.1 ± 0.7	28.6 ± 0.1	216 ± 12
Mutant	1.27 ± 0.1	1.4 ± 0.7	28.7 ± 0.1	208 ± 11
Monomer	1.7 ± 0.2	1.6 ± 1.2	22.0 ± 0.1	103 ± 8

The WT dimer is remarkably stable during the 100ns simulation time, in contrast to the isolated monomer (Figure 35 (a.)). The (former) interface of the monomer is much more flexible, than that of the corresponding dimer, due to the lack of the stabilising interactions with the second subunit (Figure 35 (b.)). Residues located at the active site of the monomer are very flexible during the first half of the MD simulation time, and fluctuate around their average values during the last 50ns (Figure 35 (c.)).

The largest contribution to the variation of RMSD in the monomer is due to the “N-terminal loop” connecting the N-terminus of the monomer with the interface. In the dimer, this loop is very stable whereas in the monomer it fluctuates a lot (Figure 36).

The monomer shows a slightly larger average Rgyr than the single subunit of the dimer, corresponding to an increase of the monomer’s overall shape and size due to its higher flexibility (Figure 37).

The RMSD of the mutant is relatively stable during the course of the 100ns MD simulation time. However, the interface of the mutant is very flexible, due to the bulkiness and the positive charge introduced at the interface by the Arginine residues. In contrast, active site residues of the dimer and the mutant show very little deviation from the starting structure (Figure 33). The results confirm that the bulky and charged Arginine residues at the interface of the mutant, slightly push the subunits apart, resulting in the increase of the overall shape and size of the mutant (Figure 34).

4.2.2.2 Interface Dynamics

The stability of the (former) interface of the monomer is studied by means of the changes in the secondary structure elements and RMSF values of the residues comprising those elements. The H3 alpha helix (residues: 425-430), that folds into the turn and 3-10 helix-like secondary structure in the WT dimer, shows significantly different preference towards folding into the alpha helix in the monomer (Figure 38). Similarly, Sh3 (residues: 431-434) beta sheet partly unfolds and resembles the coil-like secondary structure in the monomer.

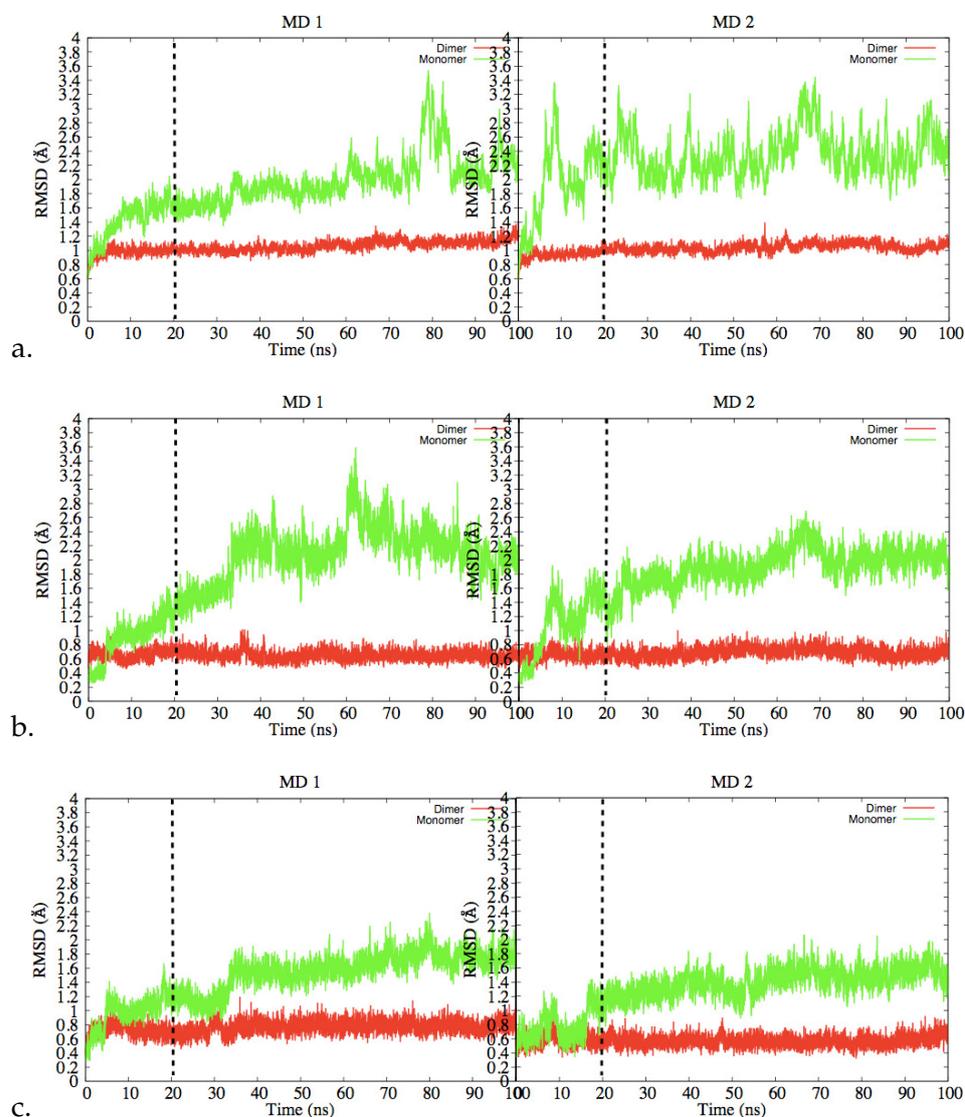


Figure 35: Time evolution of the CA atoms RMSD with respect to the starting structure as calculated from the two independent MD simulations of: the entire dimer and the monomer (a.), residues located at the interface of a single subunit (dimer A) of a dimer and the monomer (b.), the residues located at the active site of the dimer and the monomer (c.). Dashed line indicates the fraction of the trajectories not considered in the further analysis.

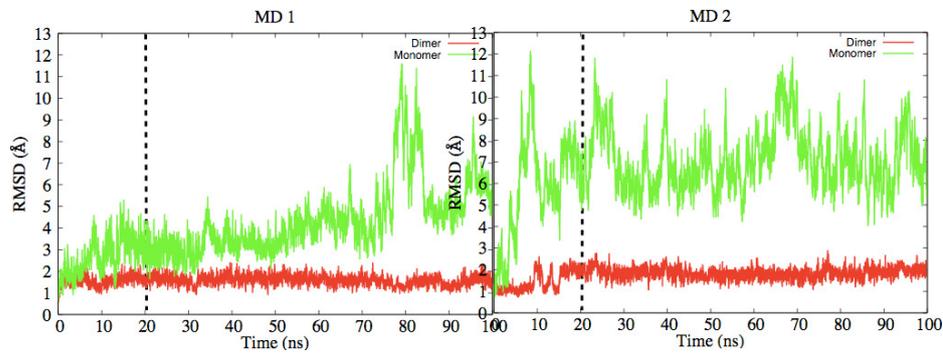


Figure 36: RMSD time-trajectory of N-Terminal loop with respect to the equilibrated structure as calculated from two independent MD simulations.

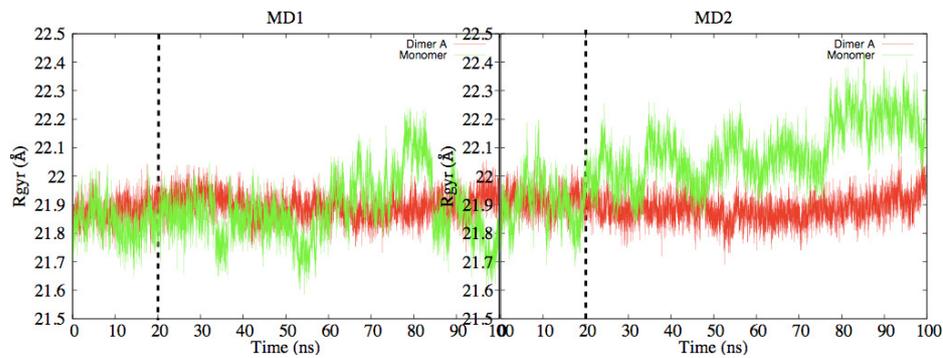


Figure 37: Time evolution of the Rgyr as calculated from the two independent MD simulations for the single subunit (dimer A) of the dimer and the monomer. Dashed line indicates the fraction of the trajectories not considered in the further analysis.

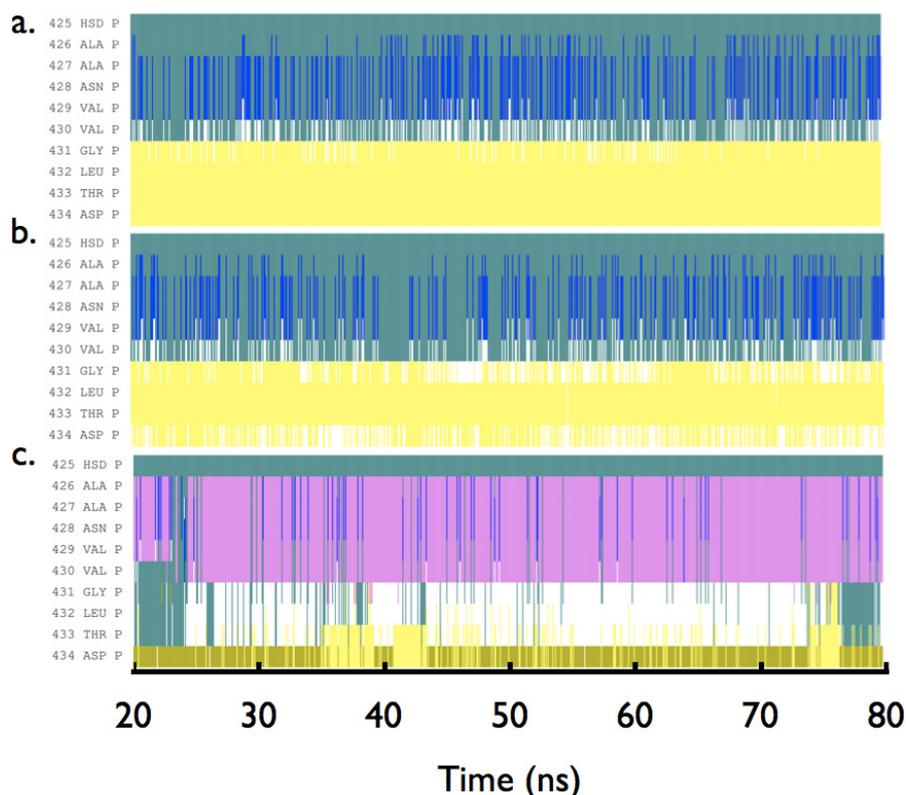


Figure 38: Secondary structure elements of the most flexible residues, located at the interface, as a function of time, as calculated from the last 80 ns of the dimer (a.), the mutant (b.) and the monomer (c.) simulations. H3 alpha helix corresponds to the residues: 425-430 and Sh3 beta sheet - residues: 431-434. Colour-coding is as follows: alpha helix - pink, beta sheet - yellow, coil - white, turn - green and 3-10 helix - blue.

The monomer shows greater flexibility than the dimer, in terms of the loss in the secondary structure elements and increased RMSF values, especially of the residues located at the (former) interface. In contrast, residues located at the wing region and the crown domain are similarly flexible, as manifested by the comparable RMSF values (Figure 39). Average RMSF per residue for the monomer and the dimer is large in the loop and terminal regions of the enzyme. The largest flexibility is observed for the loop connecting the N-terminus with the H1 helix of the interface (Figure 43).

At the interface of the mutant, disturbed by the mutations, some unfolding/refolding events take place as demonstrated by the loss in secondary structure elements and increased RMSF values (Figures 43 and 41). Average RMSF per residue for the mutant is large in the loop and terminal regions, in addition to the interface, the crown domain

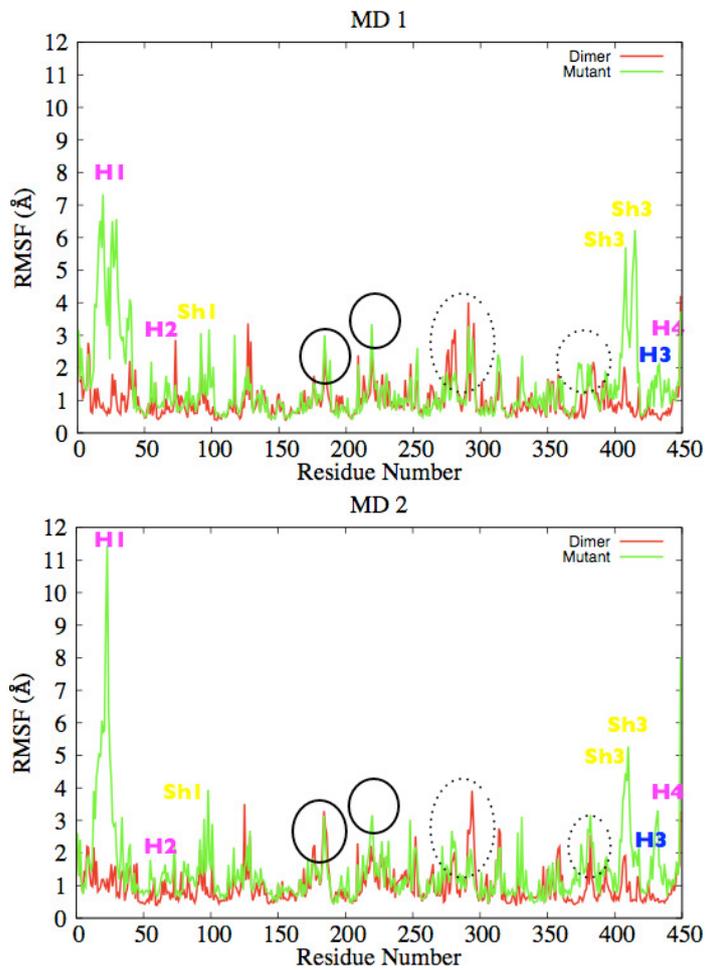


Figure 39: RMSF values of the WT dimer and the monomer, with labels at the interface of the monomeric subunit. Alpha helices are labeled as H1 (29-35), H2 (54-66), H3 (425-430) and H4 (435-447) while beta sheets are labeled as Sh1 (80-85), Sh2 (417-422) and Sh3 (431-434); Residues comprising the “wing” region and the “crown” domain are marked with solid and dashed circles respectively. Labels are coloured to match the 3D representation of the residues at the interface as in Figure 15.

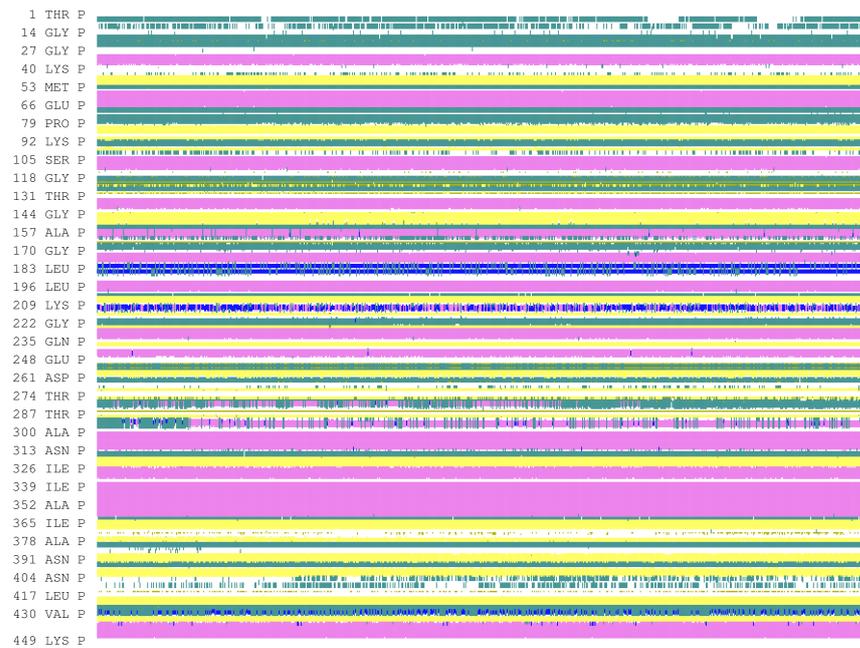


Figure 40: Secondary structure elements of the WT dimer as a function of time, as calculated from the last 80ns of the dimer simulation (chain A). Colour-coding is as follows: alpha helix - pink, beta sheet - yellow, coil - white, turn - green and 3-10 helix - blue.

and the so called wing region (comprising residues 170-186 and 224 - 232), where the mutant is most flexible and even exhibits partial unfolding (Figures 43 and 41).

4.2.2.3 Hydrogen Bonding Network

In order to understand structural and dynamic differences between the WT dimer and the monomer, we have examined the hydrogen bonding interactions in these two systems. The overall number of Hbonds as a function of the simulation time is stable, however in the monomer, as expected, the average number of Hbonds is less than half of that, in the dimer, due to about forty hydrogen bonds located at the interface between the two subunits, that are not present in the monomer (Figure 44). It is known that the extensive hydrogen bonding network stabilises the structure of the WT dimer [15]. The overall number of hydrogen bonds as a function of time is stable in all 3 models, however in the mutant this number is slightly reduced compared to that in the dimer (Figure 44) (Table 7).

