Computational investigation on the mechanism of oxygen tolerance in the radical SAM enzyme lysine 2,3-aminomutase

Submitted November 2019, in partial fulfilment for the degree of Doctor of Philosophy.

Damiano Spadoni

4243336

Supervised by Dr. Anna Croft

School of Chemical and Environmental Engineering,

University of Nottingham

I hereby declare that this dissertation is all my own work, except as indicated in the text:

Signature _____

Date ____ / ____ / _____

This thesis is dedicated to the memory of my father
Gianfranco, who always encouraged me during my
studies and whose words of support are missed since
he left this world during my PhD, and Massey who was
more a brother than a dog to me.
I miss you everyday.

Acknowledgements

Abstract

The radical *S*-adenosylmethionine (AdoMet or SAM) enzyme superfamily represents an ensemble of proteins that are able to catalyse biochemical reactions involving organic radical intermediates. These intermediate radicals then undergo a wide range of reactions, many of them difficult to accomplish in the laboratory. The products of such reactions are bioactive compounds of pharmaceutical interest that can be more over used as building blocks for other compounds. Virtually every enzyme belonging to this family is known to be unstable in air due to the requirement for a catalytic, oxygen sensitive [Fe_4-S_4] cluster that decomposes after oxidative attack by reactive oxygen species (ROS), inactivating the enzyme. The study of these enzymes and their possible application in biotechnology is difficult due to the oxygen sensitivity of the [Fe_4-S_4] cluster forcing their usage under strictly inert atmosphere. *C. subterminale* lysine 2,3-aminomutase (*Cs*LAM) is a widely studied radical AdoMet enzyme and a natural oxygen-tolerant variant from *B. subtilis* (*Bs*LAM) was discovered that catalyses, in presence of air, the interconversion between α- and β-lysine.

This project utilised computational methodologies for the assessment of how radical SAM enzymes may manage to shield their FeS cluster from air degradation. Comparison was made of the *Cs*LAM crystal structure with the structural model of oxygen-tolerant *Bs*LAM, here obtained through homology modelling using the 3D structure as its template. Following the validation of the model through PCA, both the *Cs*LAM and the *Bs*LAM enzyme structures were compared and molecular dynamics simulations were used to identify different ways the two enzymes might deal with oxygen. The tunnel searching software CAVER was used to identify those amino acid residues that could obstruct oxygen flow due to their size or their ability to trap oxygen in *Bs*LAM. Quantum mechanics calculations were then performed on the [Fe_4-S_4] sub-system to retrieve information about the difference in the electrostatics governed by the protein environment. The difference in electrostatics could account for different redox potentials in the two enzymes by making the *Bs*LAM less prone to oxidation by ROS. A selection of amino acid residues were identified as likely to affect the redox potential and

mutants of the *Cs*LAM bearing such residues were created. The observation of their dipole moment suggested that the double mutation H131Y/A138S could positively affect the oxygen tolerance in the enzyme.

Table of contents

11

Abbreviations used

S-adenosylmethionine	AdoMet / SAM
S-adenosylmethionine radical intermediate	dAdo•
Glycyl radical activating enzymes	GRE-AEs
Pyruvate formate lyase	PFL-AE
Lysine 2,3-aminomutase	LAM
Clostridium subterminale LAM	*Cs*LAM
Bacillus subtilis LAM	*Bs*LAM
Triose-phosphate isomerase	TIM
Reactive oxygen species	ROS
Ferredoxin-NADP$^+$ reductase	FDR
Pyridoxal-*5′*-phosphate	PLP
Lysine	Lys
Coenzyme pyrroloquinoline quinone synthesis protein E	PqqE
Oxygen-independent coproporphyrinogen-III oxidase	HemN
Biotin Synthase	BioB
Molybdenum cofactor biosynthesis protein A	MoaA
tRNA 4-demethylwyosine synthase	TYW1
Thiamine biosynthesis protein C	ThiC
Caulobacter crescentus ThiC	*Cc*ThiC
Arabidopsis thaliana ThiC	*At*ThiC
Glutamate 2,3-aminomutase	GAM
Ornithine 2,3-aminomutase	OAM
Fumarate and nitrate reductase	FNR
Ferredoxin	Fd

High Potential Iron-sulfur protein   HiPIP

Critical assessment of protein structure prediction   CASP

Quantum Mechanics   QM

Molecular dynamics   MD

Vertical electron affinity   VEA

1 Introduction

1.1   Generating a greener future

Biological catalysts have emerged as potentially greener and more sustainable alternatives to traditional manufacture processes in the pharmaceutical and fine chemicals industries.[1,2] They confer advantages ranging from the use of mild conditions of temperature, pH and pressure, to exquisite stereo- and regio-selectivity of reaction. Both catalyst discovery, and ensuring that biological catalysts have appropriate properties for large-scale industrial processes, are still bottlenecks in this process, and underscore the slow uptake of these technologies.[3]

Major challenges in synthesis that would constitute important targets for biological catalysts, still include carbon-hydrogen activation and carbon-carbon bond formation.[4] These processes are easily achieved through radical chemistry, where a highly reactive unpaired electron is able to remove unactivated hydrogen atoms, or undergo coupling to form new carbon bonds. This radical intermediates can thus mediate chemistries that would be impossible to access by polar reactions. However, this reactivity can be marred by a lack of selectivity.

Radical enzymes offer a highly controlled environment through which these extremely reactive and useful reactions can take place.[5] They have recently been demonstrated as useful industrial candidates for a wide range of processes including complex biosyntheses and environmental remediation.[3] Amongst the radical enzymes, there is one highly adaptable class that has demonstrated a broad substrate and product range, with good potential for further enzyme engineering that would be needed to realise their industrial potential. This class is the radical SAM enzyme superfamily.

1.2   The radical SAM enzyme superfamily

In 2001 bioinformatics methods revealed a wide variety of metalloenzymes that utilised a reduced iron-sulfur cluster in combination with *S*-adenosylmethionine to initiate this kind of transformations and classify them in a superfamily.[4] These became known as Radical *S*-adenosylmethionine (AdoMet or radical SAM) enzymes and are a superfamily of proteins that make use of organic radical intermediates generated from a [Fe_4-S_4] cluster and *S*-adenosylmethionine cofactors. Such enzymes were first reported over 40 years ago and their reactions cover a wide variety of uncommon chemical transformations.[6]

Currently over 250,000 enzymes are estimated belonging to this enzyme superfamily.[7] These enzymes are found in all three domains of life and are likely of ancient origin, among the earliest biological radical catalysts that evolved in an anaerobic world.[8] Anaerobic microorganisms (bacteria and archaea) have provided the largest proportion of radical SAM enzymes, probably because of the extreme sensitivity and reactivity to oxygenated species of both radical intermediates and the iron-sulfur cluster.[9] The radical SAM members catalyse a diverse set of chemical reactions and are found in several metabolic pathways such as enzyme activation, DNA repair, primary metabolism, biosynthesis of complex cofactors, biodegradation pathways, synthesis of antibiotics and modification of tRNA (Figure 1).[10, 11, 12]

The SAM radical enzymes exhibit a very limited sequence homology among each other.[13] The most conserved sequence is represented by the CxxxCxxC motif although few members can show variations on this feature too. The cysteine residues coordinate three of the four iron atoms of a [Fe_4-S_4] cluster, while SAM binds to a unique iron site not coordinating to the backbone, *via* a chelate bond using its α-amino and α-carboxylate moiety.[14]

In contrast, the radical SAM enzyme crystal structures described so far seem to show structural analogies rather than sequence homology. They all adopt a full or partial $(\beta\alpha)_8$ triose-phosphate isomerase (TIM) barrel fold[15] with inner β-strands surrounded by peripheral α-helices laying on the protein surface thus forming the barrel-like structure. These common features are likely play a role in sealing off the active site during catalysis

inside the SAM radical core to prevent side reactions (see section 1.5).[16] Among other common characteristics there is a cluster binding site at one end of the barrel to keep it far from possible solvent interactions and the presence of a flexible loop moving on top of the TIM barrel that moves into place upon binding of the substrate (Figure 1).[17] The conserved CxxxCxxC motif is found on the loop that follows the first β-strand and coordinate three of the four iron atoms of the [Fe$_4$-S$_4$] cluster with the "unique" iron oriented towards the centre of the barrel.[17] The coordination of SAM to the unique iron partially seals off the cluster from
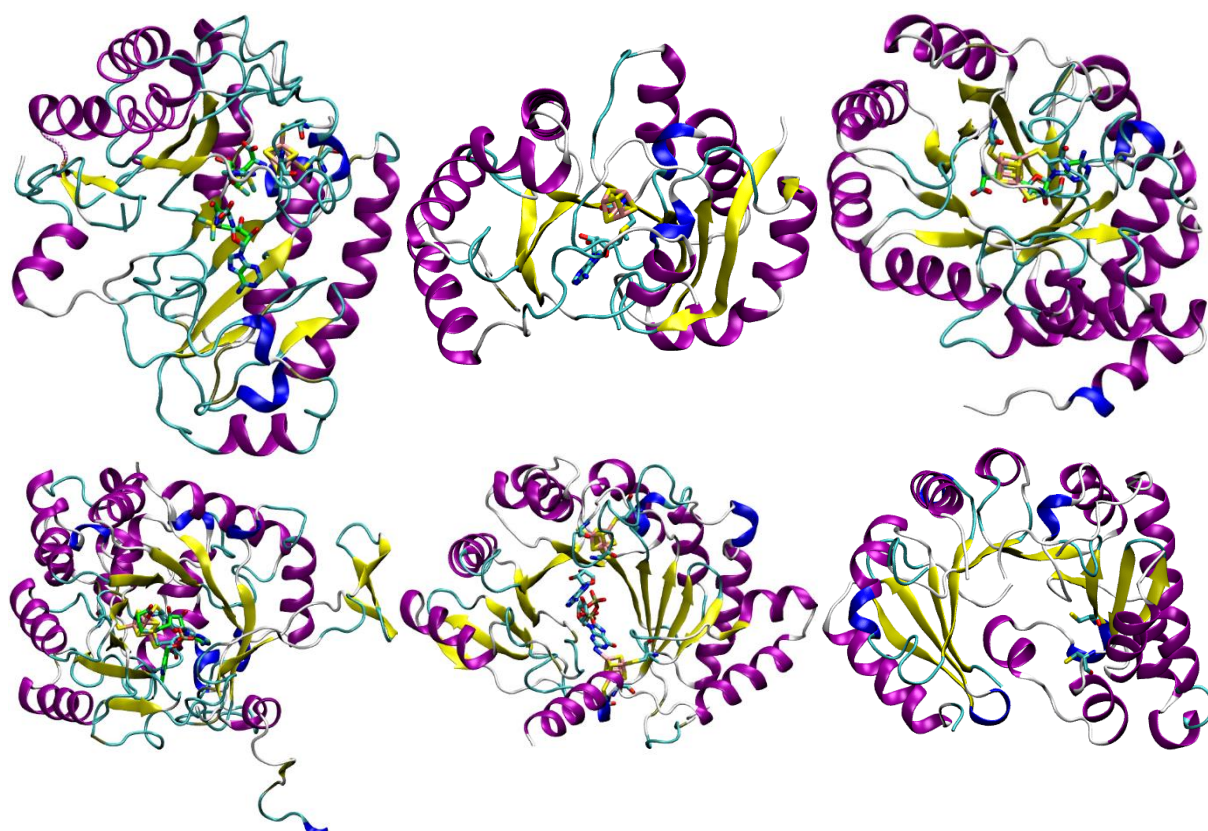
Figure 1. Representative views of the X-ray crystal structure of AdoMet radical enzymes from top of their TIM barrel. From left to right order, starting from top row: HemN (PDB code: 1OLT), PFL-AE (PDB code: 3CB8), BioB (PDB code: 1R30), LAM (PDB code: 2A5H), MoaA (PDB code: 2FB3) and TYW1 (PDB code: 2Z2U) with both SAM, [Fe$_4$-S$_4$] cluster and relative substrate bound (except TYW1). The TIM-barrel helices are coloured in purple whilst the strands are in yellow. Substrate and cofactors are represented in licorice with atom coloured as follow: Fe atoms in pink, S atoms in yellow, SAM-C atoms in green and Substrate-C atoms in cyan. The lateral opening of the barrel increases to enable the entrance of the substrate in the active site: from a small lysine in LAM to the bulkier peptide in PFL-AE. Image adapted from *Vey et al.*[17]

solvent during catalysis protecting the labile oxygen-sensitive cluster from degradation and preventing potential side reactions of the radical reactive species.[18] SAM is also kept in place by several conserved residues that are involved in electrostatic, hydrophobic, hydrogen-bonding, and π-stacking interactions. Among the structural motifs that help orient the SAM molecule to efficiently bind the [Fe$_4$-S$_4$] cluster, the "GGE motif" is known as glycine-rich region, and is located at the C-terminus of strand β2. This motif makes hydrogen-bonding interactions with the amino group of the SAM cofactor and facilitates correct orientation between SAM and the unique cluster Fe atom. Charged and polar protein sidechains arise from strands β4 and β5 that interact with the ribose hydroxyls of SAM molecule.[19] The GxIxGxxE is a SAM adenine-binding motif located at the β5 sheet, that controls the arrangement of SAM in the active site for optimal control of catalytic function.[17] The crystal structures also show, the [Fe$_4$-S$_4$] cluster and substrates bound together. It shows how important the tertiary fold is for these enzymes to help minimise the deleterious effect of incidental exposure of other cellular components to the highly reactive radical intermediates. The substrate binding helps to further shield the site of radical chemistry, and in many cases it appears that movement of protein loops during catalysis provides an additional means to protect radical intermediates from unwanted side reactions.[20]

Studies on mutants of these motifs confirm that the involved residues play critical roles in binding and cleavage of SAM and in positioning the formed 5′-deoxyadenosyl radical for reaction with the substrate.[21] Sharma *et al.*[19] showed the importance of polar residues in the catalytic effect in related B$_{12}$ radical enzymes, which can catalyse the homolytic cleavage of the Co-C bond of the B$_{12}$ cofactor leading to the formation of 5′-deoxyadenosyl radical and cob(II)alamin. The electrostatic interactions between the ribose moiety and the protein residues is crucial to achieve the catalytic effect as, in a radical process the charge distribution of the reacting fragments does not change during the reaction. Normally, in a heterolytic enzyme-catalysed reaction, the transition state presents a difference in the charge distribution of the broken bond of the substrate and the enzyme has evolved in order to

maximise favourable non-covalent interactions in the transition state. That is different from a radical process where the charge distribution does not change on transitioning to the intermediate, as the two unpaired electrons of the broken bond reside respectively on both the reactive fragments created from the bond cleavage. This electronic contribution makes it more difficult for the enzyme to apply electrostatic effects to stabilise the transition state of the radical process. What has been thought about the B_{12} radical (but also applicable to SAM) enzymes, is that nature attached a very polar adenosyl group to the leaving carbon in order for the enzyme to stabilise the transition state through electrostatic interactions with charged residues in the active site. Mutants replacing those residues that interact with the -OH groups of ribose (usually charged amino acids)[21] showed a reduction or loss in activity. These observation provided evidence that the disrupted catalysis is due to improper radical localisation and transition state stabilisation.