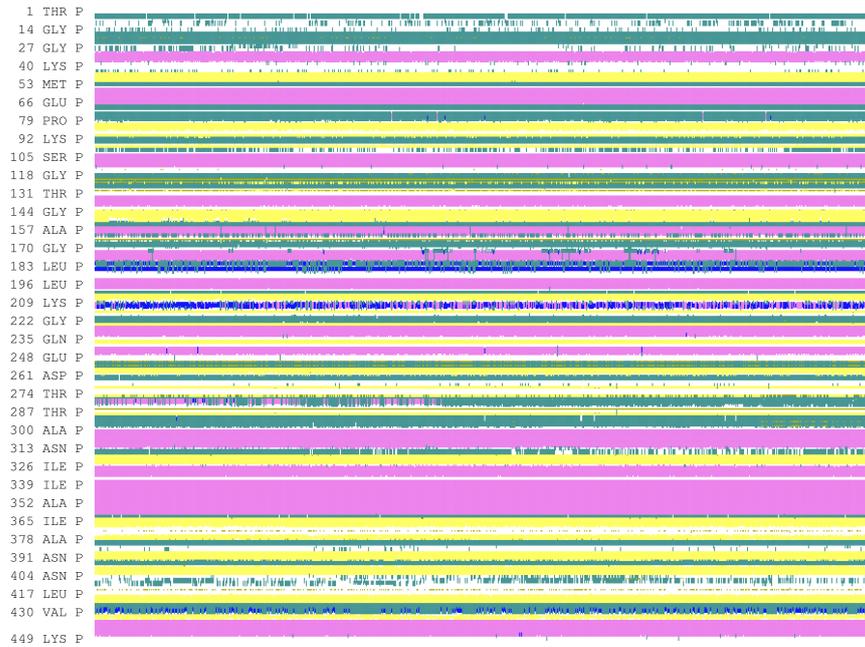


Figure 41: Secondary structure elements of the T59R mutant as a function of time, as calculated from the last 80ns of the mutant simulation (chain A). Colour-coding is as follows: alpha helix - pink, beta sheet - yellow, coil - white, turn - green and 3-10 helix - blue.

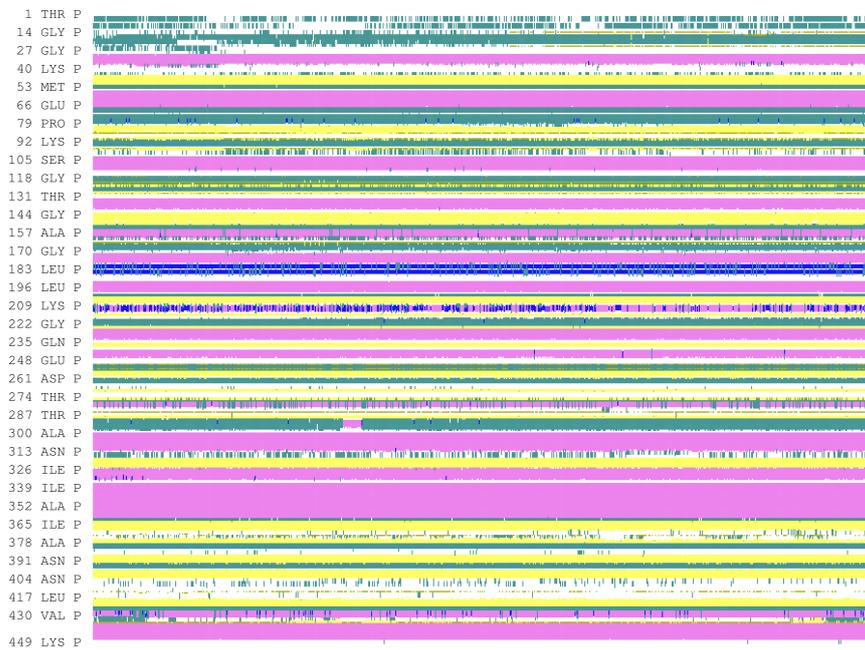


Figure 42: Secondary structure elements of the residues, located at the interface, as a function of time, as calculated from the last 80ns of the monomer simulation. Colour-coding is as follows: alpha helix - pink, beta sheet - yellow, coil - white, turn - green and 3-10 helix - blue.

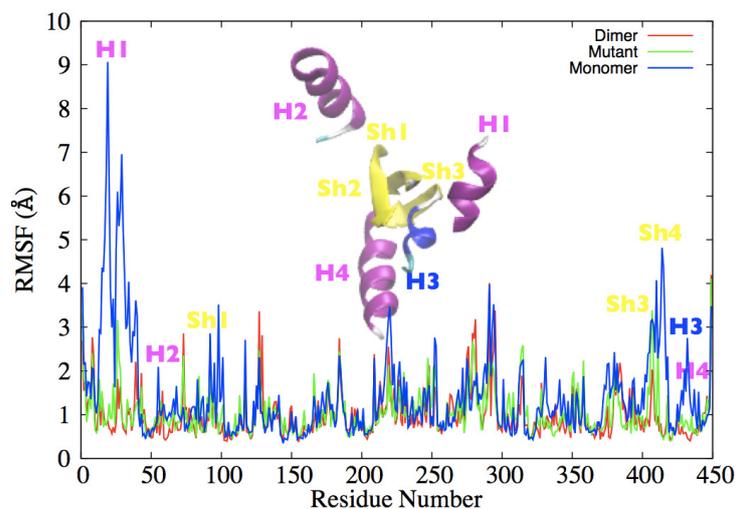


Figure 43: RMSF values of the dimer, mutant and monomer, with labels at the interface of the monomeric sub-unit. Alpha helices are labeled as H₁, H₂, H₃ and H₄ while beta sheets are labeled as Sh₁, Sh₂ and Sh₃.

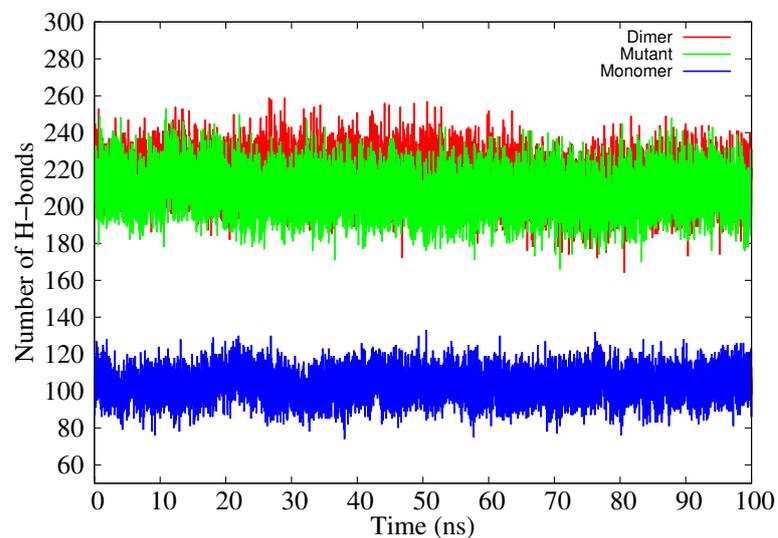


Figure 44: Number of hydrogen bonds as a function of time in the dimer, the mutant and the monomer calculated from 100ns MD simulations.

Analysis of the hydrogen bond occupancies shows that in the monomer, the hydrogen bonding network across the interface is grossly distorted, as manifested by decreased occupancies of the important hydrogen bonds. Some of those Hbonds that show high occupancies in the dimer are completely lost in the monomer, and some have significantly reduced occupancies (Figure 46).

The importance of the T59 residue at the interface was studied by experimentalist and the mutant T59R was characterised under the physiological conditions and was found to be monomeric [15]. From our MD simulations, we see that T59->D55 hydrogen bond is indeed very strong in the dimer, in contrast to the monomer where the average occupancy of this hydrogen bond is reduced significantly (Figure 46). The hydrogen bonding network of the residues at the (former) interface (especially H3 helix), and the active site of the monomer, is disrupted. The exception is the residues coordinating the Mg ion (especially E322) that have only slightly affected hydrogen bond occupancies resulting in the relatively stable coordination dynamics of the Mg ion (Figure 46).

The hydrogen bonding network in the active site is altered upon separation to monomers; A highly occupied Y402->D330 hydrogen bond, that "connects" the active site and the dimer interface, is broken in the monomer. The beta sheet of the interface - Sh2 (417-422) is connected to the crown domain by a flexible loop that is located right above the active site. This, so called "crown domain loop" is "capping" the active site by providing an electrostatic environment required for the stabilisation of the active site Zn1 ion. The Zn1 ligand, H412 is also located on that loop. It can be speculated that the motion of this loop, and hence the flexibility of the interface it is connected to, plays an important role in the modulation of the metal ion coordination geometry. In the monomer Y402 is no longer bound to the active site residue D330 that is located at the alpha helix together with two other catalytic Aspartic acid residues (D327 and D331) (Figure 47). Due to the fact that the interface, the crown domain and the active site are connected via the "capping" loop, as the interface becomes flexible Y402->D330 hydrogen bond breaks resulting in the alteration of the active site dynamics (Figure 48).

In order to understand differences between the dimer and the monomer, we have examined the hydrophobic and hydrogen bonding interactions, as well as the salt bridges all across the interface (Figure 45). Analysis of the hydrogen bond occupancy is performed on the mutant and hydrogen bonds with an occupancy of 60% or higher are examined further. Upon mutation the extensive hydrogen bonding net-

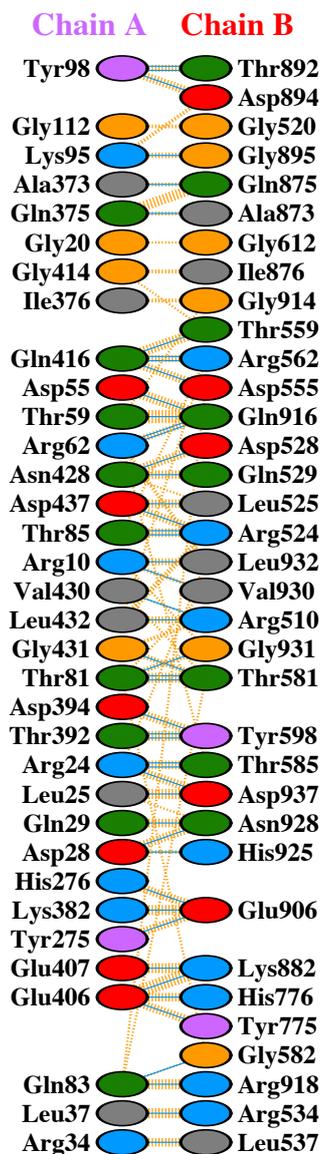


Figure 45: Residue-residue interactions across the interface coloured by residue type: Light blue - positive (H,K,R), red - negative (D,E), green (S,T,N,Q) - neutral, gray (A,V,L,I,M) - aliphatic, purple (F,Y,W) - aromatic, orange (P,G) - Proline and Glycine, yellow C - cysteine. The joining lines are coloured light blue for hydrogen bonds and orange for non-bonded contacts. Diagram is generated only for the residues at the interface of each chain, contributing to the hydrogen bonding network, as defined by the PDBSum [44].

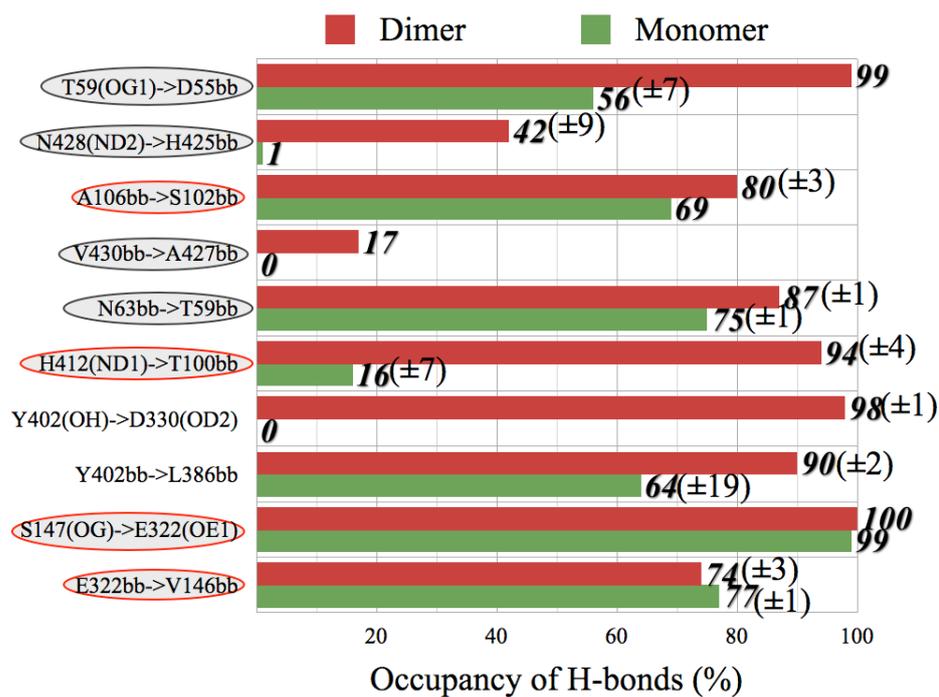


Figure 46: Occupancies of hydrogen bonds of the residues in the active sites (circled) of the dimer (red) and the monomer (green), as well as Hbonds at their (former) interfaces. Arrows point from hydrogen donors to hydrogen acceptors. Residues interacting with the backbone atoms are marked with “bb”, while the side chain atoms are given in parentheses. Averages and standard deviations are calculated from 2 independent MD runs.

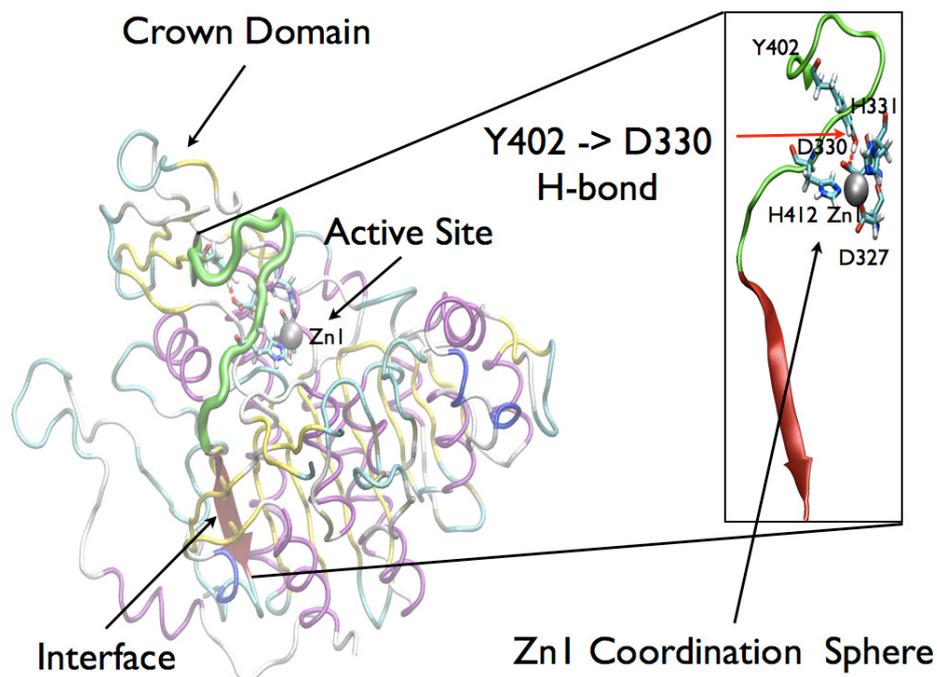


Figure 47: Single subunit of the dimer with the active site and the Y402->D330 hydrogen bonds highlighted. Sh2 interfacial beta sheet is thicker and coloured in red while the “crown domain loop” is shown in green. Zn1 and the residues in its coordination sphere are shown in licorice representation and coloured by atom type.

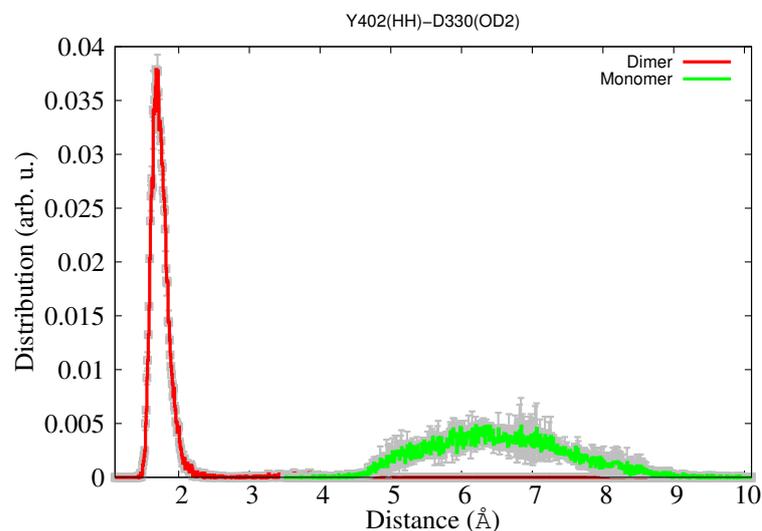


Figure 48: Distance distributions of the T402-D330 hydrogen bond calculated from the distances between the hydrogen atom of Y402 and its hydrogen bonding partner, Oxygen atom of D330.

work across the interface is disturbed as manifested by decreased occupancies of important hydrogen bonds when compared to the same hydrogen bonds in the native dimer (Table 8).

Mutation of the T59 and T559 at the interface to the corresponding Arginine residues, results in 0% occupancy of the hydrogen bond between the side chain of T59 and the backbone of D55 that has near 100% occupancy in the wild type dimer and about 55% occupancy in the monomer. Inter-subunit hydrogen bonds such as D916bb->T59, L25bb->D937 and R24->T585 have more than 70% occupancy in the dimer, whereas they are completely absent in the mutant and in the monomer (Table 8).

4.2.2.4 *Active Site Dynamics*

The geometry of the active site metal ions and the electrostatics provided by them is crucial for the catalytic activity of Alkaline Phosphatase [35]. Binding of the metal ions takes place only upon dimerisation [52], hence the monomeric form of AP should not be able to bind zinc and magnesium ions in a fashion that allows the enzyme to fulfil its catalytic functionality. Significantly decreased metal binding affinity was also observed experimentally in the mutant T59R that was found to be monomeric [15].

To understand the dynamics of the key residues in the active sites of the dimeric, mutant and monomeric forms of AP we measured the distances between the metal ions and the coordinating ligands. Coordination of the metal ions is monitored by following the time-evolution of the distance between the metal ions and the amino acid residues within 2.49Å of the zinc and within 2.95Å of magnesium ions during the last 80ns production run. All the amino acid residues that are within the coordination sphere of the metal ions during the production run are assumed to be strongly coordinated to the corresponding metal ion and hence, weakly affected by the mutation or the separation to monomers.

The distances between Zn1 and Mg metal ions in the monomer increase during the course of the MD simulations, and the distance distribution plot becomes wider, than the one for the dimer, thereby indicating the flexibility of the Zn1 metal ion and its tendency towards the departure from the active site. A closer look at the distribution of the distances between metal ions, shows that in the dimer the distribution of the Zn1-Zn2 distances is different from the one in the monomer, while the distance between Zn2-Mg becomes smaller in the monomer than in the dimer. The distance between Zn1 and Mg ions in the monomer is larger, than that between Zn2 and Mg,

Table 8: Hydrogen bond occupancies of the residues at and across the interface, computed over the simulation time of last 80ns for the dimer, the mutant and the monomer.

Donor -> Acceptor	Dimer (%)	Mutant (%)	Monomer (%)
THR436-Main-N ASP434-Side-OD2	99	96	16
THR436-Main-N ASP434-Side-OD1	0	1	81
THR436-Side-OG1 ASP434-Side-OD2	99	0	16
ARG418-Side-NH1 ASP76-Side-OD1	71	51	0
LYS114-Side-NZ ASP434-Side-OD1	79	68	51
LYS73-Side-NZ GLU66-Side-OE2	16	7	39
GLN83-Main-N GLY431-Main-O	54	59	0
VAL430-Main-N ALA427-Main-O	17	7	0
VAL429-Main-N ALA426-Main-O	14	23	6
HSD425-Main-N SER38-Main-O	14	0	4
ASN428-Side-ND2 HSD425-Main-O	36	28	0
TYR922-Side-OH MET504-Main-O	98	97	-
ARG10-Side-NH1 VAL930-Main-O	89	0	-
ARG510-Side-NH1 LEU432-Main-O	92	75	-
ARG510-Side-NH1 VAL430-Main-O	84	85	-
ARG34-Side-NH2 ASP539-Side-OD1	43	35	-
ARG24-Main-N ASP937-Side-OD2	87	0	-
ARG524-Main-N ASP437-Side-OD2	91	37	-
ARG562-Side-NH2 GLN416-Side-OE1	88	52	-
ARG62-Side-NH1 GLN916-Side-OE1	78	98	-
GLY431-Main-N THR581-Side-OG1	54	57	-
THR581-Side-OG1 THR81-Main-O	50	63	-
THR81-Side-OG1 THR581-Main-O	53	19	-
GLU406-Main-N LYS882-Main-O	57	19	-
GLU906-Main-N LYS382-Main-O	77	62	-
LEU525-Main-N ASP437-Side-OD1	83	0	-
GLN375-Side-NE2 ALA873-Main-O	14	0	-
ASN63-Main-N THR59(ARG)-Main-O	88	93	73
THR59(ARG)-Side-OG1 ASP55-Main-O	99	0	47

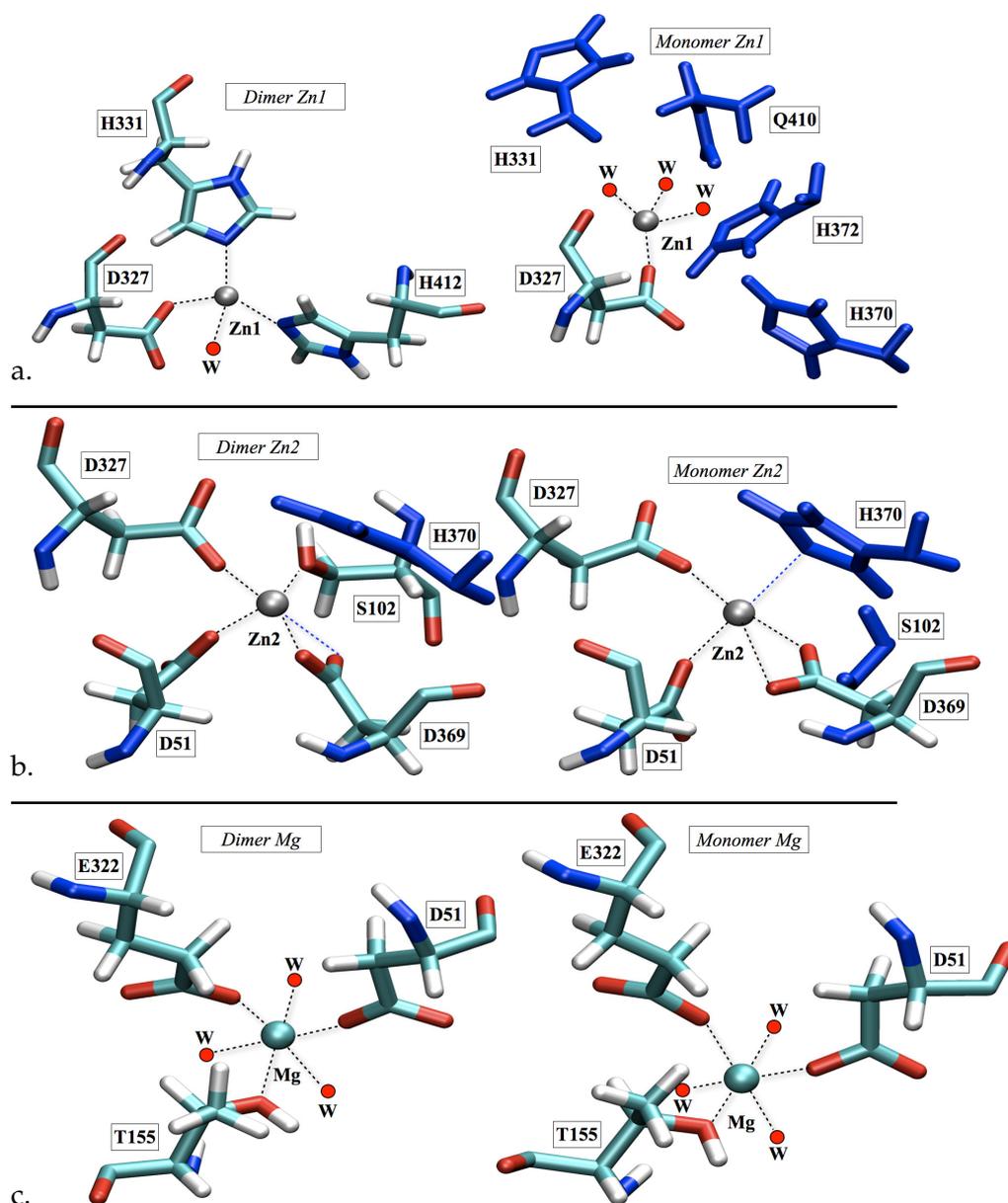


Figure 49: Coordination structure of (a.) Zn1 , (b.) Zn2, and (c.) Mg metal ions from a typical MD simulation snapshot of the dimer (left) and the monomer (right). The ligands of the metal ions are shown in licorice representation and coloured by the atom type, whereas the oxygen atoms of the water molecules coordinating the metal ions are shown as red balls. The residues that are present within the coordination sphere of one MD run and absent in another run, or present in one snapshot and absent in others - are shown in blue.

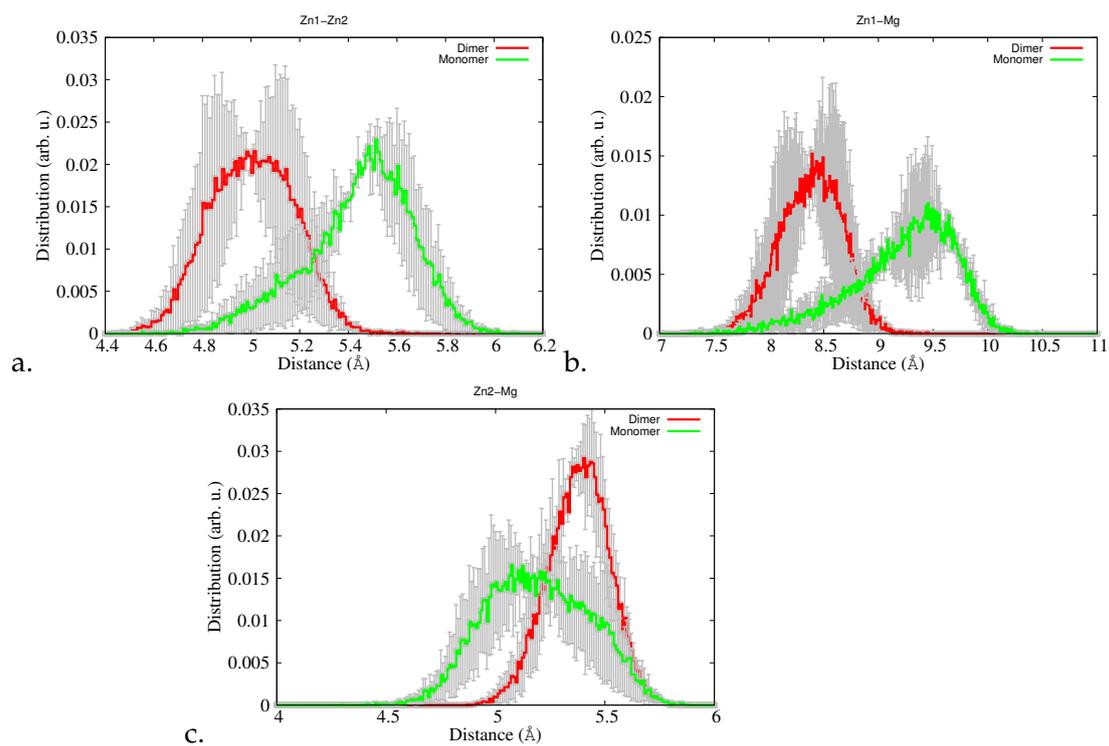


Figure 50: Intra-metallic distance distributions between Zn1 and Zn2 (a.), Zn1 and Mg (b.) and Zn2 and Mg ions (c.) and averaged over two independent MD runs.

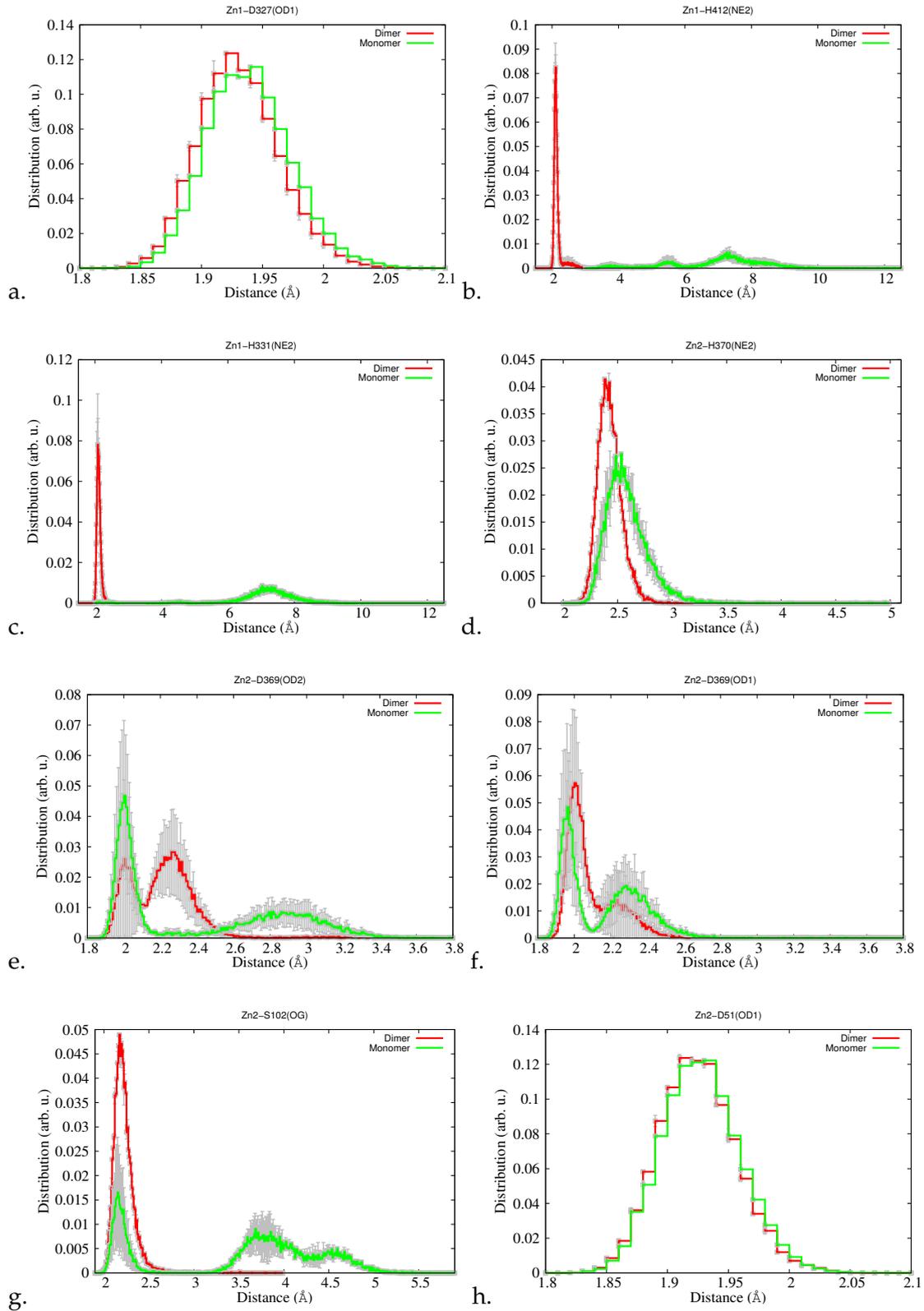


Figure 51: Distance distributions calculated from the distances between Zn1 and Zn2 ions and the corresponding atoms of their coordination shell residues: Zn1-D327(OD1) (a.), Zn1-H412(NE2) (b.), Zn1-H331(NE2) (c.), Zn2-H370(NE2) (d.), Zn2-D369(OD2) (e.), Zn2-D369(OD1) (f.), Zn2-S102(OG) (g.), Zn2-D51(OD1) (h.).

in contrast, those two distances in the dimer are comparable (Figure 50). Visual inspection of MD trajectories of the monomer confirms that the Zn1 ion becomes very flexible and moves further away from the active site. From the distance distribution plots, one can see that, during the 100ns time scale, the Zn2 and Mg ions are less flexible and dynamically more stable than Zn1 (Figure 51). Wider distance distribution plots of the monomer indicate increased flexibility of the active site residues resulting in larger distances between the metal ions and the nitrogen and oxygen atoms of the coordinating amino acids (Figure 51).

The coordination structure of Zn1 in the monomer, is very different from that in the dimer, except for the D327 that remains strongly coordinated to the Zn1 ion, in the dimer as well as the monomer, throughout the entire MD simulation time (Figure 49 (a.)). The distances of the nitrogen atoms of the residues H331 and H412, from the Zn1 ion are substantially greater in the monomer, than in the dimer, and the distribution of the distance is noticeably wider due to the increased flexibility of those residues in the monomer (Figure 51 (c.) and (d.)). This can be interpreted in terms of the weakening of coordination to H331 and H412 in the monomer; When Histidine residues leave the coordination sphere, they are replaced by water molecules to maintain tetrahedral coordination of the zinc ion (Figure 49 (b.)).

In the mutant, the distance between the atoms of the residues in the proximity to the interface region shows greater flexibility as compared to the dimer. Some important salt bridges are broken and formed during the course of the simulation time, while others are almost unaffected by the mutation or separation into monomer. The salt bridges between oxygen and nitrogen atoms of some pairs of residues remain stable throughout the course of 100ns production run in all 3 models, whereas time evolution of the salt bridges D408-K272 and E354-K357 shows that sometimes the monomer exhibits dynamics similar to the native dimer and sometimes to that of the mutant. Dynamics of the salt bridges is altered most prominently at the interface region, where important interfacial salt bridges are broken, thereby destabilising the monomer and effecting dynamics of the mutant (Figure 53).

In all models the coordination structure of Zn2 is more stable, as it remains coordinated to D327, D51, H370 and D369 throughout the entire simulation time for both, dimer and monomer. Visual inspection of the trajectories shows that the catalytic residue S102 in the

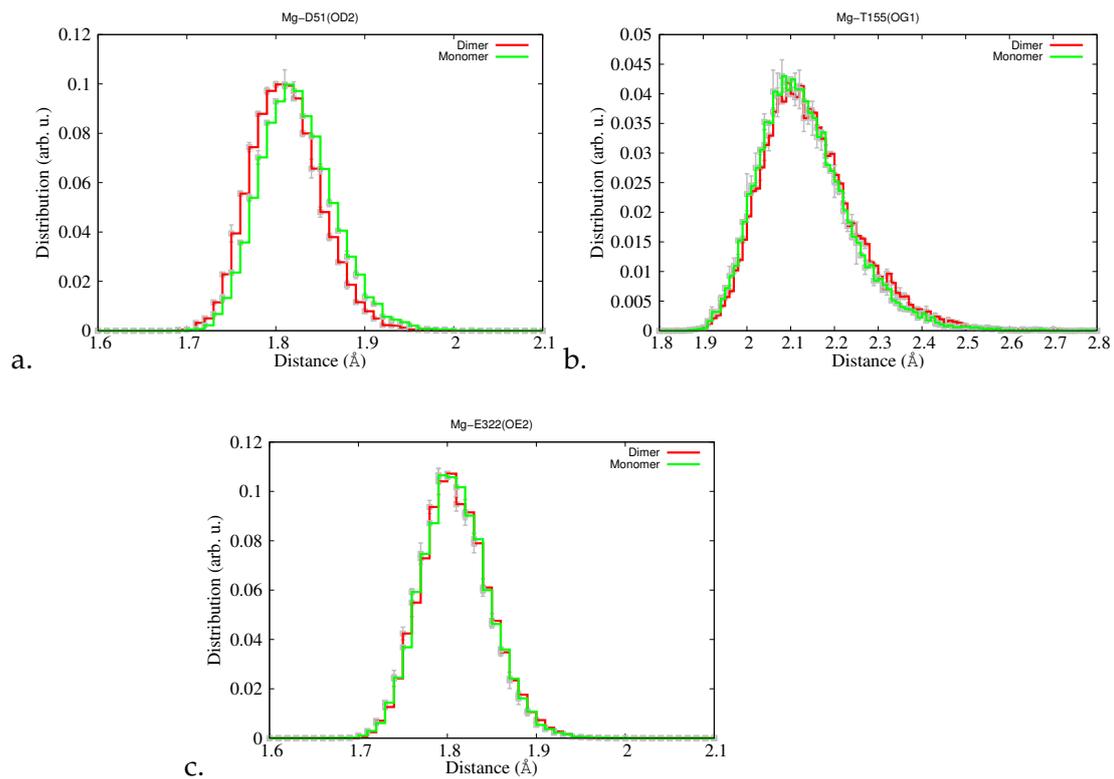


Figure 52: Distance distributions calculated from the distances between Mg ion and the corresponding atoms of its coordination shell residues: Mg-D51(OD2) (a.), Mg-T155(OG1) (b.), Mg-E322(OE2) (c.).