Before the discovery of the *S*-adenosylmethionine, adenosylcobalamin (AdoCbl, Figure 2) was the cofactor known as a source of free radical intermediates in enzymatic catalysis, and in particular *5'*-deoxyadenosyl radical formation.

For more than 30 years[22] it has been recognised as the coenzyme involved in intramolecular rearrangements undergoing a 1,2-migration between a functional group and a hydrogen atom without requiring activated bonds.[23] Those reactions begin with the homolytic cleavage of the Co-C*5'* bond of adenosylcobalamin to form Co(II) and *5'*-deoxyadenosyl radical after substrate binding to the active site (Figure 3). The *5'*-deoxyadenosyl radical extracts the migrating hydrogen from the substrate to form a radical on the substrate and *5'*-deoxyadenosine. After a rearrangement to the product-related radical, the ex-substrate abstracts a hydrogen from the methyl group of *5'*-deoxyadenosine returning to the initial radical intermediate that, after product dissociation, binds another substrate molecule to begin the cycle again.[24] *S*-adenosylmethionine operates in the same way, generating the *5'*-deoxyadenosyl moiety from SAM and mediating hydrogen transfer, but its structural simplicity with respect to coenzyme B_{12} means it was originally described as a 'a poor man's adenosylcobalamin.[25]

In the evolutionary history of life early cofactors should have been simple molecules. Similarly, early enzymes should have been simple macromolecules. Therefore, the survival of a structurally simpler alternative to adenosylcobalamin as source of the *5'*-deoxyadenosyl radical intermediate may indicate SAM as an evolutionary predecessor. It would have then have contributed to the evolution of the more complex radical initiator adenosylcobalamin.[20] SAM has long been characterised as an electrophilic methyl donor in several cellular reactions but a different function was discovered after the observation of reversible exchange of a hydrogen atom to the *5'*-methylene group of SAM from lysine and *β*-lysine in the radical SAM enzyme lysine 2,3-aminomutase by Frey *et al.*[26] Along these lines, all SAM radical enzymes are believed to commence their catalysis with the common route of the generation of the *5'*-deoxyadenosyl radical (dAdo•) intermediate with subsequent removal of a hydrogen atom from the substrate (Figure 3).[27]
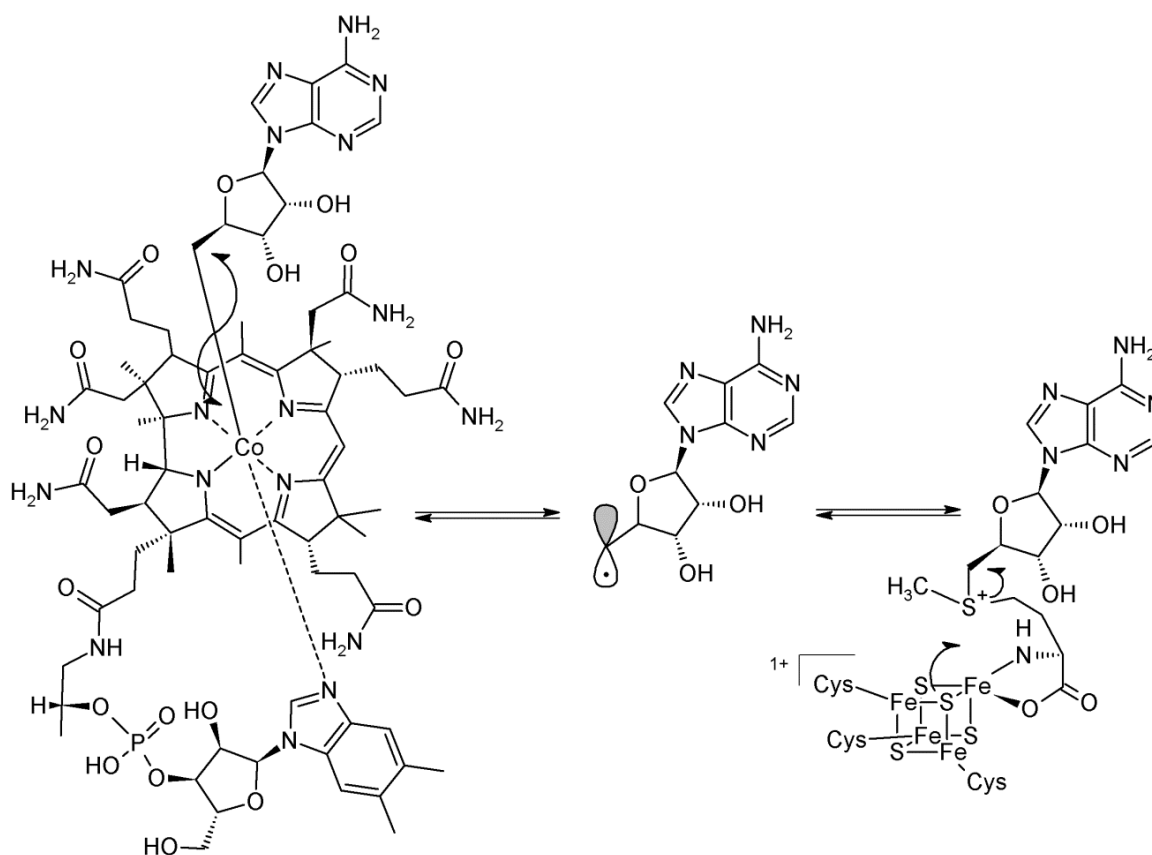
Figure 2. Homolytic cleavage and deoxyadenosyl radical intermediate (dAdo• in the middle) generation mechanism either in adenosylcobalamin, AdoCbl on the left and SAM (right) enzymes.

In the initial state, the sulfonium moiety of S-adenosylmethionine is in close proximity to the [Fe$_4$-S$_4$] cluster with direct orbital overlap between the two.[28] Subsequently, an electron is transferred from an electron donor species (that could be a reduced flavodoxin *in vivo* or a reducing agent like Na$_2$S$_2$O$_4$ in the lab) to the cluster that reduces it to its [Fe$_4$-S$_4$]$^{1+}$ state, which has been shown to be the catalytically relevant state for these enzymes.[29] Electron transfer from the iron-sulfur cluster to SAM promotes the homolytic cleavage of the S-C$_{ribosyl}$ bond of the cofactor to produce methionine bound to the unique iron site, and a 5′-deoxyadenosyl radical intermediate dAdo$^{•}$. This last radical intermediate removes a hydrogen atom (H$^{•}$) from the substrate generating 5′-deoxyadenosine (5′-dAdoH) and as a result, a substrate-radical species.[16]

Two different mechanisms of S-adenosylmethionine are then reported for these enzymes
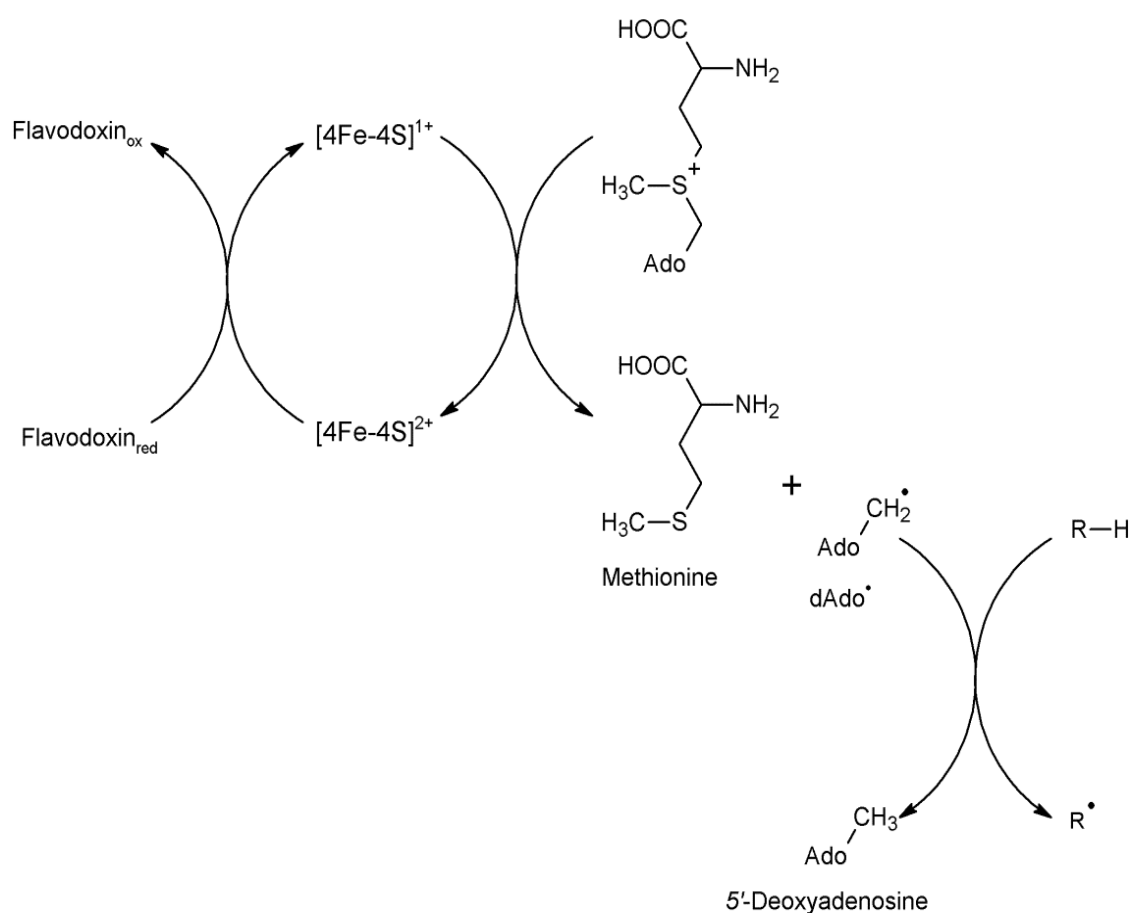
Figure 3. The cleavage of the AdoMet molecule to generate methionine and the dAdo$^{•}$ radical after one-electron transfer from biological reductant flavodoxin.

where SAM acts either as a cofactor or as a co-substrate.[18] Glycyl radical activating enzymes (GRE-AEs), such as pyruvate formate lyase (PFL-AE), make a stoichiometric use of SAM to oxidise the glycine residue of their respective protein substrates. The eventual formation of one equivalent of methionine and 5′-dAdoH species in PFL-AE is due to the irreversible removal of a hydrogen atom from a specific glycine residue present in the protein substrate upon the peptide enters the PFL-AE active site. A cysteine residue that chemically communicates through hydrogen-atom exchange with the glycyl radical thus formed, transfers a hydrogen to generate a cysteine-thiyl radical (Figure 4a). It is this last cystyl radical to initiates the mechanism attacking the substrate and undergoing its cleavage indefinite times. In the event of radical quenching of the activated enzyme, GRE-AEs can act in regeneration subsequent to the initial activation. Stoichiometric use of SAM is also required for sulfur inserting enzymes, which promote the insertion of sulfur atoms into unreactive C-H bonds (Figure 4b) such as the lipoic acid synthase (LipA)[30] and biotin synthase (BioB).[31] Information so far shows that the reaction is initiated by the 5′-dAdo• radical, but the mechanism is difficult to probe and no further detailed chemical mechanism is available so far. What is known is that these radical SAM enzymes both contain two FeS clusters: $2\times[Fe_4\text{-}S_4]$ in LipA and a $[Fe_4\text{-}S_4]$ and $[Fe_2\text{-}S_2]$ in BioB. One FeS cluster is involved in the regular AdoMet reductive cleavage, whilst the other one is proposed to provide the sulfur atoms for the sulfur insertion. It is likely that the substrate-radical intermediate, generated by attack of the dAdo• radical species after SAM reductive cleavage on the first $[Fe_4\text{-}S_4]$ cluster, captures sulfide ions from the auxiliary iron-sulfur cluster present in the protein.[32] Stoichiometric analysis indicated that at least two moles of SAM are cleaved to abstract the hydrogen atoms from their substrates; methionine and 5′-deoxyadenosine are thus released as by-products with each catalytic turnover.[8] The sulfur-inserting step is critical for the mechanism that has been widely discussed in the literature, yet it has to be confirmed because the experimental investigation of this step is rather difficult. As a matter of fact only just a single catalytic cycle can be achieved in the lab, after which inactivation of the enzyme takes place as the additional

FeS cluster is believed to function as a sacrificial S-donor during the turnover experiment.[31]

In the majority of radical SAM enzymes, abstraction of hydrogen atom from the substrate is

the starting point for more complex radical-mediated chemistry that converts the initial