Table 9: Ensemble average of the distances between metal ions and the coordinating amino acid residues along with respective standard deviations in Å. In all cases the simulation length was 100ns and only the last 80ns were used to calculate the averages. PDB refers to the coordinates of the crystal structure with ID: 1ED9.

Distance	Crystal Structure	MD Simulations				
		Dimer (MD1)	Dimer (MD2)	Monomer (MD1)	Monomer (MD2)	
Zn1-D327(OD2)	2.47	1.93(0.03)	1.93(0.03)	1.94(0.03)	1.94(0.03)	
Zn1-H331(NE2)	2.18	2.09(0.04)	2.15(0.1)	7.30(1.06)	7.12(0.82)	
Zn1-H412(NE2)	1.99	2.11(0.05)	2.19(0.18)	7.64(1.35)	6.23(1.24)	
Zn2-D369(OD2)	3.17	2.14(0.24)	2.29(0.11)	2.29(0.41)	2.65(0.43)	
Zn2-D369(OD1)	1.97	2.19(0.21)	2.02(0.07)	2.2(0.2)	2.06(0.18)	
Zn2-H370(NE2)	2.11	2.42(0.10)	2.43(0.11)	2.57(0.19)	2.62(0.18)	
Zn2-D51(OD1)	2.10	1.92(0.03)	1.93(0.03)	1.93(0.03)	1.93(0.03)	
Zn2-S102(OG)	1.90	2.22(0.12)	2.22(0.13)	3.16(0.97)	3.86(0.67)	
Mg-E322(OE2)	1.91	1.81(0.04)	1.81(0.04)	1.81(0.04)	1.81(0.04)	
Mg-D51(OD2)	2.14	1.81(0.04)	1.81(0.04)	1.82(0.04)	1.82(0.04)	
Mg-T155(OG1)	2.27	2.14(0.11)	2.15(0.11)	2.13(0.1)	2.14(0.11)	
Zn1-Zn2	4.26	5.11(0.13)	4.9(0.15)	5.5(0.17)	5.33(0.22)	
Zn1-Mg	7.00	8.56(0.2)	8.21(0.24)	9.33(0.32)	9.12(0.57)	
Zn2-Mg	4.60	5.42(0.12)	5.33(0.14)	5.28(0.22)	5.08(0.2)	

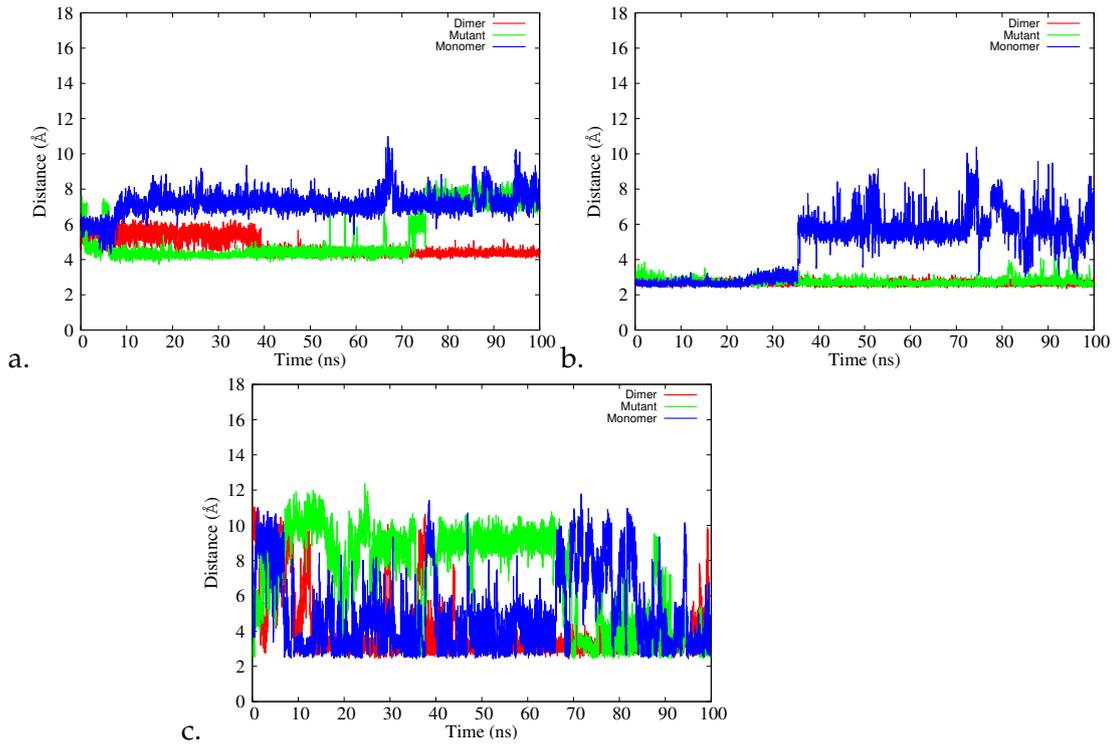


Figure 53: Time-dependent variation of the salt bridge between oxygen and nitrogen atoms of the interfacial residues D76 and R418 (a.), D434 and K114 (b.) and E66 and K73 (c.).

Distance	Dimer	Mutant	Monomer
Zn1-D327(OD1)	3.71(0.13)	3.68(0.167)	3.91(0.12)
Zn1-D330(OD1)	7.54(0.29)	7.4(0.28)	6.07(0.57)
Zn1-D330(OD2)	9.17(0.28)	9.12(0.26)	4.36(0.66)
Zn1-H370(NE2)	4.47(0.22)	4.66(0.26)	5.12(0.3)
Zn1-H372(NE2)	2.71(0.24)	4.91(0.27,)	4.22(0.65)
Zn1-D408(OD1)	14.75(0.98)	15.83(1.07)	9.24(3.05)
Zn1-D408(OD2)	14.57(1.05)	15.75(1.03)	8.9(2.86)
Zn2-D51(OD2)	3.88(0.1)	3.84(0.12)	3.75(0.13)
Mg-E322(OE1)	3.56(0.16)	3.47(0.1)	3.4(0.1)
Mg-D51(OD1)	3.9(0.08)	3.95(0.07)	3.81(0.13)

Table 10: Ensemble average of the distances between metal ions and the coordinating amino acid residues along with respective standard deviations in Å. In all cases the simulation length was 100ns and only the last 80ns were used to calculate the averages.

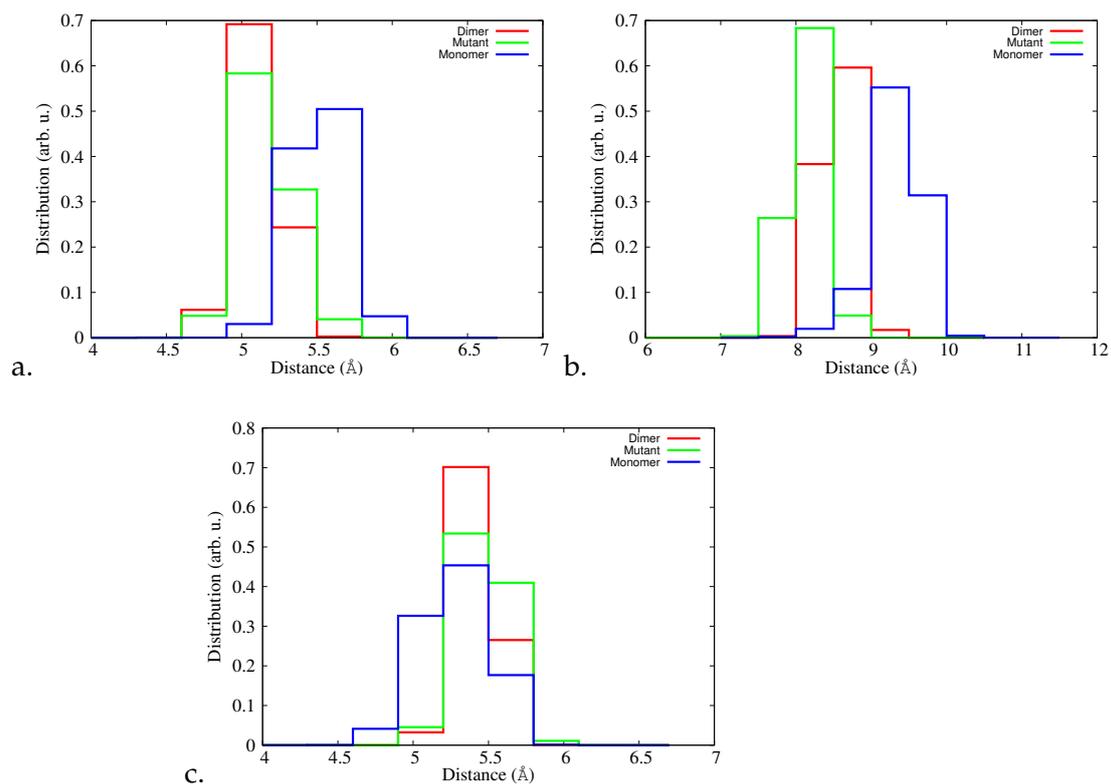


Figure 54: Time-evolution of the inter-metallic distances as a function of simulation time (right) and corresponding distance distributions, calculated from 80ns trajectories of MD simulations. Zn1-Zn2 (a.), Mg-Zn1 (b.), Mg-Zn2 (c.).

active site of the monomer, moves away from the Zn₂ ion by about 1.5Å and is no longer considered to coordinate the zinc ion. However, from the shape of the S₁₀₂(OG)-Zn₂ distance distribution plot, one can see that S₁₀₂ samples at least two distinct conformations in the monomer and spends part of the simulation time coordinated to the Zn₂ ions (Figure 51 (g.)).

The Mg coordination to E₃₂₂, D₅₁ and T₁₅₅ and oxygen atoms of the three neighbouring water molecules is remarkably stable, both in the dimer and in the monomer (Figure 52). D₅₁ is an important residue due to the fact that its oxygen atoms are coordinating both, Zn₂ and Mg metal ions. The distance distribution plot shows that those distances remain stable within two runs and in both models (Figure 51).

From the distance distribution plots (Figure 54) one can see that upon mutation coordination structure of metal ions changes and the dynamics of the key residues resembles more that of the monomer, than of the dimer. Some residues remain coordinated while others are substituted by their neighbouring amino acids or the water molecules. Average distances between metal ions, and those residues that are not directly coordinating them, are shown in Table 10.

In all three models, the D₃₂₇ ligand remains strongly coordinated to the Zn₁ ion, throughout the entire simulation time. The coordination structure of Zn₁ in the monomer, is very different from the WT dimer. In the monomer coordination to H₃₃₁ and H₄₁₂ is weak, as a result, those Histidines leave the coordination sphere and are replaced by water molecules to maintain tetrahedral coordination of the Zinc ion (Figure 57). The distance of the nitrogen atoms of the residues H₃₃₁ and H₄₁₂, from the Zn₁ ion is substantially greater in the monomer, than that of the dimer and the mutant, and the distribution of the distance is noticeably wider due to the increased flexibility of those residues in the monomer (Figure 56 (c.) and (f.)).

The coordination structure of Zn₂ in all three models is more stable, as it remains coordinated to D₃₂₇, D₅₁, H₃₇₀ and D₃₆₉ ligands throughout the production dynamics. In the mutant, however, D₅₁ moves further away from Zn₂, letting S₁₀₂ enter its coordination sphere. Interestingly, the catalytic residue S₁₀₂ of the monomer moves away from the Zn₂ ion by almost 1Å, enough to be no longer considered to be coordinated to the zinc ion (Figure 54).

From the distance distribution plots, we see that the Zn₂ and Mg ions are less flexible and dynamically more stable than Zn₁ (Figure 56). Wider distance distribution plots of the monomer indicate an increased flexibility of the active site residues resulting in larger dis-

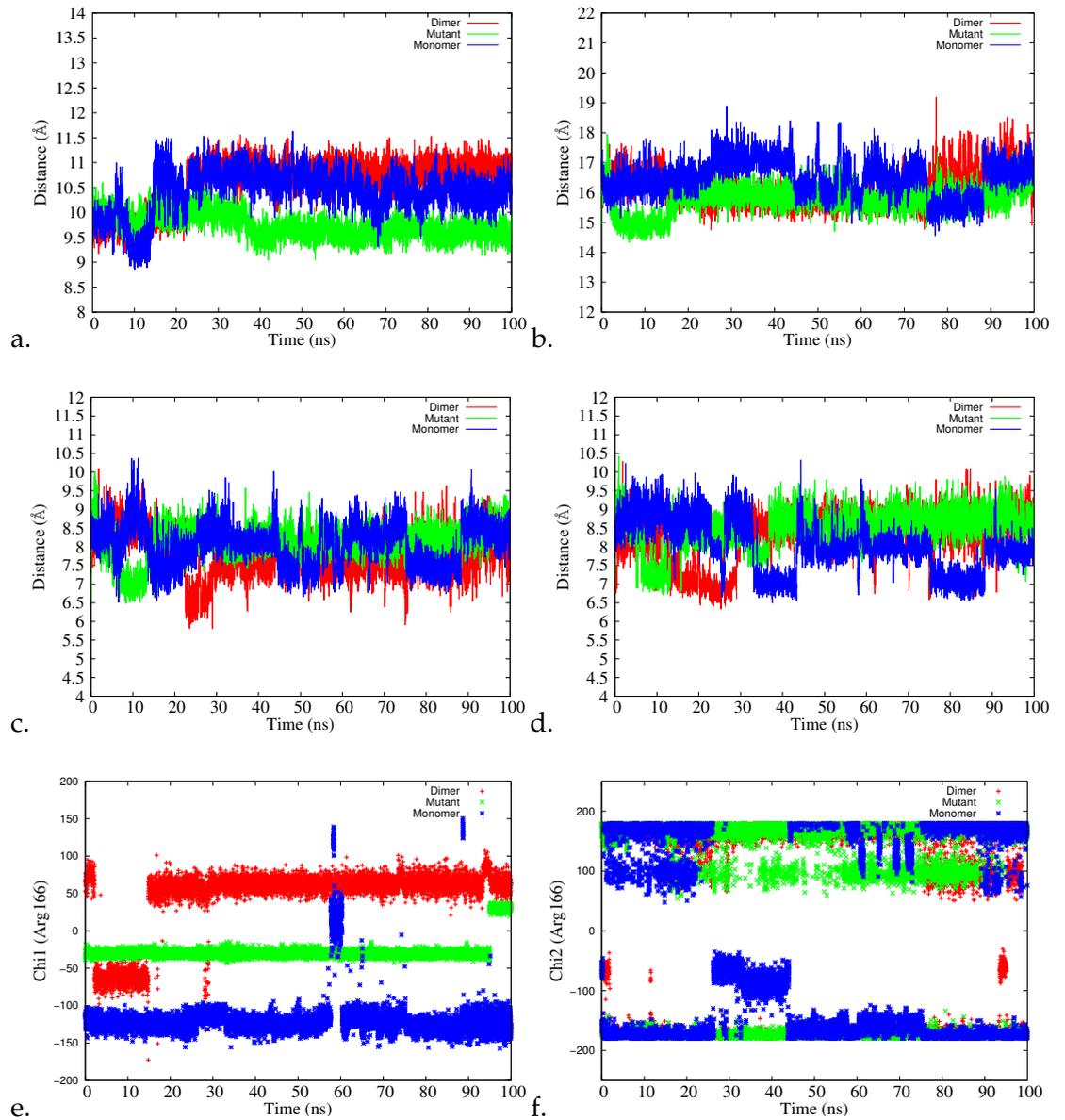


Figure 55: Time-trajectories of the centre of mass distances between subunit A and the residue S102 (a.), subunit A and residue R166 (b.) and the residue R166 and residue S102 (c.); residues R166 and D101(d); Time-trajectories of the Chi1 (e.) and Chi2 (f.) dihedral angles of the residue R166.

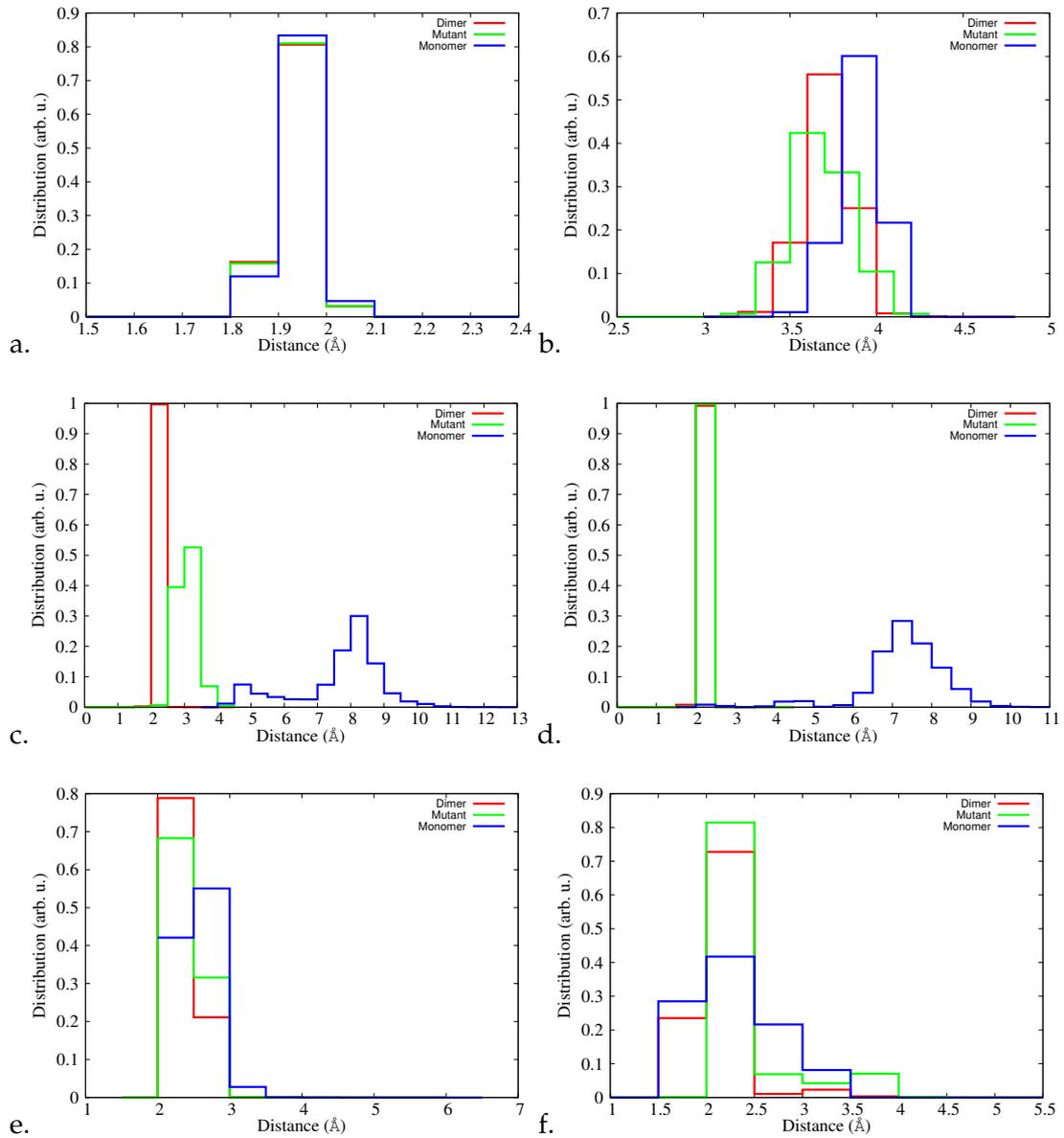


Figure 56: Distance distributions calculated from the distances between Zn1, Zn2 and Mg ions and the corresponding atoms of their coordination shell residues of the T59R mutant: Zn1-D327(OD1) (a.), Zn1-D327(OD2) (b.), Zn1-412(NE2) (c.), Zn1-331(NE2) (d.), Zn2-H370(NE2) (e.), Zn2-369(OD2) (f.).

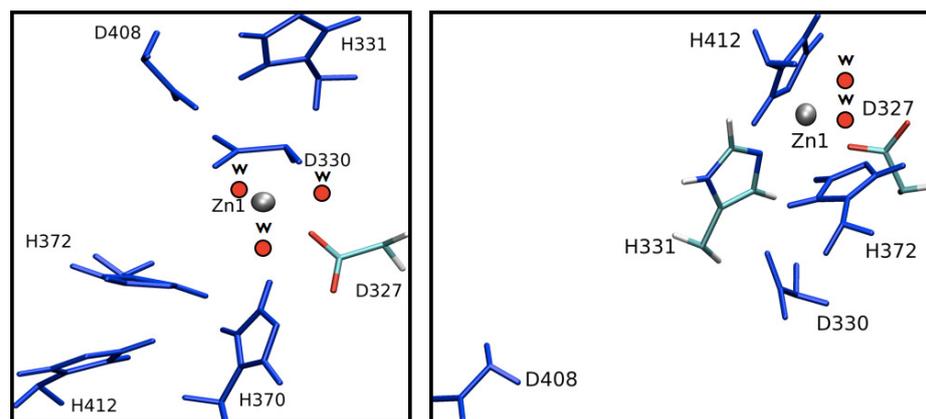


Figure 57: Zn1 metal ion coordination sphere in the last frame of the monomer (left) and the mutant (right) shown in licorice representation and coloured by the atom type. Amino acids that are in the proximity of the metal ions, but not directly coordinated to them are shown in blue.

tances between the metal ions and the nitrogen and oxygen atoms of the coordinating amino acids (Figure 54).

A closer look at the distribution of the distances between metal ions, shows that in the dimer the distribution of the Zn1-Zn2 distances is different from that of the monomer and the mutant, while the distribution between Zn2-Mg distances is similar in all three models (Figure 54). The distance of Zn1 from Mg is larger, than that of Zn2, and the distribution of this distance is noticeably wider due to the flexibility of the Zn1 metal ion (Figure 54). Visual inspection of MD trajectories of a monomer shows that the Zn1 ion becomes very flexible and moves further away from the active site. The monomer seems to sample Zn1-Zn2 distances in a manner more similar to the mutant, than to the dimer.

Part V
DISCUSSION

DISCUSSION

5.1 THE HOLO AND APO FORMS OF THE ENZYME

Comparative analysis of the dynamics of individual subunits within the dimer can help interpreting the experimental proposal of negative cooperativity in AP. Comparison of the two subunits of the enzyme shows that the dynamic behaviour of the two subunits is indeed different from what one would expect based on their structural symmetry. There are subtle differences in hydrogen bonding patterns of individual subunits and across the interface between them. Due to the negative cooperativity, binding affinities to the product and cofactors should be different in subunits of a dimer during the course of MD simulations. Our results show different dynamics of the active site residue of the holo enzymes, demonstrated by an increased flexibility of one subunit and larger atomic distances, versus rigidity of the second subunit and shorter distances between the atoms of the key residues. Hence, one subunit has a tendency to “keep the product bound” instead of “letting it go” like its counter subunit. This result might be an indication of the asymmetry in the holo dimer; subunit A of the apo dimer shows flexibility to a degree similar to that of the subunit B of the holo dimer, whereas the subunit A of a holo dimer is more restricted, possibly due to the fact that its structure is stabilised by the product. Such an interpretation of the results goes hand in hand with the idea of the selection of the preexisting conformations by the substrate/product as opposed to the induced fit model where the conformational changes are assumed to take place upon binding. Indeed, already in the 100ns simulation time, we see significantly different degrees of flexibility in the two subunits, indicating an intrinsic asymmetry of the enzyme rather, than asymmetry induced upon binding. Such an asymmetry of the subunits, is profound near the so called “wing” region (especially the residue numbers 210 to 240) and the crown domain (residue numbers 270 to 290 and 390 to 410) (Figure 23). The increased flexibility of these regions in the enzyme might be linked to the substrate binding. In the presence of an inhibitor, such as inorganic phosphate, the activity of AP becomes highly regulated, the active site is no longer available for free binding of substrates and the “closed” enzyme becomes inactivated. Binding an inhibitor stabilises the inactive form of the enzyme. However, as the inhibitor releases, the enzyme turns back to its “open” state and the active site is available for substrate to bind and form the product.

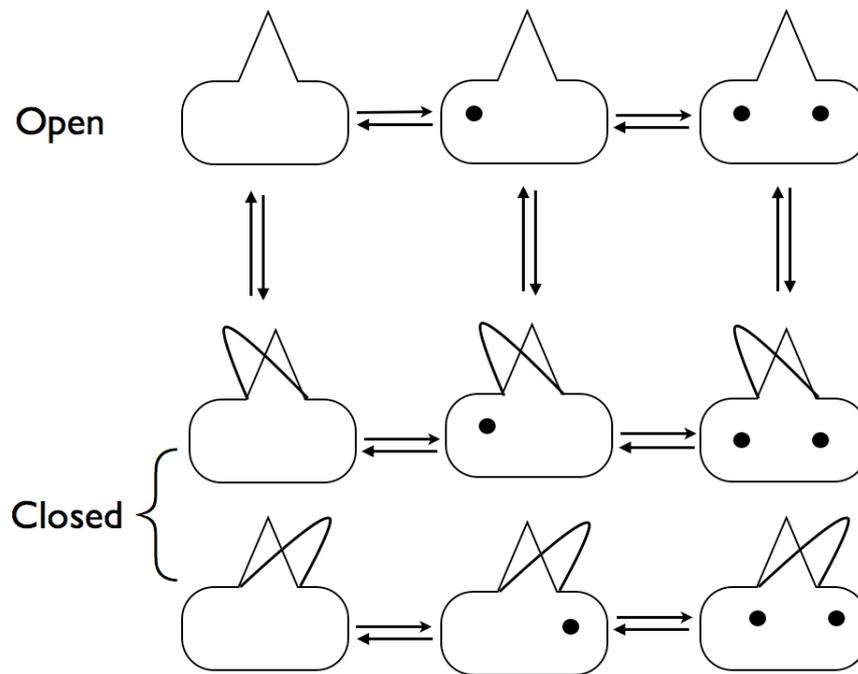


Figure 58: Cooperative binding mechanism of AP.

The subunits of an allosteric enzyme fit together in a way that a conformational change in one subunit is transmitted to another, through this interaction, an inhibitor molecule that binds to one regulatory site can affect the active sites of both subunits. Our results, obtained from ENM, using normal mode analysis suggest a mechanical coupling between the motions of individual subunits and their active sites in the wild type dimer, via the dimer interface. Correlated motions of the monomeric units of the dimer, seem to play an important role in the functionality of Alkaline Phosphatase. The global dynamics of the monomer, describes opening and closing of the active site, a motion that is correlated to the motion of the “former” interface. In the dimer, opening and closing of the active site is also present, however, most global motion in the dimer, is that of the subunits, with respect to the interface. Global dynamics results, can be interpreted, in terms of the proposed negative cooperativity in Alkaline Phosphatase. We see intrinsic asymmetry of the subunits within the dimer; when crown and wing regions of one subunit are performing opening and closing motions, another subunit is giving a “helping hand” by pushing in an opposite direction to give a mechanical momentum via the interface (Figure 58).

The function of AP is the catalysis of phosphate esters for the supply of the cell with important inorganic phosphate. The end product of the enzymatic reaction catalysed by AP is also an inhibitor, preventing AP from further binding to the substrate. When the concentration of the phosphate is high in the cell, and no additional phosphate is required, the active site of AP is “locked” and its functionality inhib-

ited. However, under the condition of cell phosphate starvation, or in the presence of phosphate acceptors, AP becomes functional again, as it transforms itself into a phosphate transfer enzyme and gives away inorganic phosphate from its active site and becomes available for the catalysis of more phosphate containing compounds, for a better supply of this essential chemical, playing multiple roles in the cell. In order to keep perfect balance and have a better control over two important conformations, open and closed, it is optimal for AP to have a dimeric quaternary structure with two subunits, each equipped with just enough catalytic machinery to be able to perform the catalysis, and communicate the information when the final product is made, to another subunit. The subunits are regulating each others function, through the rapid change in their conformation, enabling phosphate binding and release, of one subunit at a time. The feedback inhibition mechanism can be used to stop accumulations of additional inorganic phosphate and hence contribute to the biologically essential regulation of the activity of *E. coli* Alkaline Phosphatase.

5.2 THE WT DIMER, MONOMER AND T59R MUTANT OF THE APO ENZYME

It has been known for decades, that the enzyme *E. coli* Alkaline Phosphatase exists only in a dimeric quaternary structure and that its separation into monomers leads to the loss of the enzymes' biocatalytic functionality [15]. In order to understand why monomeric AP is not functional, although each subunit of the dimeric enzyme is equipped with its own catalytic site, we performed systematic analyses of MD simulations of AP in dimeric and monomeric forms. In addition, our MD simulation data shows that by introducing bulky and positively charged Arginine residue at the interface, the dimer becomes less stable and the electrostatic field induced by the mutation affects not only the interface stability, but also the overall dynamics of important residues in the active site. Even though the T59R mutant remains dimeric during the course of the 100ns MD simulation, its dynamic behaviour is more similar to that of an isolated monomer, than to the wild type dimer.

Analysis, of the overall and the interface dynamics, of MD simulations shows that the dimer is structurally more stable than the monomer (Figure 35). Increased flexibility of the monomer, most profound at the (former) interface region, can be explained by the lack of the stabilising factors, such as an undisrupted hydrogen bonding network and hydrophobic interactions.

The overall shape of the monomer, as manifested by the R_{gyr} , is changed due to the abundance of the hydrophobic residues at the interface (such as V429, V430, I58, I419, L33, L80, L417, L432, L446), being "exposed" to and trying to "escape" the solvent (Figure 37). The

hydrogen bonding network of the monomer lacks all together, those important stabilising interactions. Flexibility of the monomer is further illustrated by the RMSF values per residue, that are higher for the interface and terminal residues, than for the rest of the monomer, including the “wing” region and the “crown” domain, that are comparable in the dimer and the monomer models. This is in agreement with experimental studies, suggesting that the overall monomeric fold is stable, even though the catalytic activity of the monomer is significantly reduced [15]. High fluctuation of the interfacial residues can be understood as a result of the hydrophobic region of the monomer being exposed to the surrounding solvent instead of being stabilised by its counterpart.

The secondary structure content of the interface of the monomeric model is significantly altered, compared to the dimeric one, as shown from the changes in the secondary structure, taking place at the (former) interface region, during the course of the MD simulations. H3 that has a 3-10 helix-like secondary structure in the wild type dimer and the mutant, refolds into the alpha helix in the monomer and Sh2 beta sheet unfolds into the coil-like secondary structure (Figure 38).

The N-terminal loop connecting N-terminus with the first interfacial helix (H1) is also very flexible in the monomer. This loop is rich in Glycine and Alanine residues that contribute to its flexibility, in addition to two adjacent Arginine residues (R23 and R24) that are introducing destabilising positive charge. The importance of this loop has been reported previously, suggesting weakest resistance of the N-terminal loop to proteolysis with R10, R23 and/or R24 and R34 residues identified as sites of tryptic cleavage [11]. These Arginine residues also show hydrogen bonds with highest differences in the occupancies when apo and holo forms of the enzyme are compared (Table 4). This might be an indication of their involvement in the communication of the structural information between the subunits across the interface. The occupancy of the hydrogen bonds is a measure of catalytic competency of the enzyme. As we see from the H-bond analysis of the interface, these bonds play an important stabilising role that in turn ensures activity of the dimer. In the T59R mutant, occupancies of the hydrogen bonds between the side chain of R10 and the backbone of V930, side chains of R34 and D539, the backbone of R524 and the side chain of D437 are reduced significantly (Figure 46 (a.)). R10 and R34 are located near the “H1” helix that is exposed to the solvent and isolated from the rest of the monomer and that fluctuates most in the monomer simulation.

By introducing bulky and positively charged Arginine residue at the interface, the dimer becomes less stable and the electrostatic field induced by the mutation effects not only the interface stability, but also the overall dynamics of the important residues in the active site, in such a way that the mutant exhibits dynamic behaviour sim-

ilar to that of an isolated monomer, even though it remains dimeric during the course of the 100ns simulation time. From the comparative analysis of MD simulations of the wild type dimer, the T59R mutant and the monomer, we identified three inter-subunit hydrogen bonds: D916bb->T59, L25bb->D937 and R24->T585 that are unique to the dimeric fold. Further comparison of the MD simulations of the wild type dimer and the *in silico* monomer, reveals important hydrogen bonds such as T59->D55, N428->H425 and Y402->D330, that are twice as highly occupied in the dimer, than in the monomer (Figure 46).

The hydrogen bonding network across the interface plays an important stabilising role that in turn ensures activity of the native dimer. The hydrogen bond between Y402 and D330 is missing in the monomer, resulting in an increased flexibility of the active site residues and affecting the coordination dynamics of the Zn₁ metal ion; as a result the distance between Zn₁ and its ligands changes, allowing the zinc ion to be more flexible and move away from the active site more freely. Indeed, the Y402->D330 hydrogen bond is responsible for “linking” the active site with the interface via the so called “crown domain loop”, comprising Y402– Q416 residues and located just above the active site. The ligand of Zn₁ metal ion, H412, which appears very flexible in the monomer, is also located on this “active site capping loop”. Optimal positioning of the active site residues is ensured by the nearby Y402->D330 hydrogen bond, that is modulating the motion of the crown domain loop. It can be speculated, that when this important hydrogen bond is missing (like in case of the monomer), the crown domain loop, that is rich in negatively charged residues (E411, E407 and E406) becomes very flexible and the positioning of the active site residues is no longer optimal to ensure electrostatic stabilisation of the divalent zinc ions. Based on the comparative analysis of the apo and holo APs, it has been reported previously that the conformational changes taking place at metal binding sites are transmitted from the active site to the E406-N10 surface loop, which (based on the limited proteolysis and MALDI-TOF mass spectrometry studies) appeared disordered in the absence of metal ions [11]. Our results also support this idea, and propose a mechanism, through which the electrostatic destabilisation allows the otherwise immobile Zn₁ ion to move more freely and eventually escape the active site; without tightly bound zinc ions in the active site, no catalysis is possible and hence the monomer of AP becomes catalytically deactivated.

The differences in metal ion coordination dynamics between the dimer and the monomer, are most visible in the case of the Zn₁ ion that becomes very flexible in the monomer (Figure 49). The distance distribution plots are also noticeably wider, due the flexibility of the Zn₁ metal ion and the coordinating ligand atoms. Wider distance

distribution plots of the monomer indicate increased flexibility of the active site residues resulting in larger distances between the metal ions and the nitrogen and oxygen atoms of the coordinating amino acids. Amino acids involved in the octahedral coordination of the Mg ion, are only slightly further apart in the monomer and in the mutant, than in the wild type dimer. This is not surprising, as the magnesium ion, as well as the less flexible Zn₂ ion, are more deeply buried into the active site, whereas Zn₁ is located closer to the surface and hence, is more affected by the motion of the “crown domain loop”.

Analysis of the intra-metallic distances reveals dynamic features of the dimer that are missing in the monomer. This can lead to a reduction of catalytic activity of the monomer, as the proper positioning of the metal ions is directly linked to the functionality of the enzyme [42]. Structural and dynamic changes taking place in the monomer result in dynamic switches between two distinct conformational states: one similar to the native dimer, and another similar to the T59R mutant. Two distinct conformations are observed in the monomer, as it switches dynamically and samples intra-metallic distances closer to those of the dimer and of the mutant.