Figure 4. Mechanisms of action of radical SAM enzymes. SAM used stoichiometrically in a) Glycyl radical activating enzymes (PFL-AE), b) Sulfur inserting enzymes (Biotin synthase). Catalytic use of SAM c) Substrate rearrangement (Spore Photoproduct Lyase, LipA, Lysine 2,3-aminomutase, LAM).
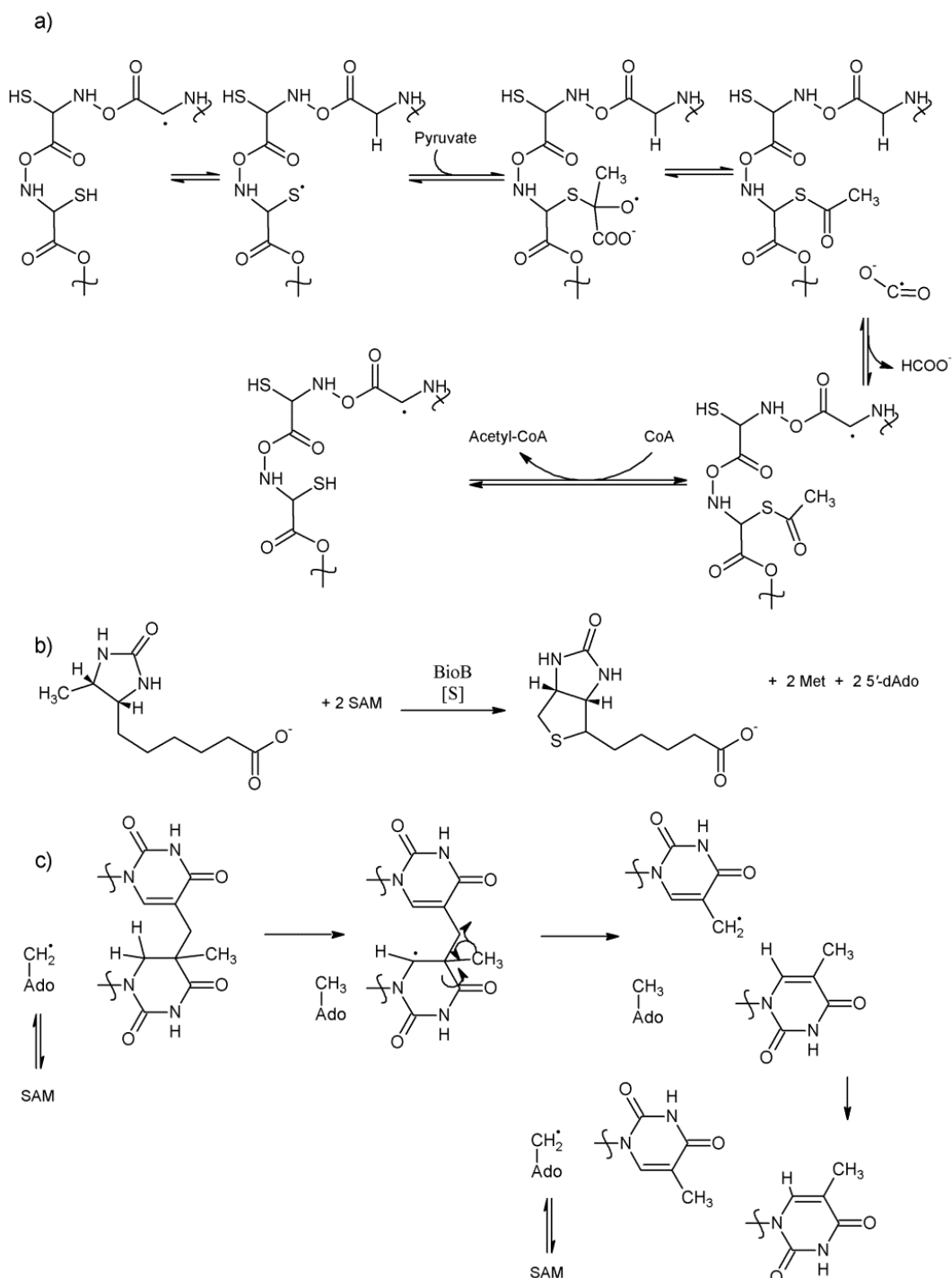
substrate radical into the final product (Figure 4c). In these reactions SAM chemically participates as a reversible source of radical species and it is renewed in each turnover. After cleavage of the S-C bond, the 5′-dAdo• radical abstracts a hydrogen atom from the substrate which, in turn, after rearrangement in the product radical, abstracts hydrogen atom from 5′-dAdoH to regenerate the 5′-dAdo• radical. Eventually, this last radical species reversibly binds to methionine, restoring SAM and replacing the electron in the $[Fe_4$-$S_4]$ cluster (Figure 4c). In this last case, SAM functions as a radical-generating cofactor in catalysis and is not consumed during the turnover. This last class of enzymes could be of considerable interest for biotechnological purposes given their potential to make use of SAM as a cofactor in a less wasting and cost-effective way. These catalytic SAM enzymes could then be useful in large-scale applications in the synthesis of natural products and pharmaceutical agents. Unfortunately, due to their air sensitivity, they require handling under strictly anaerobic conditions, and their industrial use is currently extremely limited.

1.3   Radical SAM enzymes in biotechnology

Members of the radical SAM superfamily have attracted interest for the biotechnological production of chemical compounds. Enzymes are generally preferred in industrial chemical syntheses for their high substrate-selectivity and their chiral, positional and functional group specificity.[33] Their use as biocatalysts could find application for the production of bulk and fine chemicals, antibiotics and bioactive molecules of interest or building-block molecules that can be used to pre-pare other biologically active compounds. An example is given by the production of β-aminoacids that can be obtained with the usage of recombinant microorganisms. β-aminoacid production is time-consuming to achieve by traditional chemical synthesis and requires expensive starting materials, resulting in a racemic mixture.[34] Radical SAM aminomutases catalyse the migration of the amino acid amino group from the α-carbon to the β-position (Figure 5). Glutamate 2,3-aminomutase (GAM) is a SAM

enzyme widely spread in bacterial and archaeal species and it is involved in the generation of β-glutamate in cells which serves as an osmoregulator. A biocatalytic process has been created to allow the formation of β-glutamate through isomerisation of its α-precursor.[35] Variants of 2,3-GAM from *T. tengcongensis* and *M. thermoacetica* were discovered that performed the isomerisation of glutamate and were thermostable with activity at 65 °C.[35] Lysine 2,3-aminomutase (LAM) catalyses the interconversion between α- and β-lysine. β-lysine is a precursor in the biosynthesis of bioactive compounds of interest such as nourseothricin,[36] viomycin[37] and streptothricin.[38] It also is a constituent of antibiotics produced by the fungi Nocardia, such as mycomycin[34] and it plays an important biological role as a precursor in the biosynthesis of antibiotics, anticancer agents, neurotransmitters and polymers.[39-40] The LAM enzymes from *C. subterminale*, *B. subtilis* and *E. coli* were overexpressed into the robust host *C. glutamicum* for the production of β-lysine. LAM from the first two bacteria give (S)-β-lysine as product of the bio-catalysed reaction whilst LAM from *E. coli* produces the (R).[41] The bioproduction of β-lysine through the LAM enzyme was already patented but it has not been industrialised due to the anaerobic conditions required for its handling. Of particular interest is the production of β-alanine achieved through directed evolution of 2,3-LAM. Ten mutations in 2,3-LAM from *B. subtilis* afforded conversation of α-alanine into β-alanine.[42] The engineered alanine 2,3-aminomutase (AAM, figure 5) was inserted in a biocatalytic process for the production of 3-hydroxypropionic acid, which it can be used as precursor for the production of acrylic acid, 3-hydroxypropionaldehyde, 1,3-propanediol as precursor for adhesives, polymers and other useful building blocks.[43] Several examples of potential usage of the radical SAM enzymes employed for biotechnological purposes have been reported over the years as they show potential for the biosynthesis of chiral compounds.[3] Unfortunately, one of the main obstacles for the large-scale bio-production of novel chemical compounds is represented at the enzyme purification stage. For isolated SAM enzymes, strict anaerobic conditions limit industrial use of these biocatalysts.[43] As radical SAM enzymes have gained interest in their industrial

development, current and increasing technologies might play a role for their progress as biocatalysts. The everyday-increase in genomic information could help in uncovering new radical SAM enzymes from their genetics and mechanistic information. Computer-led design technologies, coupled with experimental feedback loops or information retrieved from complementary experimental approaches offer a way to rapidly screen and redesign enzymes towards a specific use. Computational approaches are particularly attractive in the case of radical enzymes because they circumvent the need for anaerobic conditions during experimental handling.
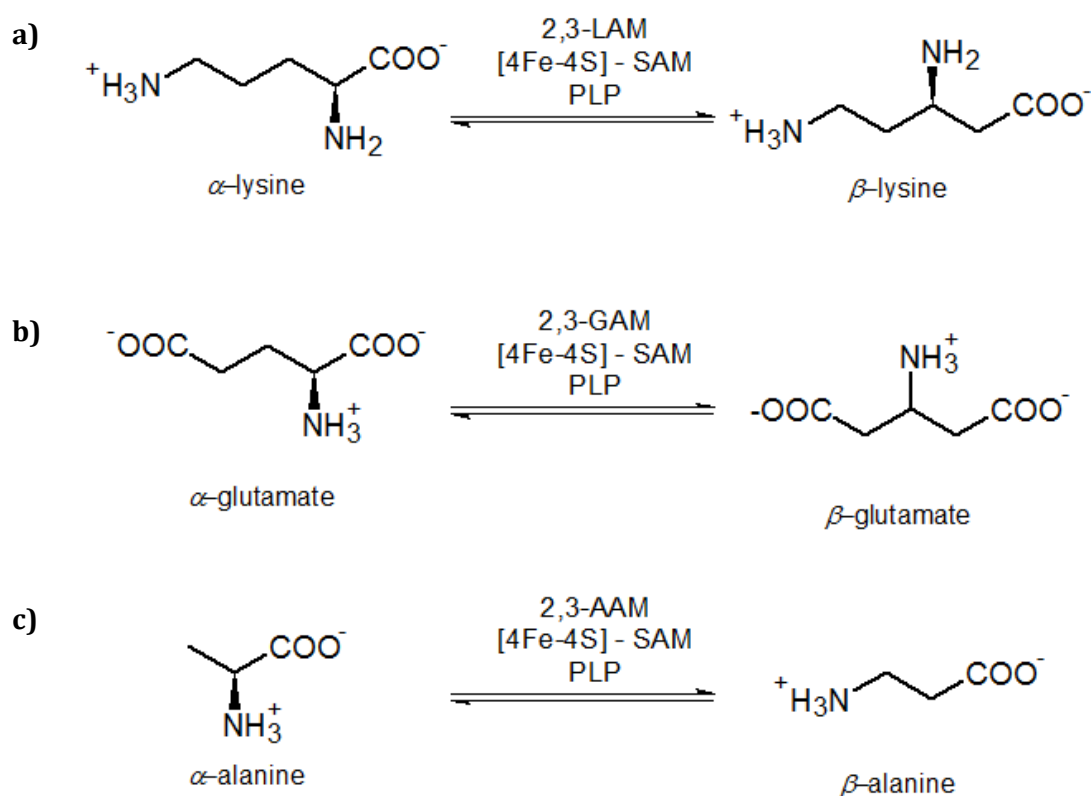
Figure 5. Three different examples of possible usage of radical SAM enzymes for biotechnological manufacture of compounds of interest. LAM 2,3-aminomutase (LAM, a), glutamate 2,3-aminomutase (GAM, b) and alanine 2,3-aminomutase (AAM, c).

1.4 The oxygen sensitivity of [Fe$_4$-S$_4$] clusters

Nearly every SAM enzyme that has been studied and characterised so far was found to be extremely sensitive toward oxygen. Reactive oxygen species (ROS) like hydrogen peroxide (H$_2$O$_2$), superoxide (O$_2^-$), hydroxyl radicals (OH$^\bullet$) and molecular oxygen (O$_2$) directly react with the [Fe$_4$-S$_4$] cluster, causing oxidative stress that decomposes the metal cofactor, inactivates the enzyme, and disrupts its catalysis (Figure 6). Because of the high affinity of oxygen for electrons, molecular oxygen and other ROS perform a one-electron oxidation, attacking the unique labile iron atom of the radical SAM enzyme [Fe$_4$-S$_4$] cluster.[44] The

$$[Fe_4S_4]^{2+} + O_2 \longrightarrow [Fe_3S_4]^+ + O_2^{\bullet-} + Fe^{2+} \tag{1}$$

$$[Fe_4S_4]^{2+} + O_2^{\bullet-} + 2H^+ \longrightarrow [Fe_3S_4]^+ + H_2O_2 + Fe^{2+} \tag{2}$$

$$[Fe_4S_4]^{2+} + H_2O_2 + H^+ \longrightarrow [Fe_3S_4]^+ + H_2O + Fe^{2+} + OH^\bullet \tag{3}$$
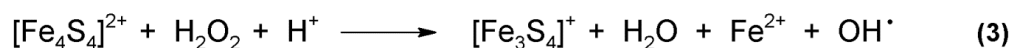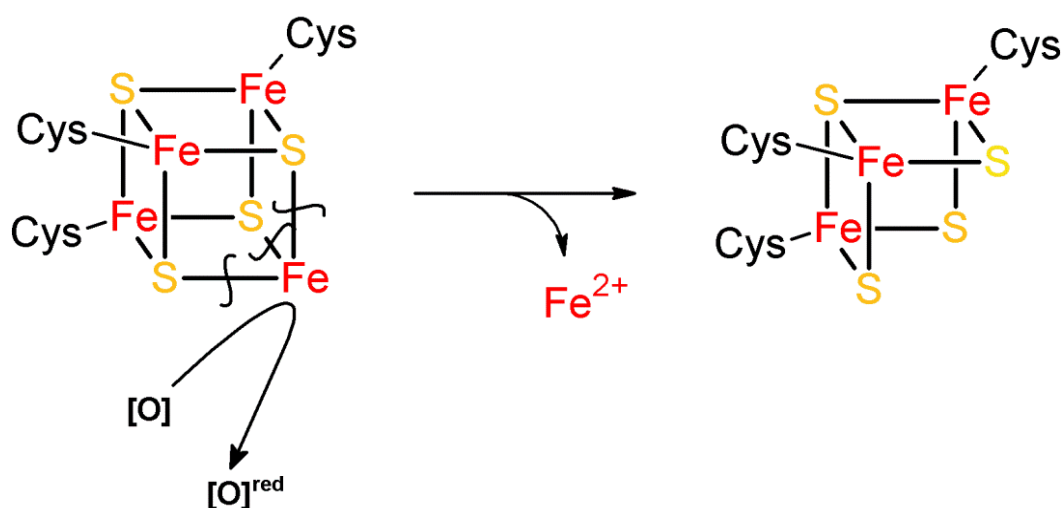
Figure 6. Cluster degradation by oxidative damage (above). Univalent oxidants, including superoxide, hydrogen peroxide, molecular oxygen (collectively denoted [O]), convert the exposed [Fe$_4$-S$_4$]$^{2+}$ cluster to the unstable +3 oxidation state. The cluster spontaneously decomposes to the inactive [Fe$_3$-S$_4$]$^{1+}$ form. Image adapted from Imlay.[45] Below, redox equations of reactive species oxidation (ROS) on the cluster.

thermodynamically favourable coordination of the oxygen atom oxidises the reduced cluster from the 1+ to 3+ oxidation state, which is extremely unstable and readily decomposes to the $[Fe_3-S_4]^+$ form releasing Fe^{2+} (Equation 1, figure 6).[45] The resultant $[Fe_3-S_4]^+$ cluster lacks the iron atom necessary for the catalysis, thus the enzyme is inactive and the pathway fails. The reactions of superoxide and hydrogen peroxide with iron-sulfur clusters are much quicker than molecular oxygen ($k_{O2} \approx 300$ M^{-1}·s^{-1},[46] $k_{O2^-} \approx 10^7$ M^{-1}·s^{-1}, and $k_{H2O2} \approx 10^5$ M^{-1}·s^{-1}[44]). As shown in equations (2) and (3), these reactions are proton-coupled electron transfer processes and the high reduction potentials of these two ROS species, make the reactions thermodynamically favoured and able to irreversibly oxidise $[Fe_4-S_4]^{2+}$ clusters to $[Fe_3-S_4]^+$ clusters and Fe^{2+}.[47,48]

Cluster protection strategies, which would be evidenced by oxygen-tolerant radical SAM enzymes, are rare, but not unknown.