Conformational changes of individual residues in the active site allow us to establish a structural and dynamic fingerprint corresponding to the active form of the dimeric enzyme. Our MD simulation data suggest that mutation and separation to monomers affects the catalytic activity of AP by changing the overall architecture of the active site. Without tightly bound zinc ions in the active site, no catalysis is possible and hence the monomer of AP becomes catalytically deactivated.

Part VI
CONCLUSIONS

CONCLUSIONS

E. coli Alkaline Phosphatase has a dimeric quaternary structure with two juxtaposed subunits related with 2-fold axes of symmetry [42]. Comparative structural analysis of the apo and holo forms of this enzyme suggests the intrinsic asymmetry of the subunits, encoded by the dimeric shape, as reflected by very different flexibilities of the individual subunits of the dimer. NMA of low resolution, coarse-grained models, such as the ENM of AP in apo and holo forms suggests a mechanical coupling between the collective motions of individual subunits and their active sites, through the dimer interface. Together with the correlated motion of the subunits, in the opposite direction to the interface, opening and closing of one subunit at a time is also apparent.

The experimental proposal of negative cooperativity of the subunits is tested and explained by the analyses of MD simulations, that reveal subtle differences in the hydrogen bonding network of the subunits and dynamics of their active site residues, demonstrating a distinct asymmetric behaviour, with increased flexibility of one subunit, versus rigidity of a second one. Both, apo and holo forms of the enzyme, show asymmetry in the square fluctuations from NMA the coarse grained models and RMSF of the atomistic ones. Distances of the key residues in the holo enzyme clearly show the asymmetry. The active site of one subunit of the holo enzyme is wider, and more open, possibly: to let the product out, and make the space available for substrates to enter, and, to make the binding of the product in the counter subunit tighter, by releasing the statin and giving momentum required for the closing the active site in the counter subunit. Information about structural changes is transmitted between the subunits via the hydrogen bonding network across the interface. At least two such communication pathways can be proposed, based on the analysis of MD simulations, both via the interfacial alpha-helix containing residues T55 and T59.

Although each subunit of the enzyme is equipped with its own catalytic sites, a monomeric AP does not exist in nature and the engineered mutants have significantly reduced activity. Due to the absence of a crystal structure, a model of a monomeric form of the enzyme is constructed based on the crystal structure of the apo dimer. In addition, a T59R mutant is built, where interface residues T59 and T559 are substituted by a bulky and charged Arginine residues. Our MD simulation results confirm that the overall dynamic behaviour of the monomer is different from that of the corresponding dimer and resembles more that of the T59R mutant. Breaking of the hydrogen

bond between residues Y402 and D330, as observed in the MD simulations of the monomer, results in the incorrect positioning of the catalytically important, divalent, zinc ion, which explains the reduced enzymatic activity of the monomer. Our findings confirm that the structural stability of dimeric AP, provided by the hydrogen bonding network across the interface, is essential for the enzymatic activity. Understanding the nature of the correlated motions of the subunits within the dimer, and their connection to the enzymes' activity, is an important step in expanding our knowledge on structure-dynamics-function relationship of *E. coli* Alkaline Phosphatase and related enzymes.

Part VII
FUTURE WORK

FUTURE WORK

In this thesis, the interplay between structure, dynamics, and function of the enzyme Alkaline Phosphatase is investigated, using molecular modelling and simulations. Analysis of MD simulations in conjunction with the normal mode analysis of low resolution, elastic network models, provides means of understanding the importance of dimerisation in the enzymes and the advantage of dimeric enzymes over their monomeric counterparts. Despite the systematic knowledge gained by modelling and simulation of *Escherichia coli* AP there is still lot to be learned about, before completing our understanding of the relationship between quaternary structure and biocatalytic functionality of this important enzyme.

The global dynamics analysis of the coarse-grained models suggests a mechanical coupling between the motions of individual subunits with respect to the dimer interface. Elastic network models of the apo and holo enzymes, describe a well defined opening and closing of the active site. In order to gain a comprehensive insight into the functional significance of those motions, an analysis of the large amplitude motions should be carried out on an ensemble of different conformations. Due to the fact that *E. coli* AP is well studied and many wild type and mutant crystal structures are available, it is feasible to perform normal mode analyses on the ensemble of crystal structures using anisotropic network models. The ability of this simple computational approach for generating potential ligand-bound conformers starting from the apo state, has been reported previously [7]. Alternatively, such motions could be examined by performing principal component analyses on a sufficiently long trajectory from molecular dynamics simulations at atomic detail.

The primary structure of Alkaline Phosphatase from a cold-active, marine *Vibrio splendidus* (*Vibrio sp.*) has been reported to be monomeric and very similar to *E. coli* AP. Magnesium metal binding site of *Vibrio sp.* is different from that of the *E. coli* AP, whereas the two zinc metal binding sites are identical [3]. The amino acid sequence of *Vibrio sp.* can serve as a template for homology modelling. Mutating D153 and K328 of *E. coli* AP to H153 and W328 respectively, and modelling additional insert loops, as they are in *Vibrio sp.* AP, should enable the prediction of the structure of this yet unknown psychrophilic enzyme. Moreover, the comparison of the D153H and K328W models can be made with the crystal structures of the corresponding mutants from *E. coli* AP (1anj.pdb and 2anh.pdb, respectively), that are available from PDB [57].

Having learned more about the interface and the nature of communication between the subunits, from modelling and simulations, we can propose different mutations for destabilising the active site and the interface. In the experimental setup, site-directed mutagenesis is typically used to test the importance of proposed functional sites, which can be time-consuming, especially when it is not clear which mutations to make. In contrast, *in silico* mutation of the large number of residues is easy to do and takes significantly less time, while providing atomic level detail of the structural and dynamic changes caused by the mutations.

Part VIII
APPENDIX

APPENDICES

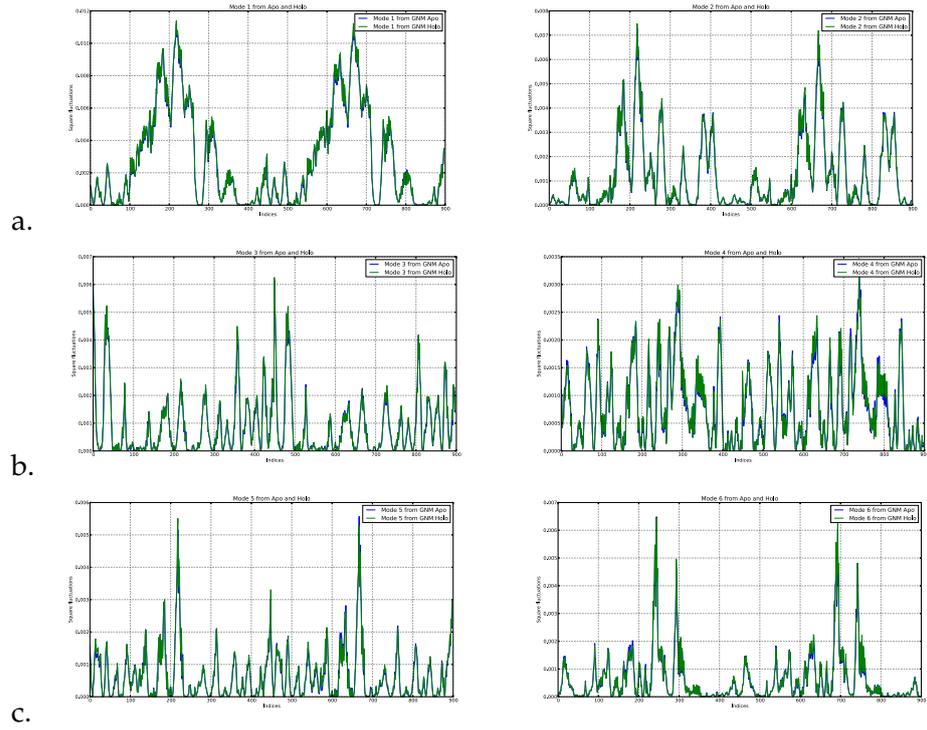


Figure 59: Square-fluctuations of the first 6 normal modes of holo and apo enzymes calculated from GNM.

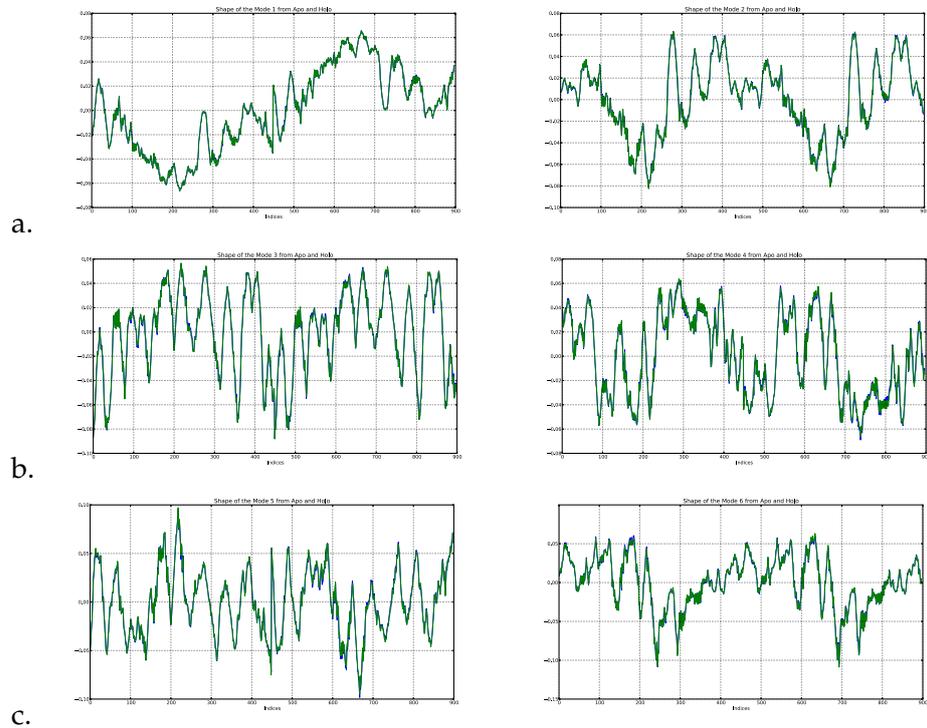


Figure 60: Shapes of the first 6 normal modes of the holo and apo enzymes calculated from GNM.

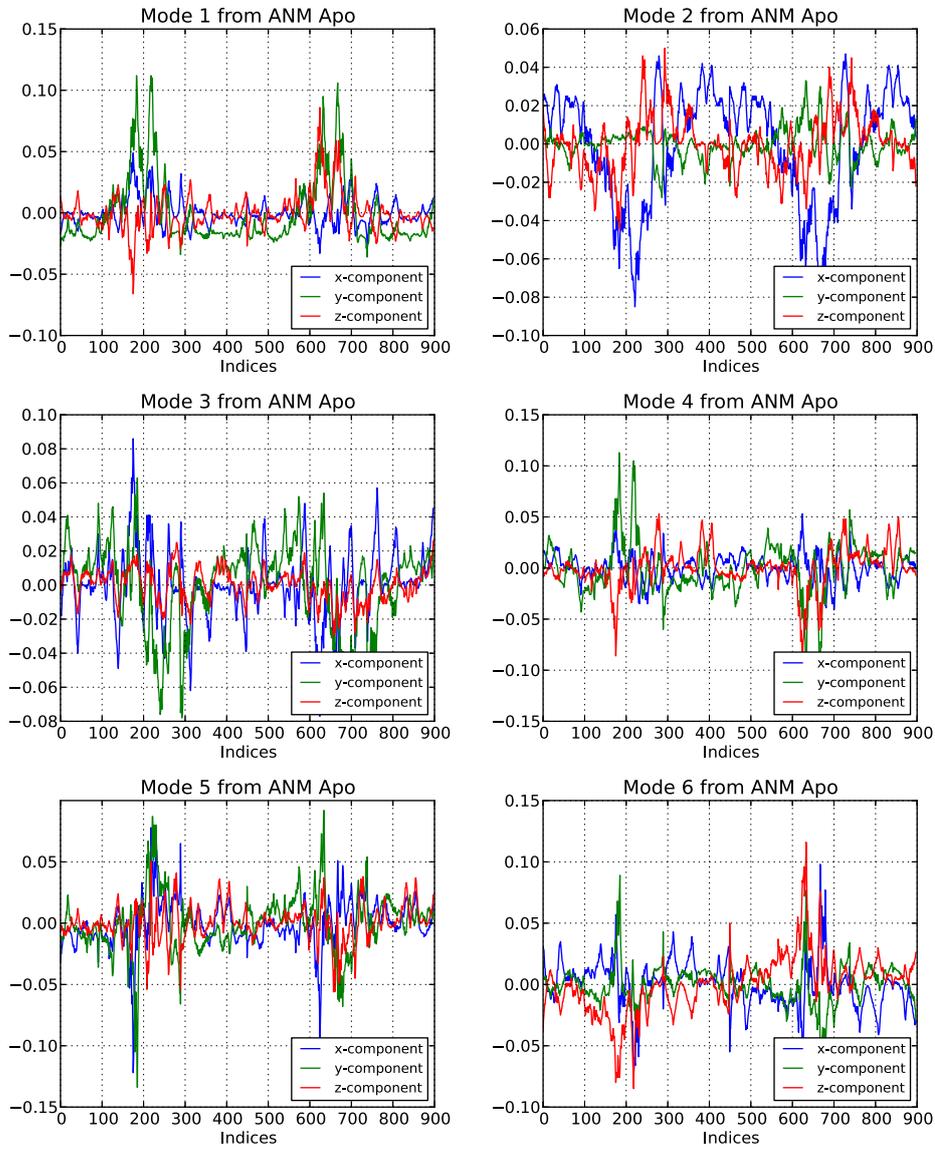


Figure 61: The mode array of the first 6 normal modes of the apo dimer based on the ANM.

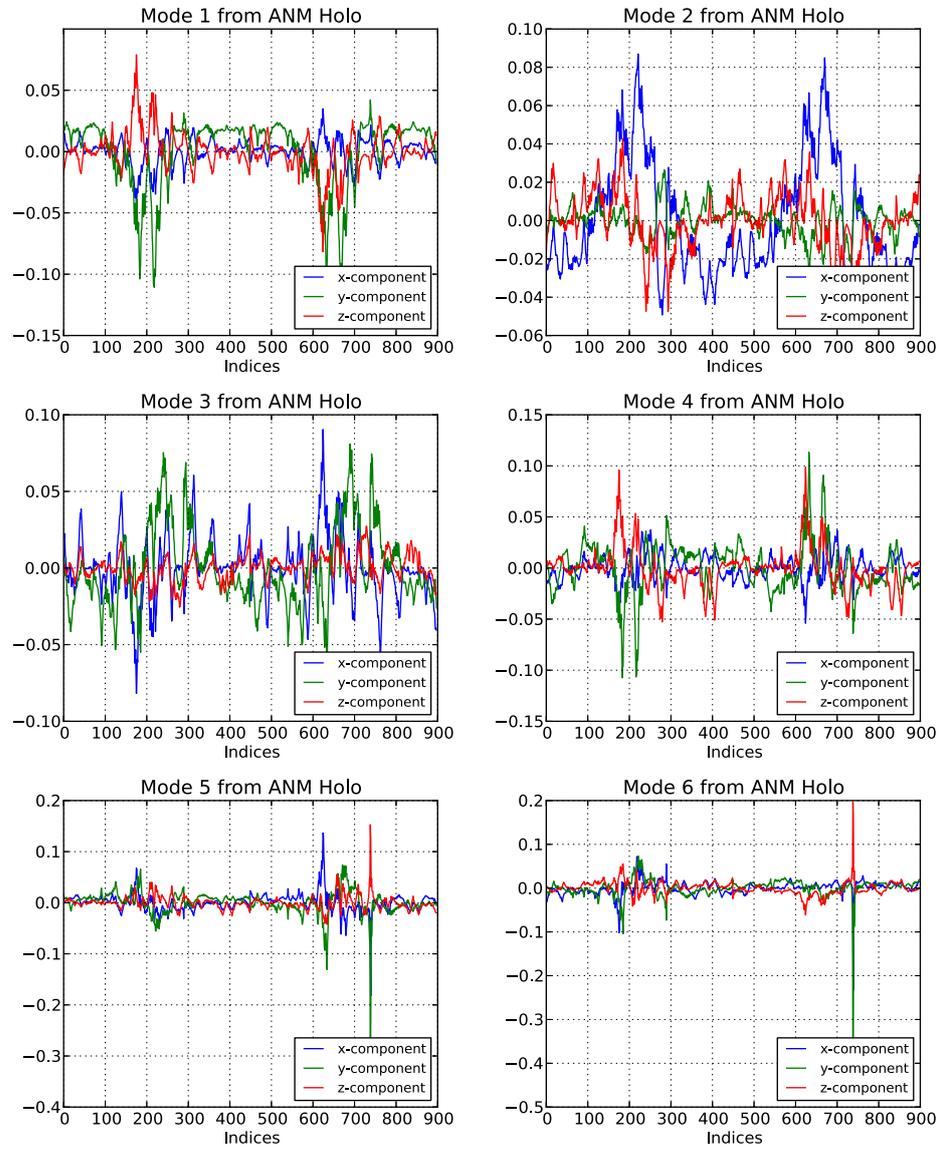


Figure 62: The mode array of the first 6 normal modes of the holo dimer based on the ANM.

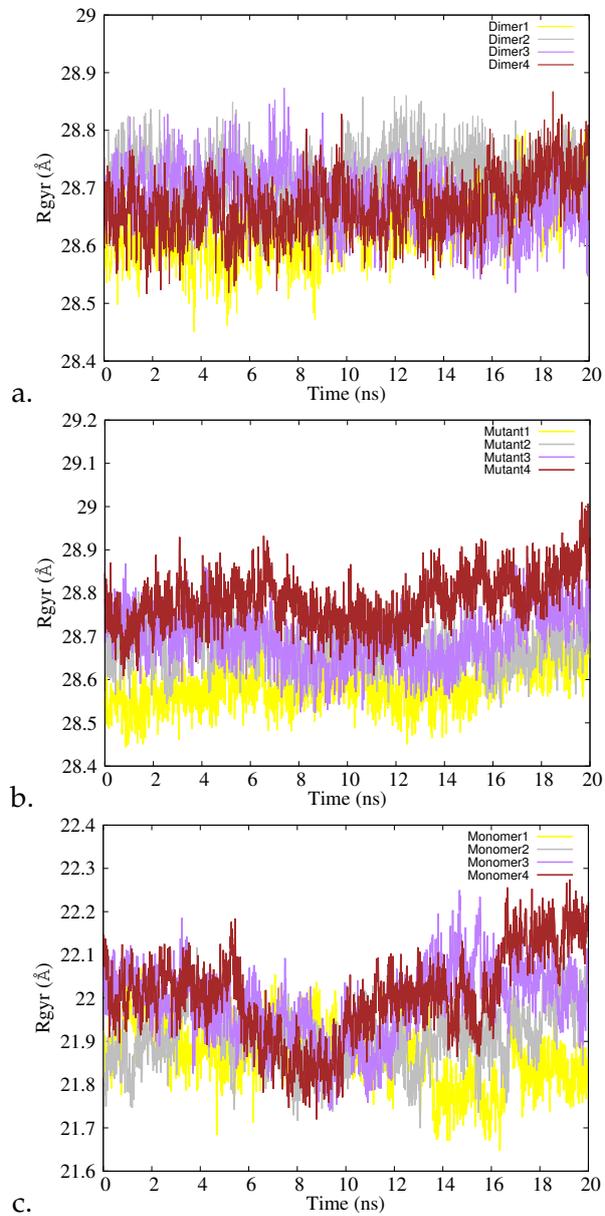


Figure 63: Rgyr as a function of time in the dimer (a.), the mutant (b.) and the monomer (c.) calculated from four individual, 20ns long chunks of the last 80ns MD simulations.

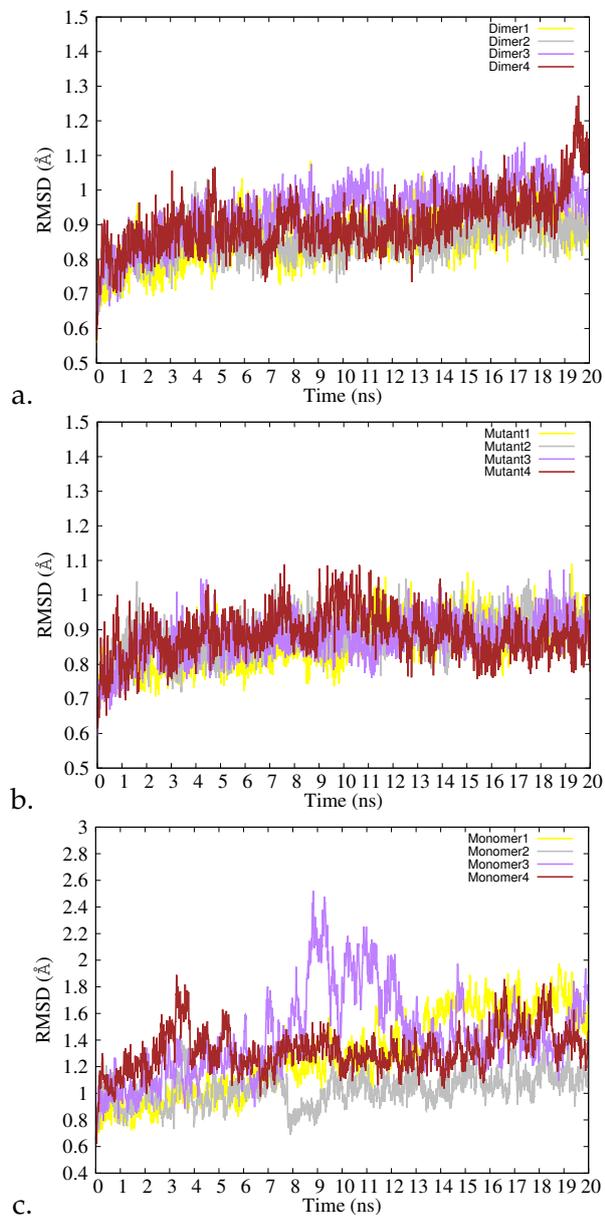


Figure 64: RMSD as a function of time in the dimer (a.), the mutant (b.) and the monomer (c.) calculated from the four individual, 20ns long chunks of the last 80ns MD simulations.

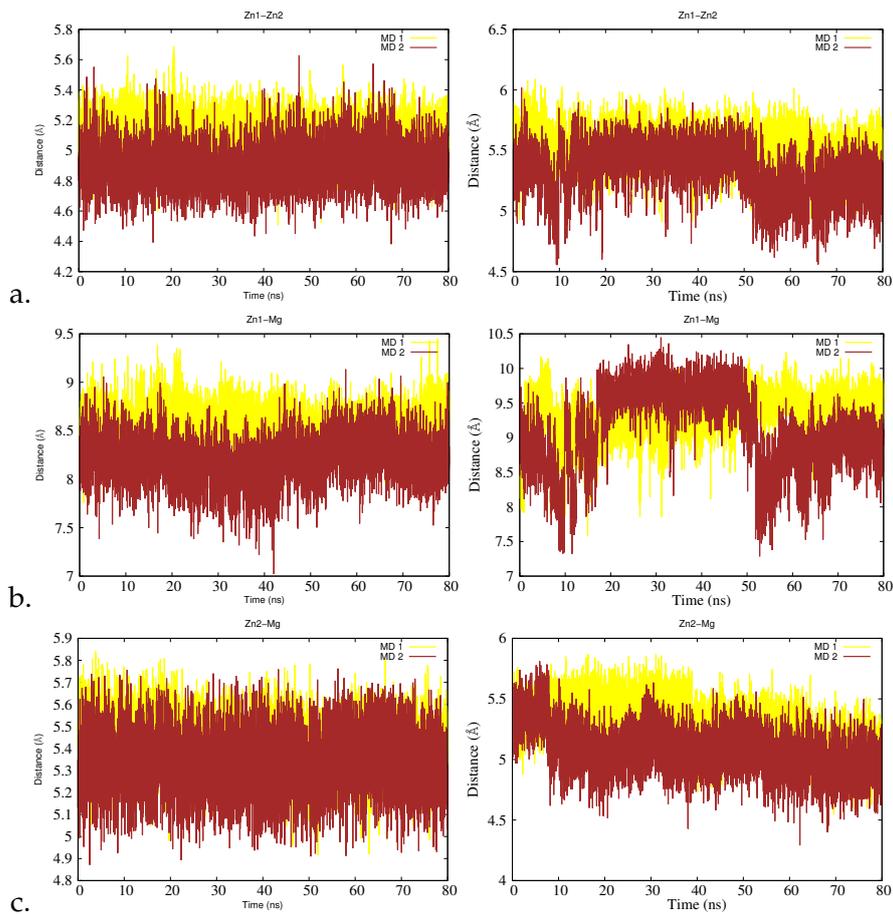


Figure 65: Time-evolution of the intra-metallic distances of the dimer (left) and the monomer (right) as a function of simulation time, calculated from the last 80 ns of two independent MD trajectories. Zn1-Zn2 (a.), Mg-Zn1 (b.) and Mg-Zn2 (c.).

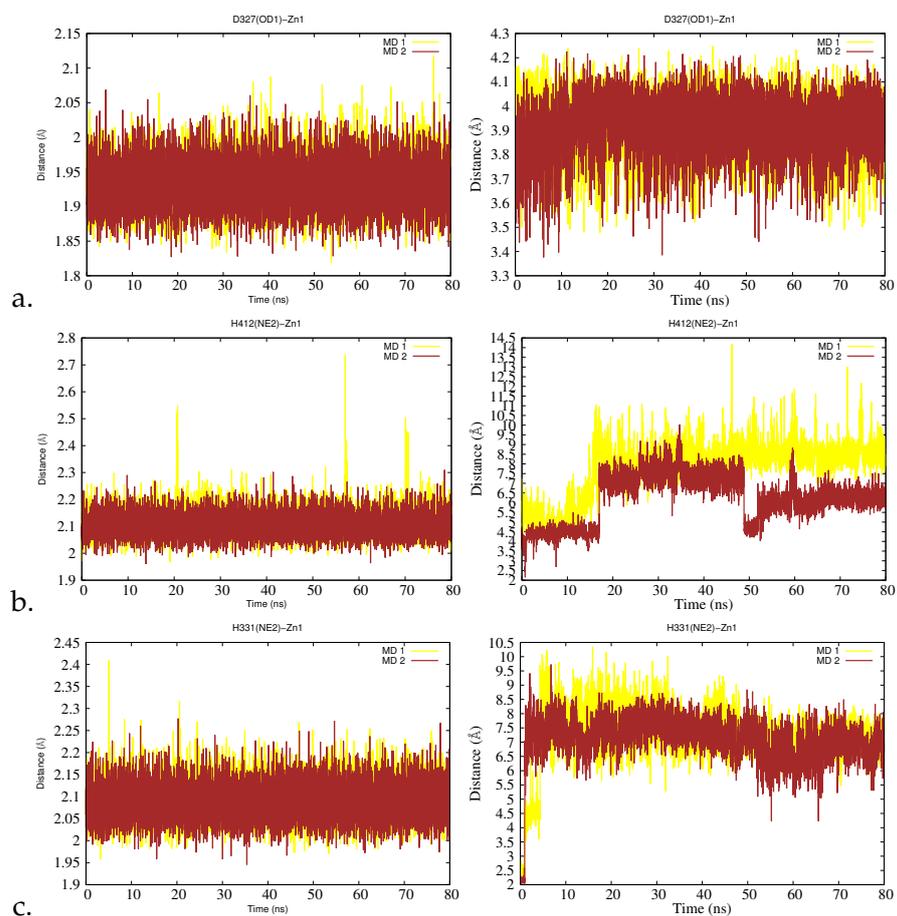


Figure 66: Time-evolution of the distances between Zn1 ion and the corresponding atoms of its coordination shell residues, as calculated from the last 80ns MD trajectories of the dimer (left) and the monomer (right). Zn1-D327(OD1) (a.), Zn1-412(NE2) (b.) and Zn1-331(NE2) (c.).

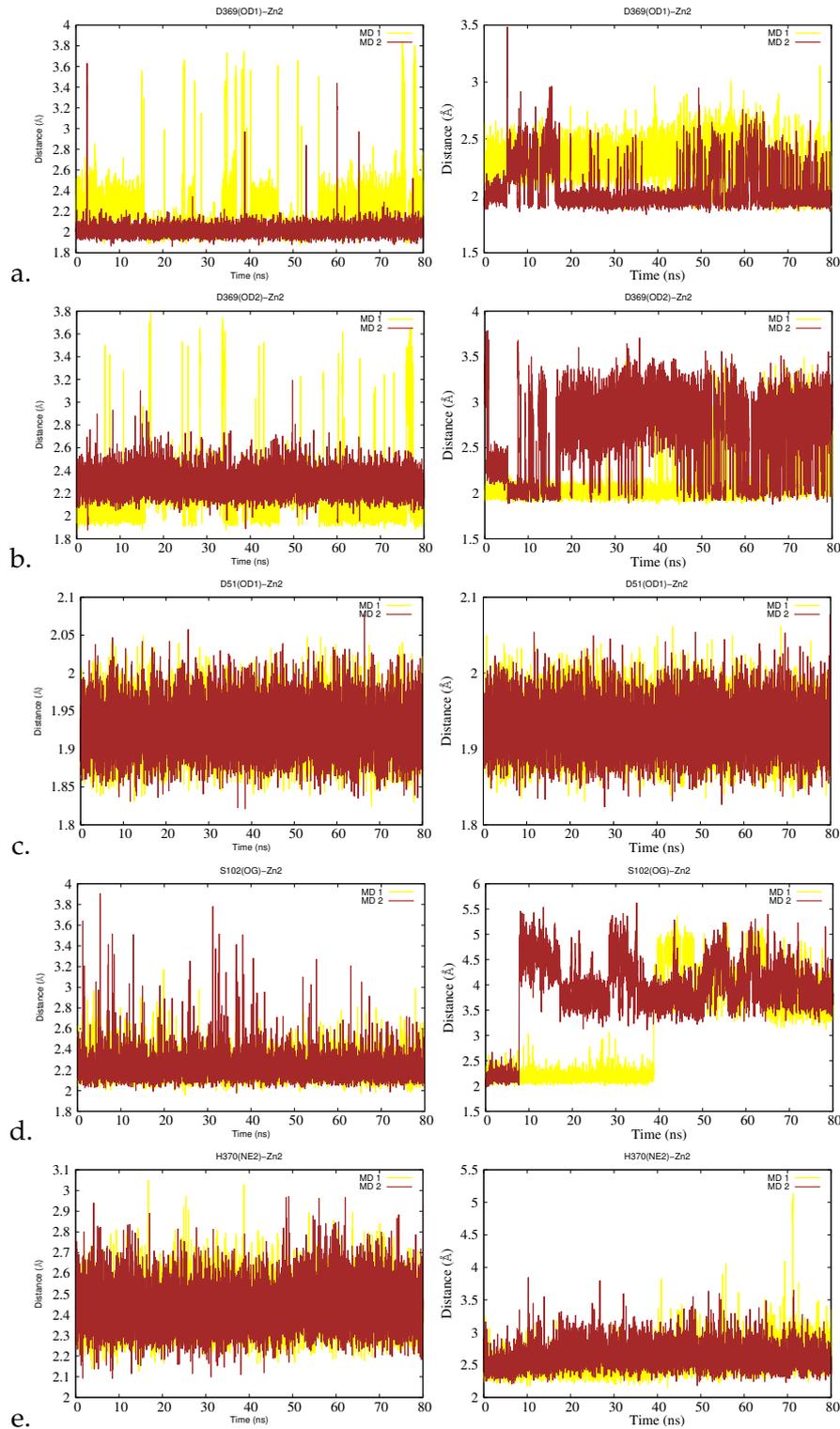


Figure 67: Time-evolution of the distances between Zn²⁺ ion and the corresponding atoms of its coordination shell residues, as calculated from the last 80 ns MD trajectories of the dimer (left) and the monomer (right). Zn²⁺-369(OD1) (a.), Zn²⁺-D369(OD2) (b.), Zn²⁺-51(OD1) (c.), Zn²⁺-S102(OG) (d.), Zn²⁺-370(NE2) (e.).

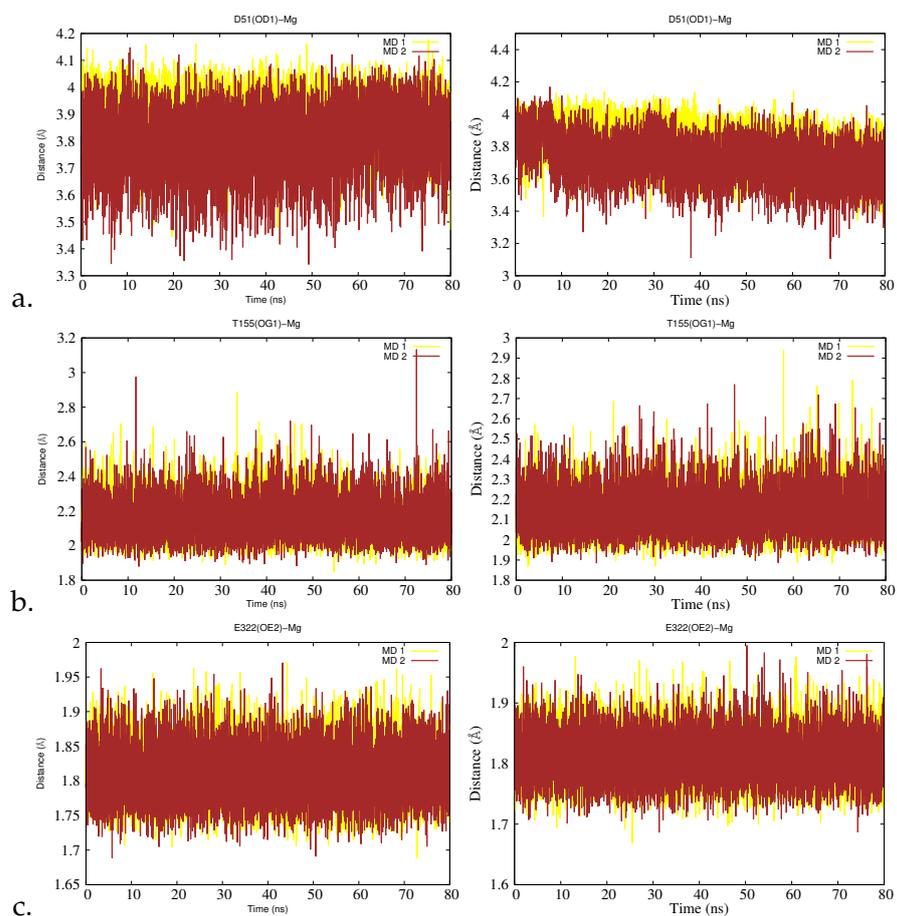


Figure 68: Time-evolution of the distances between Mg ion and the corresponding atoms of its coordination shell residues, as calculated from the last 80ns MD trajectories of the dimer (left) and the monomer (right). Mg-D51(OD1) (a.), Mg-T155(OG1) (b.) and Mg-E322(OE2) (c.).