1.5 Naturally occurring O$_2$-tolerant radical SAM enzymes

Within the several thousand enzymes belonging to the radical SAM superfamily, only two examples have been reported that maintain their activity in air. The first one to be reported was a variant of lysine 2,3-aminomutase from *Bacillus subtilis* that was identified to be structurally similar to its oxygen-sensitive counter enzyme from *Clostridium subterminale*.[49] These two enzymes share a tetrameric structure, as well as the presence of $[Fe_4-S_4]$ cluster, SAM and PLP cofactor in each enzyme subunit. The *Bs*LAM was identified as being stable to oxygen, allowing it to produce β-lysine under normal aerobic conditions and making the enzyme particularly useful as a natural biocatalyst. In contrast to its oxygen-sensitive counterpart, its β-amino acid synthesis activity is three orders of magnitude smaller, resulting in only millimolar concentrations of product under the conditions assayed.

The PqqE from *Methylobacterium extorquens* AM1 is a radical SAM-enzyme that catalyses the first step of the pyrroloquinoline quinone (PQQ) biosynthesis, a cofactor present in many

bacterial dehydrogenases.[50] This SAM enzyme contains three FeS clusters and SAM bound to the [Fe$_4$-S$_4$] cluster in the N-terminus region, and the SAM cleavage activity was recorded as retaining up to 48% and 27% activity in the His$_6$-tagged PqqE as-purified and reconstituted enzymes, respectively. This particular strain might be adopted for the elucidation of the molecular mechanism of the PQQ biosynthesis without the need to employ anaerobic conditions. Recently, an X-ray characterisation of the oxygen-tolerant PqqE was carried out[51] and its crystal structure solved (pdb code: 6C8V). The enzyme would constitute another interesting case study of an air-sensitive radical SAM protein, so furthering the field of the unusual tolerance to oxygen in the radical SAM superfamily and gain access to novel molecular machines.

1.6 Natural and radical cluster protection through protein structural changes

Only a few examples of cluster protection strategies are known among radical SAM enzymes. Nearly all members of the superfamily have their [Fe$_4$-S$_4$] cluster solvent-exposed during the catalytic cycle to better facilitate the entrance of the substrate inside the active site. When the SAM cofactor binds to the cluster and the substrate is inside the active site, the cluster is sealed from water to maintain its catalytic activity. Recently, a Mössbauer study on PFL-AE showed that it can keep on with its normal function inside whole cells under aerobic conditions, while cluster degradation occurs for the purified protein. More than 44% of the total iron content in the whole cells was found to be in the [Fe$_4$-S$_4$] cluster form after two hours of over-expression in recombinant *E. coli* cells under aerobic conditions.[52] Degraded cluster forms were instead detected in the purified protein directly after oxygen exposure. A cluster protection mechanism is speculated to occur *in vivo*, perhaps due to coordination of the unique iron atom of the cluster with abundant small molecules in cells, such as AMP.[52] This coordination may prevent the cluster from oxidative damage upon air exposure without

affecting the enzyme activity, as it can be displaced by the co-substrate SAM.

Additionally, several of the crystal structures solved so far revealed the presence of loops and helices that may undergo conformational changes upon the substrate entering the active site. The loop movements help in burying the active site for radical chemistry, avoiding oxygen exposure of the cluster in its catalytically-active form during catalysis and to protect radical intermediates from unwanted side reactions.[16]

In biotin synthase, an interaction between substrate's carboxylate group and the amides of Thr293 and Thr292 upon substrate binding, may promote a movement of a loop, closing the top of the barrel.[53] A loop containing a conserved sequence motif among glycyl radical activating enzymes, has been observed moving, upon substrate binding, in crystal structures of pyruvate formate-lyase activating enzyme (PFL-AE).[14] The motif interacts with the peptide substrate, inducing a conformational change that was proposed to be critical for the activation of the enzyme and in helping seal off the active site from the solvent. Two loops were proposed playing this role in the radical SAM enzyme coproporphyrinogen III oxidase (HemN). Two molecules of SAM are bound within the HemN structure, with one bound to the [Fe_4-S_4] cluster, whilst the second one is placed deep in the N-terminal domain and close to the first coordinated SAM. The two SAM molecules, once bound, provides further protection to the oxygen-sensitive cluster and the active site from solvent by sealing them off from the solvent.[16] The SAM molecules are not only coordinated to the unique Fe atom but they are also held in position by specific amino acid residues that are involved in electrostatic, hydrogen-bonding, hydrophobic, and π-stacking interactions. All these interactions help in keeping the SAM molecules buried from solvent and placed in position for an optimal electron transfer from the [Fe_4-S_4] cluster.

In the *Clostridium sticklandii* lysine 5,6-aminomutase, an AdoCbl PLP-dependent enzyme, a locking mechanism that keeps the adenosylcobalamin out of the catalytic pocket in the absence of substrate was observed preventing unwanted radical generation.[53] The substrate-free crystal structure was solved and it was speculated that the Rossmann domain of the

enzyme, covalently binds the PLP cofactor through an aldimine adduct formed with the enzyme Lys144. The Rossmann domain works by positioning the PLP into the neighbouring TIM-barrel domain where the active site is while at the same time it places AdoCbl cofactor 25 Å away from the active site. The cofactor thus works as an anchor by hitching the Rossmann domain to the TIM barrel domain. Following substrate binding and PLP transaldimination, the Rossmann domain would have the freedom to rotate, bringing the AdoCbl cofactor, PLP and substrate close to each other and sealing them off from solvent. Another large-scale domain motion was speculated to take place after substrate-binding in the *Clostridium sticklandii* ornithine 4,5-aminomutase (OAM),[54] another AdoCbl enzyme requiring the PLP cofactor. In its substrate-free mode, an open conformation was revealed where the Rossmann domain, which harbours the AdoCbl cofactor, slopes toward the top of the TIM barrel domain where the PLP cofactor binds to a specific enzyme-based lysine. In this pre-catalytic mode, the two cofactors are held at a distance of 23 Å and it was proposed that upon substrate binding, a large domain motion of the Rossmann domain is induced to bring the AdoCbl cofactor on top of the TIM barrel so to start the catalytic turnover in a solvent-free environment.

1.7 Artificially achieved oxygen tolerance in Fe S cluster-containing enzymes by prevention of oxygen interaction

Studies on FeS cluster-containing enzyme mutants, carrying different amino acids from the wild type, have been reported with evidences of change in their air-stability.[55,56,57] As molecular oxygen itself is not very reactive, it requires a direct interaction with the [Fe – S] cluster before the electron transfer can occur. Thus, by either preventing the interaction between oxygen species and the FeS moiety or by blocking O_2 access, the cluster stability could be greatly improved.

In fumarate and nitrate reductase (FNR) from *E. coli,* oxygen resistance was able to be

improved as shown in the work by Overton et al.[55] Fumarate and nitrate reductase regulatory proteins are O_2 sensors that regulate the switch between aerobic and anaerobic metabolism. FNR harbours in its N-terminal domain four cysteine residues that can bind either a $[Fe_4\text{-}S_4]^{2+}$ cluster under anaerobic conditions or $[Fe_2\text{-}S_2]^{2+}$ in an aerobic environment. Cluster degradation occurs with oxidation of the $[Fe_4\text{-}S_4]^{2+}$ by molecular oxygen producing a ferrous ion, a superoxide ion and a $[Fe_3\text{-}S_4]^{1+}$ intermediate that spontaneously decays yielding a $[Fe_2\text{-}S_2]^{2+}$, two ferric and sulphide ions. *E. coli* FNR mutants with an amino acid phenylalanine replacement adjacent to one of the four bridging cysteine residues (S24F mutation) have enhanced aerobic activity by 3.5-fold *in vivo*.[46] Later, Jervis et al.[58] showed with a model representing the loop containing the cysteine residues that Ser-24 is placed in a manner where there is a solvent access to the FeS cluster attached to Cys-23 that might be a primary target of O_2 attack. The replacement with the bulky phenylalanine residue shields the iron atom of the cluster attached to the sulfur atom of Cys-23 that was predicted as being in close proximity to this aromatic amino acid. The increased oxygen resistance in FNR was also known for other bulky substitutions like Arg, His, Trp and Tyr.[58]

For some $[Fe_2\text{-}S_2]$ ferredoxin enzymes, replacement of specific amino acid residues next to the cluster-bridging cysteines, can affect the way the enzyme deals with oxygen, improving its resistance in air. Singh et al.[56] worked on cyanobacterial ferredoxins that function as one electron carriers. They found that in *Anabaena variabilis* 29413 cells there are two genes encoding two $[Fe_2\text{-}S_2]$ ferredoxins, FdxH1 and FdxH2 that show a different sensitivity to molecular oxygen: while FdxH2 loses 50% of its activity in air after 1.5 h, FdxH1 can retain up to 80% of its activity after 24 h. The ferredoxins' crystal structures showed both the cluster iron atoms tetrahedrally coordinated by four cysteine residues, and two inorganic sulphides to form a planar ring. They also shared the same global folding pattern and 76% of amino acid sequence identity. To identify which residues might have played a role in oxygen sensitivity, they aligned and compared the amino acid sequences of the two enzymes looking for possible residue exchanges that possibly make a ferredoxin more tolerant to oxygen attack. They

subsequently created FdxH2 mutants through site-directed mutagenesis (Figure 7). At first, they found that the mutant protein Q42H was even more oxygen-sensitive than the wild type, losing 80% of its activity in 2 hours. They subsequently created a double mutant of the FdxH2 ferredoxin carrying the I76A and V77L residue substitutions and they observed an increased stability in air that was comparable with the FdxH1 ferredoxin. By modelling the ferredoxin, it was highlighted that there was a cavity formed near the [Fe_2-S_2] cluster close to the Cys49, which provided enough space for dioxygen to get into and stay close to the cluster, whereupon it was oxidised. For FdxH1, the leucine residue at position 77 is able to fill such a cavity, sufficiently to block the entrance of dioxygen and in the same way, the V77L mutations worked on the sensitive FdxH2 enzyme to increase its tolerance to oxygen.[59]

[NiFe]-hydrogenases are enzymes that catalyse the reversible oxidation of H_2 into protons and electrons. Such a reaction is based on a complex interaction that involves several [Fe_4-S_4] clusters and the [Ni-Fe] active site. The process occurs mainly in anaerobic environment
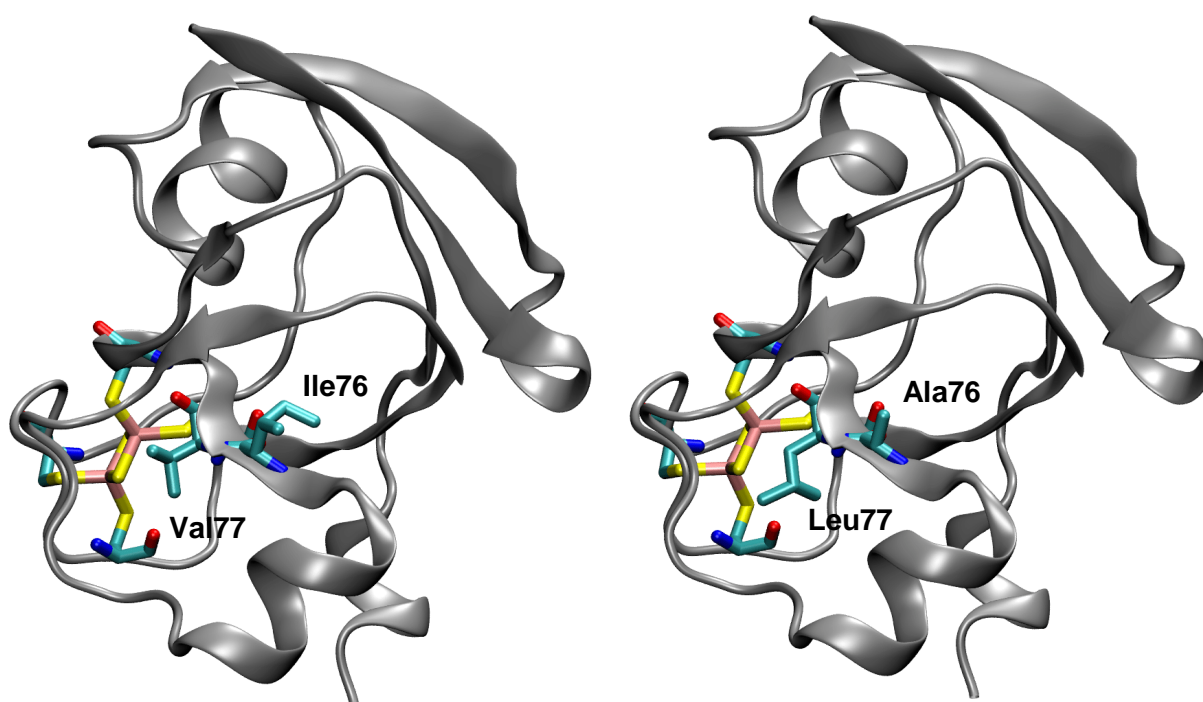
Figure 7. The active site of the FdxH2 ferredoxin crystal structure (on the left) and the model carrying the I76A and V77L mutations (on the right). The steric hindrance created by the substituted amino acids could prevent the attack of oxygen species to the metal cluster so to improve the resistance in the ferredoxin enzyme. Image adapted from Singh *et al*.[59]

because most hydrogenases are very sensitive to O_2-inhibition with inactivation of the active site. In [NiFe]-hydrogenases, a channel network allows gas diffusion between the protein surface and the active site.[60] At the end of the gas channel of O_2-sensitive hydrogenases, close to the [NiFe] site, two conserved hydrophobic residues most commonly valine and leucine, are swapped with isoleucine and phenylalanine respectively, that are present in the oxygen resistant subclass of H2-sensing regulatory hydrogenases.[61] Buhrke *et al.*[62] established that naturally occurring oxygen-resistant [NiFe]-hydrogenases had one or more conserved residues close to the active site that were larger in size than those found in the same region of the oxygen-labile [NiFe]-hydrogenases. In their work, they created enzyme mutants of an oxygen tolerant [NiFe]-hydrogenases by replacing the conserved isoleucine and phenylalanine with the less bulky valine and leucine residues by site-directed mutagenesis. The final analysis revealed a decrease in oxygen resistance, supporting the theory that enlargement of the gas channel makes the active site accessible to oxygen and ROS species. The bulkier sidechain of those specific amino acid residues in the oxygen-resistant enzyme may act as an obstruction to the flow of oxygen towards the active site that physically prevents the interaction between the metal cluster and oxygen molecules.