BIBLIOGRAPHY

- [1] Y. Akiyama and K. Ito. Folding and Assembly of Bacterial Alkaline Phosphatase in Vitro and in Vivo. *J. Bio. Chem.*, 268:8146–8150, 1993. (Cited on page 13.)
- [2] D. Antoniou and S.D. Schwartz. Protein Dynamics and Enzymatic Chemical Barrier Passage. *The Journal of Physical Chemistry B*, 115(51):15147–15158, 2011. (Cited on page 12.)
- [3] B. Ásgeirsson and Ó.S. Andrésón. Primary Structure of Cold-adapted Alkaline Phosphatase From a *Vibrio* sp. as Deduced From the Nucleotide Gene Sequence. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1549(1):99–111, 2001. (Cited on page 105.)
- [4] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*, 80(1):505–515, 2001. (Cited on pages 20 and 22.)
- [5] I. Bahar, A.R. Atilgan, and B. Erman. Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-parameter Harmonic Potential. *Folding and Design*, 2(3):173–181, 1997. (Cited on page 20.)
- [6] I. Bahar, A. Wallqvist, D.G. Covell, and R.L. Jernigan. Correlation between Native-State Hydrogen Exchange and Cooperative Residue Fluctuations from a Simple Model. *Biochemistry*, 37(4):1067–1075, 1998. (Cited on pages 21 and 24.)
- [7] I. Bahar, T.R. Lezon, A. Bakan, and I.H. Shrivastava. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chemical Reviews*, 110(3):1463–1497, 2010. (Cited on pages 20, 24, and 105.)
- [8] I. Bahar, T.R. Lezon, L.W. Yang, and E. Eyal. Global Dynamics of Proteins: Bridging Between Structure and Function. *Annual Review of Biophysics*, 39(1):23–42, 2010. (Cited on pages 18, 19, 20, and 24.)
- [9] A. Bakan and I. Bahar. Computational Generation Inhibitor-bound Conformers of P38 MAP Kinase and Comparison with Experiments. *Pacific Symposium on Biocomputing*, pages 181–192, 2011. (Cited on page 24.)

- [10] A. Bakan, L.M. Meireles, and I. Bahar. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*, 27(11):1575–1577, June 2011. (Cited on pages 22, 31, 32, and 38.)
- [11] S.J. Benkovic. A Perspective on Enzyme Catalysis. *Science*, 301(5637):1196–1202, 2003. (Cited on pages 12, 96, and 97.)
- [12] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000. (Cited on page 31.)
- [13] T.L. Blundell and N. Srinivasan. Symmetry, Stability, and Dynamics of Multidomain and Multicomponent Protein Systems. *Proceedings of the National Academy of Sciences*, pages 14243–14248. (Cited on pages 3 and 9.)
- [14] M. Bortolato, F. Besson, and B. Roux. Role of Metal Ions on the Secondary and Quaternary Structure of Alkaline Phosphatase from Bovine Intestinal Mucosa. *Proteins: Structure, Function, and Bioinformatics*, 37(2), 1999. (Cited on page 14.)
- [15] R.R. Boulanger, Jr. Kantrowitz, and E.R. Kantrowitz. Characterization of a Monomeric Escherichia coli Alkaline Phosphatase Formed upon a Single Amino Acid Substitution. *The Journal of Biological Chemistry*, 278(26):23497–23501, 2003. (Cited on pages 10, 14, 70, 73, 77, 95, and 96.)
- [16] L. Brocchieri and S. Karlin. Protein Length in Eukaryotic and Prokaryotic Proteomes. *Nucleic Acids Research*, 33(10):3390–3400, 2005. (Cited on page 19.)
- [17] B. Brooks and M. Karplus. Normal Modes for Specific Motions of Macromolecules: Application to the Hinge-bending Mode of Lysozyme. *Proceedings of the National Academy of Sciences*, 82(15):4995–4999, 1985. (Cited on page 22.)
- [18] Bruccoleri R.E. Olafson B.D. States D.J. Swaminathan S. Brooks, B.R and M. Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1982. (Cited on pages 32 and 37.)
- [19] C.L. Brooks, M. Karplus, and B.M. Pettitt. *Advances in Chemical Physics, Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, volume 4. Wiley-Interscience, 1990. (Cited on page 22.)
- [20] D.A. Case. Normal Mode Analysis of Protein Dynamics. *Current Opinion in Structural Biology*, 4(2):285–290, 1994. (Cited on page 22.)

- [21] J.P. Changeux and S. Edelstein. Conformational Selection or Induced Fit? 50 Years of Debate Resolved. *F1000 Biology Reports*, 3(19):15, 2011. (Cited on pages 4, 10, and 11.)
- [22] D. Chappelet-Tordo, M. Iwatsubo, and M. Lazdunski. Negative Cooperativity and Half of the Sites Reactivity. Alkaline Phosphatases of *Escherichia coli* with Zn^{2+} , Co^{2+} , Cd^{2+} , Mn^{2+} , and Cu^{2+} in the Active Sites. *Biochemistry*, 13(18):3754–3762, 1974. (Cited on page 18.)
- [23] J.E. Coleman. Structure and Mechanism of Alkaline Phosphatase. *Annu. Rev. Biophys. Biomol. Struct.*, 21:441–483, 1992. (Cited on pages 3, 13, 14, 16, and 17.)
- [24] Q. Cui and M. Karplus. Allostery and Cooperativity Revisited. *Protein Science*, 17(8):1295–1307, 2008. (Cited on pages 4, 11, and 18.)
- [25] R.M. Daniel, R.V. Dunn, J.L. Finney, and J.C. Smith. The Role of Dynamics in Enzyme Activity. *Annual Review of Biophysics and Biomolecular Structure*, 32(1):69–92, 2003. (Cited on page 10.)
- [26] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n^2 method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993. doi: 10.1063/1.464397. URL <http://link.aip.org/link/?JCP/98/10089/1>. (Cited on page 37.)
- [27] D.J. Evans and B.L. Holian. The Nose–Hoover Thermostat. *The Journal of Chemical Physics*, 83(8):4069–4074, 1985. (Cited on pages 26 and 37.)
- [28] A. Fersht. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. W.H. Freeman, 1998. (Cited on pages 9, 10, and 11.)
- [29] M.J. Field. Molecular Simulations of Enzyme Catalysis. In *ESCES*, pages 65–78, 2007. (Cited on page 24.)
- [30] D.S. Goodsell and A.J. Olson. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.*, 29(1):105–153, 2000. (Cited on pages 3 and 9.)
- [31] A. Grossfield and D.M. Zuckerman. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. (Cited on pages 19, 24, and 27.)
- [32] T. Haliloglu, I. Bahar, and B. Erman. Gaussian Dynamics of Folded Proteins. *Physical Review Letters*, 79(16):3090–3093, 1997. (Cited on page 21.)

- [33] T. Hansson, C. Oostenbrink, and W. Gunsteren. Molecular Dynamics Simulations. *Current Opinion in Structural Biology*, 12(2): 190–196, 2002. (Cited on pages 24, 25, and 27.)
- [34] K. Henzler-Wildman and D. Kern. Dynamic Personalities of Proteins. *Nature*, 450(7172):964–972, 2007. (Cited on page 9.)
- [35] K.M. Holtz and E.R. Kantrowitz. The Mechanism of Alkaline Phosphatase Reaction: Insights from NMR, Crystallography and Site-specific Mutagenesis. *FEBS Letters*, 462:7–11, 1999. (Cited on pages 13, 14, and 77.)
- [36] C.Y. Huang, S.G. Rhee, and P.B. Chock. Subunit Cooperation and Enzymatic Catalysis. *Annual review of biochemistry*, 51:935–971, 1982. (Cited on pages 3 and 17.)
- [37] X. Huangming. The Effect of Dimerization on the Functionality of Alkaline Phosphatase. Technical report, Department of Biological Sciences, NUS, Singapore, 2002. (Cited on pages 3 and 13.)
- [38] S. Hughes, J. Tanner, A. Hindley, A. Miller, and I.R. Gould. Functional Asymmetry in the Lysyl - tRNA Synthetase Explored by Molecular Dynamics, Free Energy Calculations and Experiment. *BMC Structural Biology*, 3(1):5, 2003. (Cited on page 11.)
- [39] S. Hughes, J. Tanner, A. Julian, A.D. Miller, and I.R. Gould. Molecular Dynamics Simulations of LysRS: An Asymmetric State. *Proteins: Structure, Function, and Bioinformatics*, 62(3), 2006. (Cited on page 11.)
- [40] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996. (Cited on pages 32, 37, and 38.)
- [41] J.H. Jensen. *Molecular Modeling Basics*. CRC Press, 2010. (Cited on pages 26 and 27.)
- [42] E.E. Kim and H.W. Wyckoff. Reaction Mechanism of Alkaline Phosphatase Based on Crystal Structures. Two-metal Ion Catalysis. *J. Mol. Biol.*, 218:449–464, 1991. (Cited on pages 13, 14, 15, 16, 98, and 101.)
- [43] D. E. Koshland, G. Nemethy, and D. Filmer. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry*, 5(1):365–385, 1966. (Cited on pages 10 and 18.)
- [44] R.A. Laskowski. PDBsum New Things. *Nucleic Acids Research*, 37(suppl 1):D355–D359, 2009. (Cited on pages xxiii, 31, and 74.)

- [45] M.H. Le Du and J.L. Millán. Structural Evidence of Functional Divergence in Human Alkaline Phosphatases. *Journal of Biological Chemistry*, 277(51):49808–49814, 2002. (Cited on page 13.)
- [46] A.R. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2001. (Cited on pages 19, 25, 26, and 27.)
- [47] Y. Levy and A. Caflisch. Flexibility of Monomeric and Dimeric HIV-1 Protease. *J. Phys. Chem. B*, 107:3068–3079, 2003. (Cited on page 9.)
- [48] Y. Levy, A. Caflisch, J. N. Onuchic, and P.G. Wolynes. The Folding and Dimerization of HIV-1 Protease: Evidence for a Stable Monomer from Simulations. *Journal of Molecular Biology*, 340(1): 67–79, 2004. (Cited on page 9.)
- [49] H. Li, A.D. Robertson, and J.H. Jensen. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins: Structure, Function, and Bioinformatics*, 61(4):704–721, 2005. (Cited on pages 31 and 32.)
- [50] L. Li, V.N. Uversky, A.K. Dunker, and S.O. Meroueh. (Cited on page 11.)
- [51] X. Lopez, D.M. York, A. DeJaegere, and M. Karplus. Theoretical Studies on the Hydrolysis of Phosphate Diesters in the Gas Phase, Solution and RNase A. *Int. J. Quantum Chem.*, 86:10–16, 2002. (Cited on page 13.)
- [52] D.C. Martin, S.C. Pastra-Landis, and E.R. Kantrowitz. Amino Acid Substitutions at the Subunit Interface of Dimeric *Escherichia coli* Alkaline Phosphatase Cause Reduced Structural Stability. *Protein Sci.*, 8(5):1152–1159, 1999. (Cited on pages 5, 13, 14, and 77.)
- [53] J.A. McCammon, B.R. Gelin, M. Karplus, et al. Dynamics of Folded Proteins. *Nature*, 267(5612):585, 1977. (Cited on page 19.)
- [54] S. Mehboob, L. Guo, W. Fu, A. Mittal, T. Yau, K. Truong, M. Johlfs, F. Long, L.W.M. Fung, and M.E. Johnson. Glutamate Racemase Dimerization Inhibits Dynamic Conformational Flexibility and Reduces Catalytic Rates. *Biochemistry*, 48(29):7045–7055, 2009. (Cited on page 10.)
- [55] L. Meireles, M. Gur, A. Bakan, and I. Bahar. Pre-existing Soft Modes of Motion Uniquely Defined by Native Contact Topology Facilitate Ligand Binding to Proteins. *Protein Science*, 20(10):1645–1658, September 2011. (Cited on page 20.)
- [56] J.L. Millán. Alkaline Phosphatases. *Purinergic Signalling*, 2(2): 335–341, June 2006. (Cited on page 3.)

- [57] J.E. Murphy, T.T. Tibbitts, E.R. Kantrowitz, et al. Mutations at positions 153 and 328 in *Escherichia coli* alkaline phosphatase provide insight towards the structure and function of mammalian and yeast alkaline phosphatases. *Journal of molecular biology*, 253(4):604–617, 1995. (Cited on page 105.)
- [58] S. Orhanović and M. Pavela-Vrančić. Dimer Asymmetry and the Catalytic Cycle of Alkaline Phosphatase from *Escherichia coli*. *European Journal of Biochemistry*, 270(21):4356–4364, 2003. (Cited on pages 3, 17, and 18.)
- [59] Y.P. Pang. Successful Molecular Dynamics Simulation of Two Zinc Complexes Bridged by a Hydroxide in Phosphotriesterase Using the Cationic Dummy Atom Method. *Proteins: Structure, Function, and Bioinformatics*, 45(3):183–189, 2001. (Cited on page 34.)
- [60] H.G. Petersen. Accuracy and Efficiency of the Particle Mesh Ewald Method. *Journal of Chemical Physics*, 103(9):3668–3679, 1995. (Cited on page 26.)
- [61] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005. (Cited on pages 32 and 37.)
- [62] A.A. Porollo, R. Adamczak, and J. Meller. POLYVIEW: a Flexible Visualization Tool for Structural and Functional Annotations of Proteins. *Bioinformatics*, 20(15):2460–2462, 2004. (Cited on page 31.)
- [63] A. Radzicka and R. Wolfenden. A Proficient Enzyme. *Science*, 267(5194):90–93, 1995. (Cited on page 11.)
- [64] A. Rahman. Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.*, 136:A405–A411, 1964. (Cited on page 19.)
- [65] D.C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2004. (Cited on pages 25, 26, and 27.)
- [66] T.D. Romo and A. Grossfield. Validating and Improving Elastic Network Models with Molecular Dynamics Simulations. *Proteins: Structure, Function, and Bioinformatics*, 79(1):23–34, September 2010. (Cited on page 19.)
- [67] T. Schlick, R. Collepardo-Guevara, L.A. Halvorsen, S. Jung, and X. Xiao. Biomolecular Modeling and Simulation: a Field Coming of Age. *Quarterly Reviews of Biophysics*, 44(02):191–228, 2011. (Cited on pages 18 and 24.)

- [68] D.E. Shaw, M.M. Deneroff, R.O. Dror, J.S. Kuskin, R.H. Larson, J.K. Salmon, C. Young, B. Batson, K.J. Bowers, J.C. Chao, M.P. Eastwood, J. Gagliardo, J.P. Grossman, C.R. Ho, D.J. Ierardi, I. Kolossváry, J.L. Klepeis, T. Layman, C. McLeavey, M.A. Moraes, R. Mueller, E.C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S.C. Wang. Anton, a Special-purpose Machine for Molecular Dynamics Simulation. *SIGARCH Comput. Archit. News*, 35(2):1–12, 2007. ISSN 0163-5964. (Cited on page 24.)
- [69] W. Shunzhou and P.V. Coveney. Molecular Dynamics Simulation Reveals Structural and Thermodynamic Features of Kinase Activation by Cancer Mutations within the Epidermal Growth Factor Receptor. *Journal of Computational Chemistry*, 32(13):2843–2852, 2011. (Cited on page 61.)
- [70] E.C. Slater, K. van Dam, and B.F. van Gelder. *Structure and Function of Energy-transducing Membranes: Proceedings of a Workshop Held in Amsterdam on August 10-13, 1977, in Honour of E. C. Slater's 60th Birthday*. B.B.A. library. Elsevier Scientific Pub. Co., 1977. (Cited on page 17.)
- [71] M. Sone, S. Kishigami, T. Yoshihisa, and K. Ito. Roles of Disulfide Bonds in Bacterial Alkaline Phosphatase. *Journal of Biological Chemistry*, 272(10):6174–6178, 1997. (Cited on page 13.)
- [72] J.M. Sowadski, M.D. Handschumacher, H.M. Murthy, B.A. Foster, and H.W. Wyckoff. Refined Structure of Alkaline Phosphatase from *Escherichia coli* at 2.8 Å Resolution. *Journal of Molecular Biology*, 186(2):417–433, November 1985. (Cited on pages 14 and 35.)
- [73] B. Stec, K.M. Holtz, and E.R. Kantrowitz. A Revised Mechanism for the Alkaline Phosphatase Reaction Involving Three Metal Ions. *J. Mol. Biol.*, 299:1303–1311, 2000. (Cited on pages 13, 15, and 16.)
- [74] F.H. Stillinger and A. Rahman. Improved Simulation of Liquid Water by Molecular Dynamics. *The Journal of Chemical Physics*, 60(4):1545–1557, 1974. (Cited on page 19.)
- [75] F. Tama and C.L. Brooks. Symmetry, Form, and Shape: Guiding Principles for Robustness in Macromolecular Machines. *Annual Review of Biophysics and Biomolecular Structure*, 35(1):115–133, 2006. (Cited on pages 9, 20, and 22.)
- [76] A. Torriani-Gorini, E. Yagil, and S. Silver. *Phosphate in Microorganisms: Cellular and Molecular Biology*. ASM, 1994. (Cited on pages 3 and 13.)

- [77] A. Warshel, P.K. Sharma, M. Kato, Y. Xiang, H. Liu, and M.H.M. Olsson. Electrostatic Basis for Enzyme Catalysis. *Chemical Reviews*, 106(8):3210–3235, 2006. (Cited on page 11.)
- [78] F. Weinhold. *Classical and Geometrical Theory of Chemical and Phase Thermodynamics*. Wiley, 2009. (Cited on page 19.)
- [79] F.H. Westheimer. Why Nature Chose Phosphates. *Science*, 235:1173–1178, 1987. (Cited on page 12.)
- [80] Thomas Williams, Colin Kelley, and many others. Gnuplot 4.4: an Interactive Plotting Program. 2010. (Cited on page 38.)
- [81] K. Wilson and J.M. Walker. *Principles and Techniques of Practical Biochemistry*. Cambridge University Press, 5 edition, 2000. (Cited on pages 3, 12, and 13.)
- [82] L. Yang, G. Song, A. Carriquiry, and R.L. Jernigan. Close Correspondence between the Motions from Principal Component Analysis of Multiple HIV-1 Protease Structures and Elastic Network Modes. *Structure*, 16(2):321–330, 2008. (Cited on page 24.)

ERKLÄRUNG GEMÄSS §8 (3) B) UND C) DER
PROMOTIONSORDNUNG:

- A) Ich erkläre hiermit, dass ich die vorgelegte Dissertation selbst verfasst und mich keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe,
- B) Ich erkläre hiermit, dass ich an keiner anderen Stelle ein Prüfungsverfahren beantragt bzw. die Dissertation in dieser oder anderer Form bereits anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.

19th of December

Heidelberg

Mari Chikvaidze