Dementin *et al.*[57] reported mutagenesis experiments to convert O_2-sensitive [NiFe]-hydrogenases into enzymes with increased oxygen-resistance based on specific residue replacement. They engineered V74I and L122F mutants of the O_2-sensitive *D. fructosovorans* [NiFe]-hydrogenase but unexpectedly the replacement with bulkier residues in the proximity of the active site did not afford increased oxygen tolerance as predicted. Inspired by the role in oxidative stress responses and affinity of their sulfur atom for ROS,[63,64] they created a mutant bearing methionine residues at positions 74 and 122 of the [NiFe]-hydrogenase. The hypothesis was that such residues placed at the entrance of the active site cavity might protect the [NiFe] site from oxidation by reacting to form weak S-O bonds and thus facilitating the evacuation of the bound oxygen species from the active site.[65] They showed that the oxygen diffusion rate in the V74M/L122M mutant decreased by nearly three orders of magnitude

when exposed for several minutes to high O_2 concentrations indicating that mutated hydrogenases become more tolerant to oxygen and confirming chemical trapping of oxygen as a valid mechanism.

1.8   Achieved oxygen-tolerance by redox modulation

For some other [Fe$_4$-S$_4$] cluster-containing enzymes, the ability to develop tolerance to oxygen in nature, resides in a difference in redox potential of the metal cluster, regulated by individual amino acid residues, as in the case of ferredoxins.

In some cases, the stability of the [Fe$_4$-S$_4$] cluster against degradation by oxidative stress under aerobic conditions can be correlated with the reduction potential of the oxidative process, as shown by Tilley *et al.*[66] and summarised in Figure 8. They examined the process of oxidative degradation of an all-cysteine-ligated [Fe$_4$-S$_4$] cluster into the [Fe$_3$-S$_4$] form in ferredoxins and their mutants. Ferredoxins are enzymes that mediate electron transfer in a variety of metabolic reactions and they can contain FeS clusters of different shapes and oxidation states. A selection of ferredoxins were taken into account and they either contained a cuboidal-type [Fe$_4$-S$_4$] cluster that operates between +2 and +1 oxidation levels or, after degradation upon exposure to air, a [Fe$_3$-S$_4$] cluster in a cuboidal conformation that switches between +1 and 0 levels, while some of the ferredoxins analysed contained two [Fe$_4$-S$_4$] clusters. Typically, the redox potential of [Fe$_3$-S$_4$] clusters differs by between 93 to 350 mV from that of [Fe$_4$-S$_4$] clusters, toward more positive values.[67] In their work, Burgess and co-workers, showed that the stability of [Fe$_4$-S$_4$] clusters towards oxidation, in reconstituted 8Fe ferredoxins containing two [Fe$_4$-S$_4$] clusters, under normal aerobic conditions was determined by the reduction potential of the oxidative processes. They presumed to have found a relationship between cluster stability upon exposure to air and the potential of the [Fe$_4$-S$_4$] cluster oxidation process (the 'transition potential') that converts it to the corresponding [Fe$_3$-S$_4$] derivative. Among the ferredoxins analysed there was the *Azotobacter*

vinelandii FdI which contains an extremely air-sensitive $[Fe_4-S_4]$ cluster that degrades at a potential of +420 mV, while the mutant of the wild type carrying a deletion of two residues next to a cluster-binding cysteine, $\Delta Thr^{14}/\Delta Asp^{15}$ *Av*FdI, shows oxidation at potential +532 mV. The ferredoxin III from *Desulfovibrio africanus* displayed an oxidation peak relating to cluster degradation at +434 mV and, as well as the previous proteins, it is known to be air-sensitive (Figure 8).[66] The reconstituted 8Fe form of *Av*FdI mutant with the wild-type threonine residue at position 14 substituted with a cysteine, *Av*T14C, inserts an additional cysteine residue into the $[Fe_3-S_4]$ cluster-binding region. This mutant attempted to create a typical cluster-binding motif CxxCxxC. This variant revealed an oxidation peak at +610 mV that was broader than the previous proteins and it resulted with an increased oxygen-tolerance compared to the wild-type. In addition, the mutant *Av*C42D, which carried the
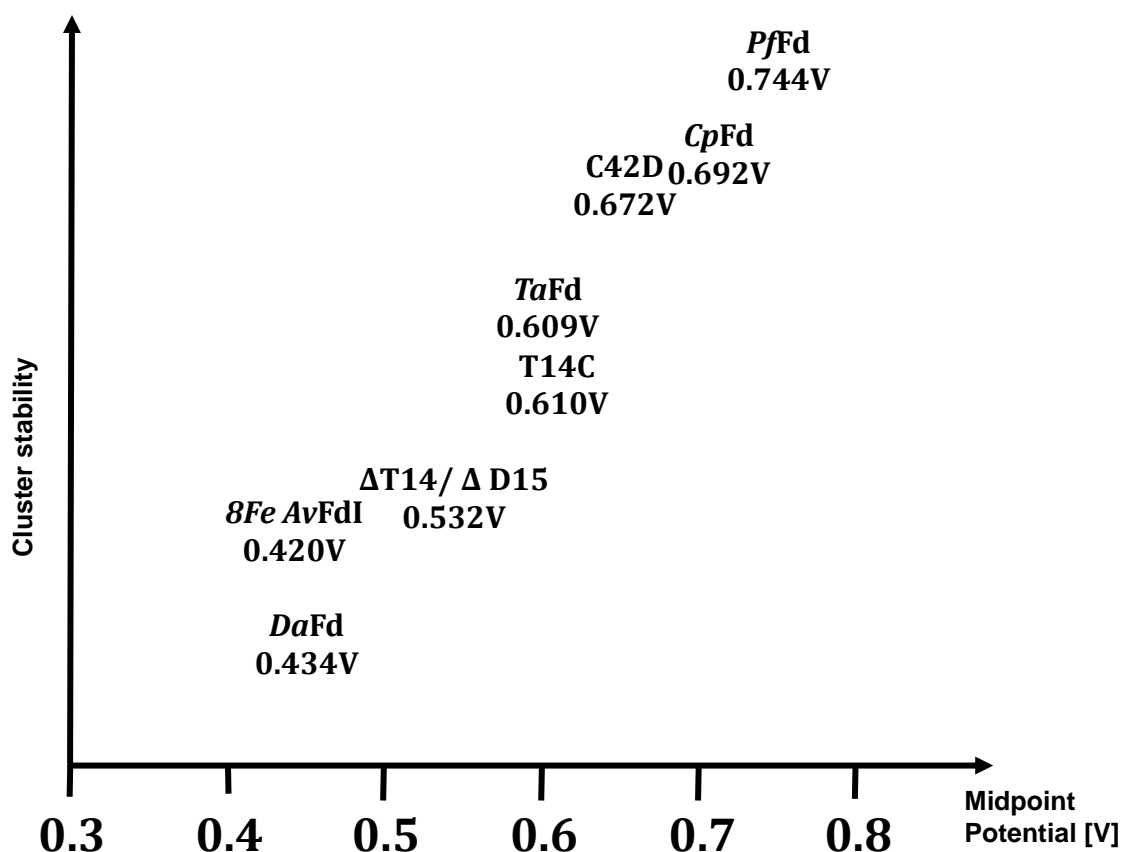
Figure 8. Correlation between oxygen-sensitivity in Ferredoxin variants and midpoint potential values of the $[Fe_4-S_4]$ cluster oxidative transformation potentials. The image was adapted from the work by Tilley *et al*.[66]

substitution of an aspartate residue into the central FeS cluster-coordinating cysteine producing a CxxDxxC motif, showed a resistance of the [Fe$_4$-S$_4$] cluster toward chemical oxidation and displayed a high oxidation potential of +672 mV.[68] The ferredoxin from *Pyrococcus furiosus*, *Pf*Fd, contains a [Fe$_4$-S$_4$] cluster *in vivo* that is slowly converted into the [Fe$_3$-S$_4$] form in air. *Pf*Fd presented an oxidation peak at +744 mV and was reported to be electrochemically inert, as there was no evidence of the formation of a [Fe$_3$-S$_4$] cluster.[69] The ferredoxin from *Thauera aromatica, Ta*Fd contains two [Fe$_4$-S$_4$] clusters and is not known to be sensitive to oxygen *in vivo*,[70] reporting a peak at +609 mV. The *Clostridium pastearium* ferredoxin, *Cp*Fd, contains two all-cysteine-ligated [Fe$_4$-S$_4$] clusters that degrade into [Fe$_3$-S$_4$] form upon extended contact with air while showing a peak at +692 mV.[71] In summary, there seems to be a good relationship between air-sensitivity reflected by the facile formation of the [Fe$_3$-S$_4$] cluster and the redox potential of the ferredoxins analysed that appeared to be sensitive to oxidation at midpoint potentials lower than +450 mV.

Experiments on ferredoxins (Fd) and HiPIP (high potential iron-sulfur proteins) enzymes,[72] both carrying the [Fe$_4$-S$_4$](Cys)$_4$ cluster in their active site, suggested that the difference in their midpoint potential was due to a different number of hydrogen-bonding interactions around the cluster. The redox potential of the [Fe$_4$-S$_4$](Cys)$_4$ clusters contained in Fds have the characteristic low reduction potentials that ranges from -650 to -250 mV at the oxidation couple of +2/+1.[73] On the other hand, HiPIPs have more positive potential ranging from +50 to +450 mV,[74] and switch between +3 and +2 oxidation states despite the identical cluster geometry in the two enzymes. There is a difference in the number of hydrogen-bond interactions, springing from the backbone amide moiety to the terminal and bridging cluster sulfur atoms (N-H--S), with eight of these interactions in Fds and five identified in HiPIPs. The larger number of N-H--S bonds in ferredoxins was speculated as contributing to the lower midpoint potential because more hydrogen-bonds could help stabilise the more negatively charged reduced form of the cluster by delocalising the additional negative charge on sulfur atoms.[75] Later calculations showed that the FeS bond is of high covalency,[76] with greater

covalent character for interactions with the Fe^{3+} atoms than for the Fe^{2+} sites of the metal cluster, whilst the bridging cluster sulfides were observed to be better donors than thiolates. Furthermore, the sulfur charge becomes more negative in the reduced state because the extra electron density is mainly localized on the sulfur atoms. In this state, the iron−sulfur cluster should participate in more hydrogen-bonding interactions with the surrounding protein environment. These hydrogen bonds would compete with the sulfur-to-metal charge transfer that leads to a decreased covalency. The amide backbone groups are prone to extend the hydrogen-bonding network on the metal cluster in its reduced state. Carvalho *et al.*[77] later confirmed that the protein environment plays a major role in stabilising the reduced cluster by shortening the FeS bonds, highlighting that the coordination sphere provided by the protein environment is the most significant determinant of electronic and functional properties in FeS cluster-containing enzymes.

Stephens *et al.*[78] speculated that the hydrogen-bonding network is a main element in tuning the redox potential and the more the amide groups are oriented to favour Coulombic interactions, the higher the redox potential. However, the midpoint variation derives from a reciprocal interaction of factors that does not easily correlate with any specific variable. It is then more difficult to understand the small redox variations among similar proteins (e.g. among ferredoxins) rather than a bigger variation in midpoint potential between unrelated enzymes (e.g. between ferredoxin and HiPIP enzymes).

From the site-directed mutagenesis studies of Swartz *et al.*,[79] the polar groups surrounding the FeS cluster were taken as a factor to explain the redox potential modulation between homologous proteins. The polar contributions derive from the protein backbone and polar side chains, and different residues were observed as providing different contributions to the electrostatic potential, thus a different enzyme redox potential. All uncharged amino acid residues except for Ala, Gly, Ile, Leu, Phe, Pro, and Val have polar properties and the backbone contribution arises from both parts of the peptide bond with N, H and C atoms on the amide end, C and O on the other end. A group of closely-related rubredoxin enzymes (Rd, small

enzymes carrying the prosthetic group made by a unique iron atom coordinated by four cysteine ligands) have been used to show the replacement of residues close to the metal cofactor can either, positively or negatively affect the cluster midpoint potential. Replacement of valine 44 in *C. pasteurianum* Rd with alanine, which is present at the same position of different rubredoxins, increases its midpoint potential by 20mV. Mutation A44V on other Rds with native Ala44, decreased the midpoint potential by the same amount: the valine side chain is bigger than alanine and it moves the amide portion of the residue farther from the metal cluster so the hydrogen bond is weaker, correlating with a lowering of the cluster potential. The threonine at position 5 of *Cp*Rd (or the equivalent position in other Rds) was found to bring a negative contribution to the electrostatic potential and if replaced by the valine that is present in other Rds, the redox potential increased by at least 90 mV. The asparagine at position 22 present in two rubredoxins, with the dipole moment of its sidechain gives a positive contribution of 70 mV and its replacement with leucine decreased the cluster redox potential. Through these assessments, the polar groups speculated to be relevant for the electrostatic contribution were narrowed to within 8 Å from the metal cluster. Short-range interactions were confirmed to be the most relevant factors in the redox potential modulation in rubredoxins, with a stronger hydrogen bonding network leading to more positive midpoint potential.[80]

A study focused on the *Saccaromyces cerevisiae* Rieske-type proteins containing the [Fe_2-S_2] cluster,[81] identified two redox-modulating residues, Ser183 and Tyr185, which were found to form hydrogen bonds with a sulfur atom of the cluster and a cysteinyl ligand (Figure 9). The residues took part in the hydrogen bond network and were likely to contribute to the positive midpoint potential of the Rieske protein (+285 mV) by decreasing the negative charge density on the sulfur atoms through their respective hydrogen bonds. The two amino acids were swapped with few other residues but only substitutions with alanine and phenylalanine were observed to result in a significant decrease of the cluster redox potential. Substitutions of serine with alanine and tyrosine with phenylalanine, disrupted the hydrogen bond made by

the hydroxyl moiety of the native residues with consequent lowering of the protein midpoint potential and reduction in enzyme activity. The S183A mutation alone contributed to the redox potential lowering by 130 mV, whilst the Y185F contributed by decreasing the redox potential by 65 mV. In the wild-type, the –OH group of Ser183 was found to be hydrogen-bonded to S in the cubane cluster and Tyr185 binds to S of a cysteinyl ligand. Such hydrogen-bonds in the native protein are likely to decrease the electron density on the metal cluster to ease the delocalisation of the extra negative charge upon reduction of an electron.

The midpoint potential of menaquinol-oxidising *bc* complexes is comparable to the S183A Rieske mutant where the alanine residue was found among the native sequences instead of serine.[82, 83]

Yeast strain	$E°_{(mV)}$
WT	+285
S183T	+259
S183A	+155
Y185F	+217
Y185F, S183A	+105
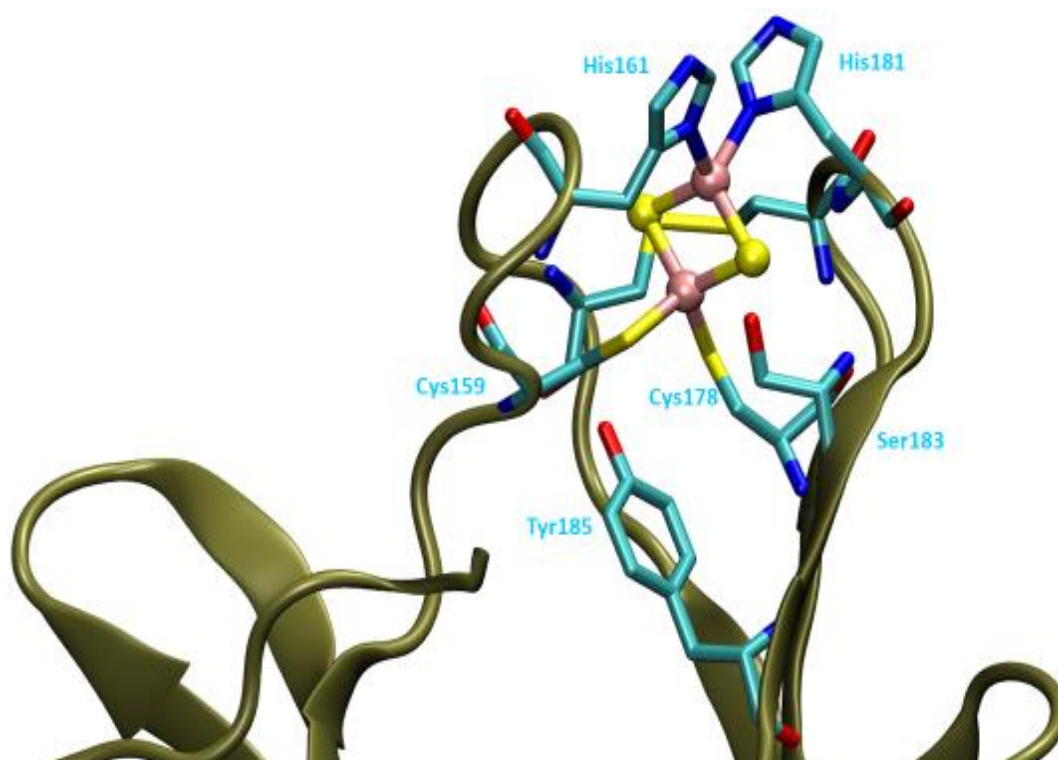
Figure 9. The table at the top reports few of the redox potential measured on the *S. cerevisiae* Rieske-type mutant proteins. Picture at the bottom, view of the *S. cerevisiae* Rieske-type protein near the [Fe₂-S₂] cluster. Mutations on Tyr185 and Ser183 lowered the midpoint potential of the metal cluster because of the disruption of the H-bonds created by the –OH group of the amino acids with sulfur atoms determination of the reduction-oxidation midpoint potential of the *bc₁* complex of the cluster and ligand. Image adopted from Denke et al.[81]

1.9   *Clostridium subterminale* lysine 2,3-aminomutase (*Cs*LAM)

The interconversion of the amino acids L-α-lysine and L-*β*-lysine is catalysed by the enzyme lysine 2,3-aminomutase (LAM) that was initially identified in *Clostridium subterminale* bacteria by Barker.[84] This oxygen-sensitive protein is one of the main representatives of the radical SAM enzyme superfamily and, like many amino acid metabolising enzymes, LAM was found to contain pyridoxal-*5'*-phosphate (PLP) as a co-substrate. LAM's role is split between the production of *β*-lysine for the biosynthesis of antibiotics and the metabolism of lysine in anaerobic organisms.[85] LAM has a tetrameric structure with an overall molecular mass of ~192 kDa, which is composed by two domain-swapped dimers linked by zinc-coordination by three cysteine residues of a monomer and a fourth from the opposite subunit. Each monomer consists of an external N-terminal domain and an internal C-terminal loop that enclose the central globular domain composed by a partial $(\beta\alpha)_6$ TIM barrel where the active site is found along with the buried [Fe$_4$-S$_4$] cluster, SAM and PLP in the catalytic pocket, an N-terminal domain and a C-terminal domain.[86] The enzyme requires a PLP cofactor molecule to facilitate the 1,2-imino rearrangement of the substrate *via* formation of the lysine-PLP aldimine radical. When the substrate is not present in the catalytic pocket, the PLP is covalently bound to a backbone lysine present in the PGGGGK motif (Lys337) that is typical of the LAMs of several (if not all) species identified so far. PLP then undergoes a transaldimination reaction with the lysine substrate once the latter one is bound to form the PLP-Lysine aldimine adduct that is in close contact with S-adenosylmethionine to start the catalytic cycle. The mechanism of LAM is illustrated in Scheme 1. Reduction of the cluster in nature, occurs by electron transfer from a physiological reductant such as flavodoxin, ferredoxin or flavodoxin-NADP+ reductase[87] with a binding site proposed to be the surface nearest the cluster-binding loop, formed mainly by this loop, with in some cases additional residues from loops following strands β2 and β4.[88] The secondary structure representation of LAM is presented in Appendix 6.2. In the lab, thiolate sets the cluster in the catalytically active state where it transfers an electron to SAM and initiates the cleavage to methionine to

obtain the dAdo• active radical (Scheme 1). The radical abstracts the 3-pro-*R* H-atom of the bound lysine to yield a substrate radical intermediate with the unpaired electron at lysine C3 and the dAdoH species.

The resulting radical is expected to undergo rearrangement to an aza-cyclopropylcarbinyl radical that eventually turns into the *α*-radical of *β*-lysine following a ring opening. The *α*-radical species abstracts a hydrogen from the dAdoH intermediate which forms back the SAM cofactor while the *β*-lysine will be released and PLP goes back to binding to the protein Lys337 until the next catalytic cycle.[89] The reaction works with migration of the hydrogen
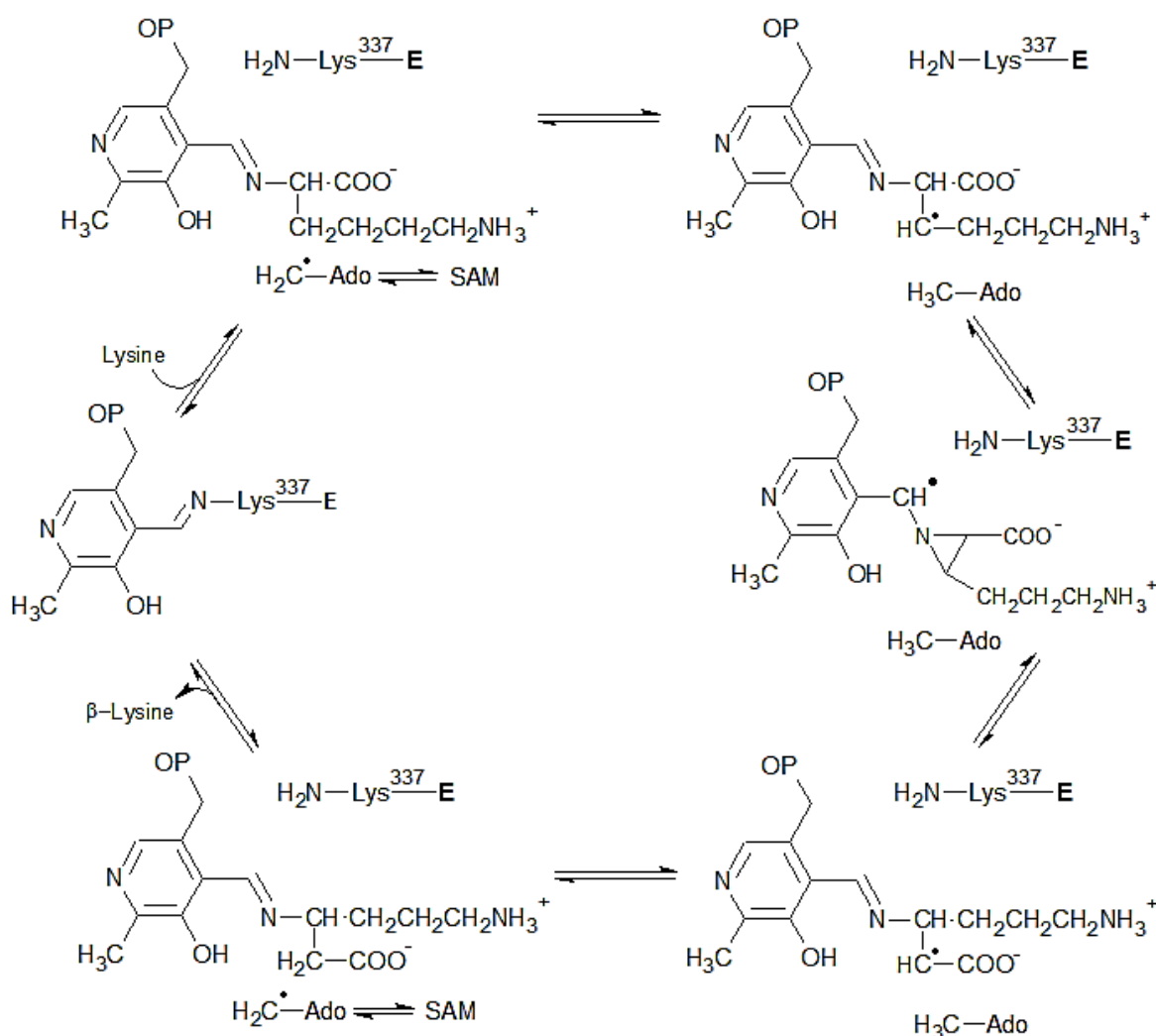
Scheme 1. Reaction mechanism of lysine 2,3-aminomutase (LAM). The radical intermediate dAdo• takes an electron from the PLP-Lysine adduct triggering the complex rearrangement that will provide the final product, β-lysine.

atom from carbon 3 of lysine with consequent interchange of the α-amino group of lysine to the β-carbon with inversion of configuration both at C2 and C3 carbon atoms.[90]

PLP and α-lysine are kept in position by a series of ionic contacts and hydrogen-bonding interactions in order to correctly position the lysine for abstraction of the 3-pro-*R* hydrogen by the dAdo$^\bullet$ radical intermediate. Some of the conserved residues necessary for LAM catalysis are Asp293 and Asp330, which bind to the ε-aminium group of lysine, and Arg134, which binds to the lysyl-carboxylate group. Mutation of any of these residues completely inhibits the enzyme activity.[91]

In order for the catalytic reaction to start, the SAM molecule must correctly bind to the FeS cluster. The correct binding occurs when SAM is positioned at a closed distance to the [Fe$_4$-S$_4$] cluster and it coordinates when both the amino- and carboxylate-SAM moiety are both at a distance of ~2 Å from the unique Fe atom of the cluster.[92] The distance of the SAM-S atom binds at a distance of ~3.2 Å to the unique Fe atom to which SAM is bound.

Energetic considerations regarding the SAM cleavage and the *Cs*LAM enzyme redox properties, highlighted by Prof. Frey and coworkers[93] indicated that the midpoint potential of the irreversible one-electron transfer to generate trialkylsulfonium ion in aqueous solution was on the order of -1.8 V (Figure 10). The [Fe$_4$-S$_4$]$^{2+/+}$ midpoint reduction potential exhibited a value of -430 mV at the enzyme resting state, i.e. when SAM and PLP are bound, whilst binding of the lysine substrate lowered the redox potential by ~150 mV. The redox potential for the SAM cleavage though, was estimated to be -990 mV in the LAM-SAM-lysine complex, suggesting that the enzyme active site environment raises the midpoint potential by ~810 mV, while lowering the [Fe$_4$-S$_4$] redox potential upon substrate binding.[94] This data establishes that the electron transfer from the reduced FeS-cluster to the coordinated SAM cofactor is extremely unfavourable with a barrier of ~32 kcal mol^{-1} thus leaving the question of how LAM could undergo its catalytic mechanism. Furthermore, such a low midpoint potential makes the [Fe$_4$-S$_4$] cluster inaccessible to reduction by physiological electron donors, such as flavodoxin. The redox potential of the [Fe$_4$-S$_4$] cluster was then found to

increase by ~50 mV upon coordination of SAM, likely due to the electrostatic effect brought by the positively charged sulfonium ion of SAM, as cations close to the Fe centre help stabilise the ferrous form of the cluster.[93]

Binding of SAM contributes 19 kcal mol^{-1} to lowering the energy barrier. The presence in the
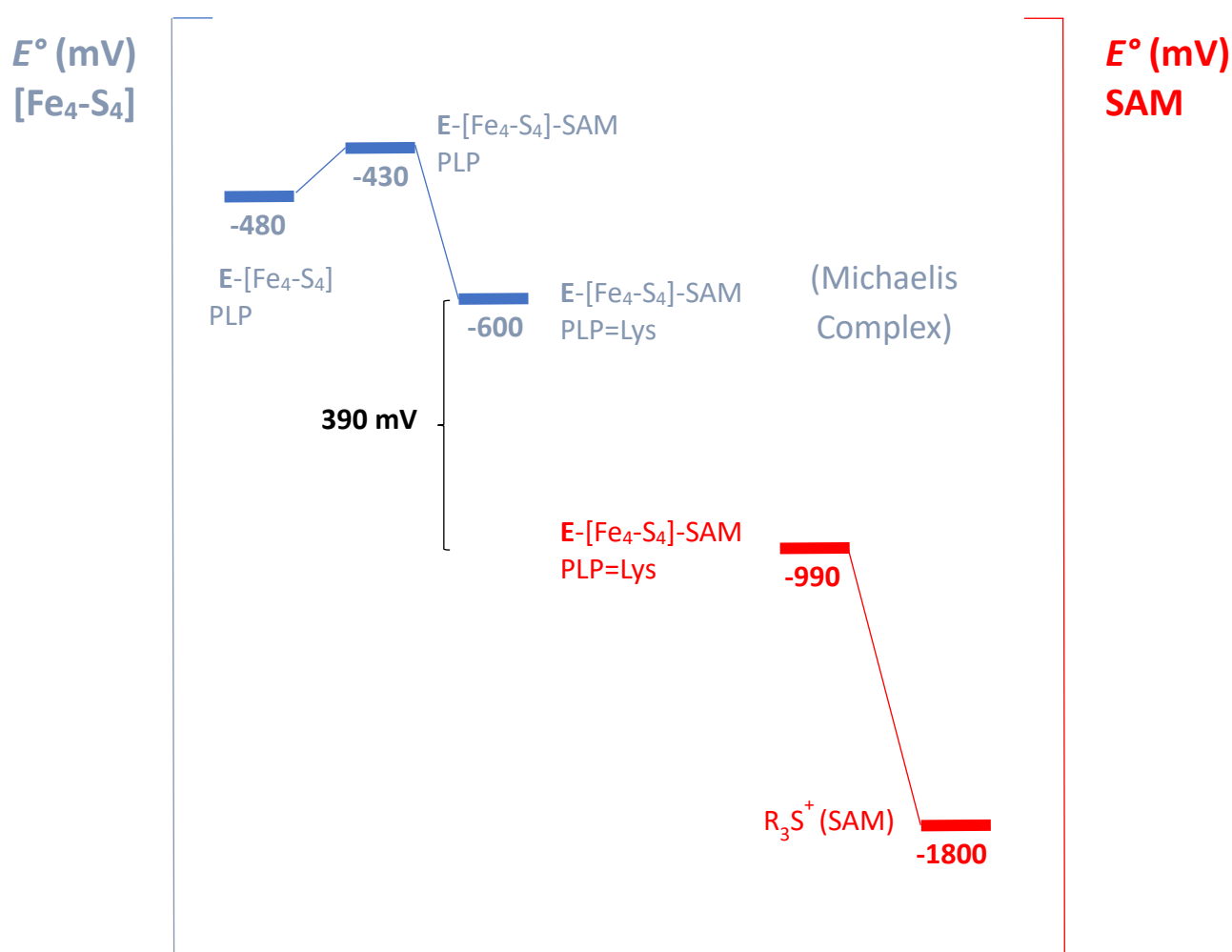
Figure 10. Energetics of one-electron reversible cleavage of SAM at the active site of LAM. The blue scale (left) illustrates the midpoint reduction potentials of the [Fe$_4$-S$_4$]$^{2+/1+}$ cluster with cysteine as the ligand, with SAM as the ligand and with both lysine and SAM bound as the ligand. The red scale (right) shows the potentials for irreversible one-electron reduction of trialkylsulfonium ions, such as SAM, in solution and for reversible one-electron reductive cleavage of SAM bound to the [Fe$_4$-S$_4$] cluster in the active site of LAM. Image adapted from Wang *et al.*[94]

active site of the aldimine adduct PLP-lysine substrate then lowers the overall energy barrier for the reductive cleavage of SAM from 32 kcal mol^{-1} in solution to 9 kcal mol^{-1}, considering both the contributions.[94] The midpoint potential of the SAM alone elevates up to -990 mV (an increase of 810 mV) following binding to the FeS cluster. Upon SAM binding, the methionine moiety coordinates the unique Fe atom that was reported changing its coordination from penta- to a more favoured hexa-coordinated geometry, which also favours the electron transfer reaction. Such a ligation is believed to contribute the additional energy to increase the SAM redox potential in the LAM active site and thus, stabilise the radical active species dAdo•. A recent study reported additional support for the mechanism of reductive cleavage of SAM by calculation of the most likely reaction trajectory using the solved X-ray structure of the HydE enzyme with bound SAM.[95] The calculated barrier of 12.9 kcal mol^{-1} agreed well with the previous experimental estimation of 9 kcal mol^{-1}. In addition, in all the radical SAM enzymes structure analysed with bound SAM, the cofactor coordinated to the FeS cluster always in the same fashion. The controlled radical fragmentation in the LAM enzyme (and broadly, in the radical SAM family) efficiently avoids the reactive dAdo• species triggering harmful side reactions. This control of reaction outcome is achieved thanks to issues that are present in most SAM enzymes: the gap between the redox potential of the FeS cluster and the free SAM,[16] the active site being deeply buried in the TIM barrel that seals off the cluster from the solvent and the presence of second-sphere residues that position the SAM cofactor and the substrate correctly.

1.10   Isolated Bacillus subtilis lysine 2,3-aminomutase (*Bs*LAM)

The LAM enzyme has also been isolated from *Bacillus subtilis* and a number of structural and mechanistic features are observed in both LAM enzymes.[49] The main difference is that LAM from *C. subterminale* is extremely oxygen sensitive with purification steps that must be carried out under strictly anaerobic conditions to ensure the maximal activity. *B. subtilis* LAM

is instead remarkably stable in air and it can be expressed and purified under regular aerobic conditions without loss of activity of the $[Fe_4$-$S_4]$ cluster, unlike most other radical SAM enzymes.[85] Like the clostridial variant, the *B. subtilis* LAM was characterised as a tetrameric structure with a molecular mass of 216 kDa, which contains four PLP molecules and $[Fe_4$-$S_4]$ clusters per tetramer. The enzyme was found lacking zinc atoms and the four cysteine residues to which zinc was coordinated in the *Cs*LAM, which reveals a different way for the *Bs*LAM structure to hold its subunits together; unfortunately no more structural data are available at the present time as no crystal structures have been yet solved for *Bs*LAM. Unlike the *Cs*LAM enzyme, which displays full activity after reductive pre-incubation with dihydrolipoate or dithiothreitol (DTT) in the presence of Fe(II), the bacillary enzyme only requires the presence of L-cysteine in the pre-incubation without added Fe(II) to achieve maximum activity. The activity of *Bs*LAM is quite low, with its full activity recorded as ~1.7% that of the clostridial variant either when the standard pre-activation was used or when the enzyme was purified under anaerobic conditions. No higher activity was recorded. The very low detected accumulation of the $[Fe_4$-$S_4]^+$ species in the *B. subtilis* enzyme may be due to lesser stabilization by the protein environment of the reduced state relative to the 2+ state of the FeS cluster. A less stabilised FeS-cluster reduced state suggests a lower reduction potential for the $[Fe_4$-$S_4]^{2+/+}$ couple in the *B. subtilis* relative to the *Cs*LAM enzyme; this may explain why the activity of the oxygen-tolerant enzyme is reported to be so low in respect to that of the *C. subterminale* variant.[96] The reason why the *Bs*LAM gained interest within this project was because of its natural oxygen-tolerance that could shine a light on the molecular mechanisms of oxygen-tolerance in radical SAM enzymes, which could be tested by transferring key modifications to the oxygen-sensitive, but more active LAM. The two enzymes were found to have 60% sequence identity implying a good level of structural similarity and making the protein pair a good example where the air-tolerant behaviour in radical SAM enzymes could be investigated.

The sequence alignment in Figure 11 shows the characteristic CX_3CX_2C consensus motif for this subtype of radical SAM enzymes, which coordinates the [Fe_4-S_4] cluster (highlighted in yellow). This sequence is situated in the N-terminus between residues 125-132 for oxygen-sensitive *C. subterminale* LAM and residues 134-141 for oxygen-tolerant *B. subtilis* LAM. In red, Arg112, Tyr113 and Arg116 (and the corresponding Arg121, Tyr122 and Arg125 for *Bs*LAM) coordinate the pyridinium ring of PLP while Arg134 (*Bs*LAM Arg143) is involved in

```
sp|Q9XBQ8|KAMA_CLOSU     ---------MINRRYELFKDVSDADWNDWRWQVRNRIETVEELKKYIPLT  41
sp|O34676|KAMA_BACSU     MKNKWYKPKRHWKEIELWKDVPEEKWNDWLWQLTHTVRTLDDLKKVINLT  50
                               . ::: .*.: .**** **: : :.::::*** : *:

sp|Q9XBQ8|KAMA_CLOSU     KEEEEGVAQCVKSLRMAITPYYLSLIDPNDPNDPVRKQAIPTALELNKAA  91
sp|O34676|KAMA_BACSU     EDEEEGVRISTKTIPLNITPYYASLMDPDNPRCPVRMQSVPLSEEMHKTK 100
                          :* ***    .:: : ***** **:* :. . *:* *::*   *::::

sp|Q9XBQ8|KAMA_CLOSU     ADLEDPLHEDTDSPVPGLTHRYPDRVLLLITDMCSMYCRHCTRRRFAGQS 141
sp|O34676|KAMA_BACSU     YDLEDPLHEDEDSPVPGLTHRYPDRVLFLVTNQCSMYCRYCTRRRFSGQI 150
                          *  ****** ******************:*:*: ******:******:*.

sp|Q9XBQ8|KAMA_CLOSU     DDSMPMERIDKAIDYIRNTPQVRDVLLSGGDALLVSDETLEYIIAKLREI 191
sp|O34676|KAMA_BACSU     GMGVPKKQLDAAIAYIRETPEIRDCLISGGDGLLINDQILEYILKELRSI 200
                         . .:  .::* ** ** :**::** *:****.**:.*: **  : .** *

sp|Q9XBQ8|KAMA_CLOSU     PHVEIVRIGSRTPVVLPQRITPELVNMLKKYHPVWLNTHFNHPNEITEES 241
sp|O34676|KAMA_BACSU     PHLEVIRIGTRAPVVFPQRITDHLCEILKKYHPVWLNTHFNTSIEMTEES 250
                         **:*::***:*.***:***** .*  .:****** :******* . *:* *:

sp|Q9XBQ8|KAMA_CLOSU     TRACQLLADAGVPLGNQSVLLRGVNDCVHVMKELVNKLVKIRVRPYYIYQ 291
sp|O34676|KAMA_BACSU     VEACEKLVNAGVPVGNQAVVLAGINDSVPIMKKLMHDLVKIRVRPYYIYQ 300
                          .**: *.:**:*:***:*:* *:**.: :**.*::.** :*********

sp|Q9XBQ8|KAMA_CLOSU     CDLSLGLEHFRTPVSKGIEIIEGLRGHTSGYCVPTFVVDAPGGGGKTPVM 341
sp|O34676|KAMA_BACSU     CDLSEGIGHFRAPVSKGLEIIEGLRGHTSGYAVPTFVVDAPGGGGKIALQ 350
                         **** *: ***:**:**:************.************** .:

sp|Q9XBQ8|KAMA_CLOSU     PNYVISQSHDKVILRNFEGVITTYSEPINYTP---GCNCDVCTGKKK--V 391
sp|O34676|KAMA_BACSU     PNYVLSQSPDKVILRNFEGVITSYPEPENYIPNQADAYFESVFPETADKK 400
                         *:**:**. .:*:*********:* ** .*           :     :.

sp|Q9XBQ8|KAMA_CLOSU     HKVGVAGLLNGEGMALEPVGLERNKRHVQE------------------- 421
sp|O34676|KAMA_BACSU     EPIGLSAIFADKEVSFTPENVDRIKRREAYIANPEHETLKDRREKRDQLK 450
                         . *: :  .  :: *  : * :*

sp|Q9XBQ8|KAMA_CLOSU     ---------------------
sp|O34676|KAMA_BACSU     EKKFLAQQKKQKETECGGDSS 471
```

Figure 11. The primary sequence of clostridial (Q9XBQ8) and bacillary (O34676) LAM aligned as proposed by Clustal Omega. The symbols below the alignment give information on the similarity of the residues: where " * " means identical residue, " : " means conserved substitutions and " . " means semi-conserved substitution (similar structure/functionality). Coloured residues are: Yellow, CX_3CX_2C [Fe_4-S_4]-binding motif; Red, residues involved in PLP coordination; Bright green, residues of β4 and β5 sheets of the TIM-barrel involved in orienting the cofactor SAM; Pink, alternate PLP binding residues in absence of substrate; Turquoise, tetramer-forming, Zn-binding residues (*Cs*LAM); Teal, C-terminal helical extension (*Bs*LAM).

PLP-carboxylate group binding. In bright green are highlighted the β4 and β5 sheets of the TIM-barrel, which contain those residues that are involved in correctly orientating the SAM molecule in front of the substrate to better drive the radical reaction. In pink, Lys337 (*Bs*LAM Lys345) resides in the PGGGGK motif and is responsible for PLP-binding when the substrate is not present in the active site. In turquoise, Cys268, Cys375, Cys377 and Cys380 in *Cs*LAM are involved in the Zn^{2+} binding to hold the four monomers together constituting the tetrameric protein. The *Bs*LAM lacks a Zn^{2+} binding cysteine motif and Chen *et al.*,[49] could not detect any zinc present in the purified *Bs*LAM thus suggesting a different way to hold its final structure together. Bacillary LAM possesses a long C-terminus helix that is not present in the *Cs*LAM enzyme, thus suggesting an alternative to hold the monomers together.

Despite the lack of *Bs*LAM crystal structure, the similarity of the catalytic pocket and SAM orientation was evidenced in the study by Mei-Chin[97] that reported the comparison between the *Cs*LAM crystal structure and the homology model of a *Bs*LAM strain (Bacillus subtilis subsp. subtilis str. 168). The *Bs*LAM structure was modelled by using the Swiss-Model server (http://swissmodel.expasy.org)[98] and the substrate and cofactors were docked in their *Bs*LAM model. They reported the arrangement of SAM and the PLP-Lys substrate adduct to be similar in both enzymes. Although they modelled the *Bs*LAM structure, they could not retrieve the whole enzyme folded chain because the server used the *Cs*LAM as a template, but was unable to model the C-terminus extension that is typical of this *Bs*LAM because of a lack of templates. Despite the lack of validation of their model their study gave support to the similarity existing between the two enzyme structures and the cofactor orientations in the active site. That encouraged us in studying the unknown structure and mechanism of the oxygen-tolerant LAM relative to the crystal structure of the oxygen-sensitive lysine 2,3-aminomutase.

1.11   Computational approaches to solve the oxygen tolerance challenge

Molecular dynamics simulations and other computational approaches are now widely used to attempt to link together the complex structures and properties of living systems to the underpinning laws of chemistry and physics that drive the conformational changes of enzyme structure impacting enzyme catalysis.[99] In the last 40 years, biological systems of relevant scientific interest that have their origin in bacterial and archaeal cells have been investigated by applying either classical or quantum mechanics to elucidate most of their properties. Through microscopic simulations is now possible to elucidate macroscopic properties of biomolecules that could be linked with the distribution and motion of the atoms constituting the biological system in a time-dependent process. As we wanted to investigate the correlation between two LAM enzyme structures and their different potential mechanisms for dealing with oxygen, the best approach to adopt was a computational study that relied on the molecular dynamics simulation method. In this way it is possible to achieve information about the dynamics and structure-related features of the LAM enzymes with no usage of classical laboratory time and equipment and without the challenges of anaerobic working. With the ever-increasing development of refined force-field potentials, along with advanced computer technologies, MD simulations can create the basis for a deeper comprehension of biological molecules. This methods can also be used to help with the interpretation of experiments concerning the biomolecule of interest and its properties, providing contribution to the knowledge of structure and functions of living systems. MD simulations have also been used to help with the interpretation of experiments concerning the dynamics of a biomolecule, through the calculation of the possible ways of motion in each particle composing the biological system. Information that arises from such calculations provides possible explanations or predictions about equilibrium averages, motional properties, thermodynamics of the system and time-dependent processes. The main limitation of the method is that even though the results are increasingly closer to experimental results, data arising from the simulations are based on a mathematical model that has approximations, and

experimental work is still needed to validate the simulation method, to test the accuracy of the calculations and improve the methodology.

As radical SAM enzymes are highly sensitive to oxidative damage, their expression, purification and catalytic studies must be run under a strictly inert atmosphere with the need for special equipment in the laboratory to avoid [Fe$_4$-S$_4$] cluster degradation and allow the enzyme to be studied. The MD simulation method was thus chosen for this project to pre-screen molecular possibilities without these experimental challenges. The aim is to obtain a more detailed understanding of oxygen-tolerance by providing a hypothesis for this rare skill among radical SAM enzymes and to design modifications for the improved resistance that can later be tested experimentally. Particularly, this project concerns the elucidation of the exceptional resistance of *Bs*LAM to such an oxidation where much more time and resources are demanded in understanding its properties with a purely experimental approach. By choosing a computational protocol, this project sets out to compare the two lysine 2,3-aminomutases and their properties in order to reveal those differences that make the *Bs*LAM resistant to oxidative damage, and identify modifications conferring tolerance that can be transferred to *Cs*LAM through design of mutants. By adopting the molecular dynamics simulation method, it has been possible to increase the knowledge of the dynamics of the enzymes and the electronic structure of the [Fe$_4$-S$_4$] cubane cluster. From the analysis of both enzymes protein environment around the active site, we could identify a few amino acids that, theoretically, tune the redox potential of the [Fe$_4$-S$_4$] cluster due to the different electrostatic interactions established by the new residues sidechain. The computational approach was used to reveal LAM's specific features and differences so to engineer new enzymes carrying amino acid residues that should contribute to an increased oxygen tolerance. Experimental work is still necessary to validate the outcomes of the calculations, i.e. that the engineered enzyme(s) can actually work in presence of oxygen, with a significant reduction in lab resources compared with a full, *a priori* lab study.

1.12   Technical background

1.12.1 Molecular Dynamics (MD) simulations

Living systems can be rendered by images at the atomic level thanks to crystallography but a limited amount of information about their dynamics, such as conformational variation, ligand binding or protein folding can be retrieved. Although, the atomic structure can be used to study biological macromolecules with regard to their chemistry and physics because most of the properties that define these systems can be represented by classical mechanics. Molecular approaches that apply either classical or quantum mechanical equations representing the fundamental physical law of matter, can be employed to solve chemical issues involving structures and reactions of small molecules. The study of the dynamical behaviour in biological systems can be realised by computer simulation of classical molecular dynamics (MD). Alder and Wainwright[100] were pioneers in the field and they use to model molecules as hard spheres interacting like pool balls. Stillinger and Rahman[101] performed the first simulation of liquid water, whilst the first simulation of a protein, the bovine pancreatic trypsin inhibitor was carried out in the Karplus laboratory.[99] Simulations at that time were performed in *vacuo* but over the past 20 years computational approaches have been developed and dynamics studies of biological systems lying between the macroscopic and the microscopic scale, can now be performed.[102] Classical molecular dynamics solves Newton's equations of motion for a system of N atoms interacting according to a potential energy *U*:

$$m_i \frac{\partial^2_{r_i}}{\partial t^2} = -\frac{\partial U}{\partial r_i} = F_i \qquad \textbf{(4)}$$

where m_i is the mass and r_i is the position of the atom *i*. On the right side the negative gradient of the potential is denoted, i.e. the force F_i acting on the atom. The equation describes the derivative of the potential energy as the changes in position over time. Thus, provided a model of the potential energy U, the force field is given, and integration of equation (4) yields the trajectory of the system. The trajectory gives insights about the structural flexibility of the

system as it explores different accessible conformations of the phase space. These molecular dynamics methods recreates the dynamics of the systems, in contrast to X-ray crystal structures of proteins which provide an average atomic position of atoms within a structure, but with such atoms having fluid-like motions around the average. Such a static structure, however, can be used as the starting 3D representation of a protein, to perform protein dynamics and so to calculate the trajectories of the atoms composing the biomolecule, which represent the protein motions and fluctuations that plays an essential role in the biomolecule mechanism and activity. The Maxwell-Boltzmann distribution assigns the initial velocities for all atoms at the desired system temperature. The choice of the integration time step Δt is crucial for the quality of the simulation because the accuracy increases with shorter time steps but at computer time expenses. The C–H bond stretching mode accounts for $\sim 10^{14}$ Hz the highest vibrational frequency in biomolecular systems. Splitting this period into 10 segments an integration time step of 1 fs is considered sufficient.

1.12.2 Force Fields and Functional Form

The functional form of the empirical potential energy that describes the interactions between the atoms, and the parameters used in that function constitute a force field.[103,104] Force fields are parameterised by comparing it to a combination of experimental and *ab initio* data, thus a force field should be able to reproduce properties of a molecule. Properties include equilibrium geometries, vibrational frequencies, dipole moments, molecular volume and thermodynamic properties like free energies of solvation. X-ray crystal structures or electron diffraction typically provide data for equilibrium geometries, whilst vibrational frequencies from infrared (IR) and Raman spectroscopy can be used to regulate force constants for bonded interactions. In case reliable experimental data are missing, like for many organic compounds, they can be obtained from quantum mechanical calculations.

A large variety of force fields is provided for different applications. The most popular force fields adopted for MD simulations are CHARMM,[105] AMBER,[106] GROMOS96[107,108] and OPLS-

AA.[109] They present differences in their parameters but the functional form of these force fields is very similar and here explained. Classical force fields use empirical functions to describe the atomic interactions. Atoms are considered as spherical particles, connected through covalent bonds forming molecules and each atom experiences a force resulting from its interaction with the rest of the system. The total energy expression, U_{total}, includes contributions from bonded and non-bonded interactions. The bonded interaction contains terms for bond stretching U_{bond}, bond angle bending U_{angle}, bond torsion U_{dihed}, and out-of-plane distortions (improper torsions) U_{impr}. Non-bonded interactions are represented by the van der Waals term U_{vdW} and the Coulomb term $U_{Coulomb}$. The final potential energy encompasses all these contributions and is expressed as follows:

$$U_{total} = U_{bond} + U_{angle} + U_{dihed} + U_{impr} + U_{vdW} + U_{Coulomb}$$

Bond, angle, and improper energies are approximated in terms of harmonic potentials with respect to bond distance r_i, bond angle Θ_i and improper dihedral angle ω_i:

$$U_{bond} = \sum_{bonds\ i} k_i^r \, (r_i - r_{0i})^2 \tag{5}$$

$$U_{angle} = \sum_{angles\ i} k_i^\theta \, (\Theta_i - \Theta_{0i})^2 \tag{6}$$

$$U_{impr} = \sum_{imprps\ i} k_i^\omega \, (\omega_i - \omega_{0i})^2 \tag{7}$$

while k_i^r, k_i^θ and k_i^ω denote the force constants. The parameters r_{0i}, Θ_{0i} and ω_{0i} mark the distance and angles at the energy minimum. Bond torsions are characterized by periodic potentials instead:

$$U_{dihed} = \sum_{diheds\ i} \frac{k_i^\Phi}{2} (1 + \cos(n\Phi_i + \delta_i)) \tag{8}$$

with the dihedral angle Φ_i, the barrier height k_i^Φ and a reference angle δ_i at which the potential is at the maximum value. The periodicity n_i counts the number of minima for a full rotation of 360°. In order to account for more complex rotational potential forms it is possible to employ a superposition of multiple sinusoid potentials like Eq. (8) with different periodicities, which add up to a Fourier expansion of the potential. The van der Waals (vdW) interaction energy is constituted of the attractive dispersion energy due to induced dipole interactions between two or more atoms and a repulsive term due to overlapping electron clouds. These phenomena are taken into account by two terms in the Lennard-Jones potential:

$$U_{\text{VDW}} = \sum_{\text{atoms } i,j} \varepsilon_{ij} \left(\frac{R_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}}{r_{ij}}\right)^{6} \tag{9}$$

that are dependent of the atom distance r_{ij}. The attractive term is proportional to r^{-12} whilst the repulsive one scales with r^{-6}; as a consequence there is an optimal distance $R_{ij} = (R_i + R_j)/2$ that corresponds to a minimum of depth $\varepsilon_{ij} = \sqrt{\varepsilon_i \cdot \varepsilon_j}$ in the potential but at a closer distance, repulsion term dominates. The second non-bonded term describes the electrostatic interaction of two point charges q_i and q_j:

$$U_{\text{Coulomb}} = \sum_{\text{atoms } i,j} \frac{q_i q_j}{4\pi\varepsilon\varepsilon_0 r_{ij}} \tag{10}$$

where ε denotes the relative dielectric constant, ε_0 is the absolute dielectric permittivity, and r_{ij} is the distance between the charges. The atom charges in this expression are an artificial construct since atoms and molecules are not charged unless they are ions, but atoms in molecules share their valence electrons and the electron density can be moved due to different electronegativity of the atoms involved. These different charge densities can be delineated into partial point charges. It is important to know that van der Waals and Coulomb interactions are usually excluded from computation whether the two atoms were separated by less than three bonds, because these interactions are modelled by the dihedral potential.

1.12.3 Homology modelling and model validation

In order to perform a computational study and comparison of biomolecules, reasonably good tertiary structures are needed so that behaviour could be predicted under the conditions of choice. Molecular structures can be determined experimentally by techniques such as X-ray crystallography and NMR spectroscopy and are accessible through "Protein Data Base" (PDB) files, where available and published.[110] For proteins without a 3D model available (*e.g.* at the PDB), reliable protein structures can sometimes be obtained through homology or comparative modelling. Homology modelling is based on the general hypothesis that proteins sharing a high degree of similarity in their amino acid sequences also share similarity in their secondary and tertiary structures. By comparing the sequence of a protein without a structure available to the sequence of a protein with known structures, a sequence alignment can be built where similar sequences are overlaid. These alignments can also reveal functional relationships among the proteins. Among homologues proteins, structures are more conserved than sequences.[111] On the other hand, proteins with a lower sequence identity generally produce poorer quality models, as there are larger deviations between those regions that can be more accurately modelled and the areas of uncertainty. In addition, structural features such as loop conformations are notoriously difficult to predict because they might differ between template and target even with many amino acids in common[112] and may lead to regions of higher error.

The root-mean-square deviation (RMSD) provides information about the "structural distance" of the homology model from the template by measuring the difference between Cα atom pairs present in the template and in the model structures when they are overlaid to each other. It is defined by the formula:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \qquad \textbf{(11)}$$

Where δ is the distance between *N* pairs of equivalent Cα atoms. For two proteins with 70% sequence identity between their amino acid sequences, the RMSD between Cα atoms of a model and the template is likely to be around 1-2 Å, but it can reach 4 Å at 25% sequence identity. RMSD values were used in the early stage of the homology model validation procedure to make a first estimation of the derived model and its quality. It was reported from the literature that at sequence alignments of least 25% for short proteins and 40% for long alignments is likely to provide a reliable homology model as it was demonstrated the structural relatedness between template and target structures at those levels of alignment.[113] Several software including exist to find the best protein template including BLAST,[114] ROBETTA,[115] I-TASSER[116] and SWISS-MODEL[117] that can also be used to generate the homology model structure of the molecule of interest.

The quality of the homology models can be assessed by measuring the distribution of the φ and ψ torsion angles of each residue composing the biomolecular system and represented in the Ramachandran plot.[118] The Ramachandran plot[118] is a way of representing the energetically allowed regions of torsion angles φ and ψ of the amino acids composing a peptide. The torsion angles are defined by $C_{(i-1)}$, $N_{(i)}$, $C\alpha_{(i)}$, $C_{(i)}$ atoms in phi and by $N_{(i)}$, $C\alpha_{(i)}$, $C_{(i)}$, $N_{(i+1)}$ atoms in psi. The Ramachandran plot graphically represents the dihedral angles population and the possible combination of angles allowed by the structure that contains them. Therefore, dihedral angles define the geometry of each residue relative to its two adjacent residues attached by positioning its planar peptide bond relative to the two adjacent planar peptide bonds, so to establish the conformation of every residues and the whole peptide. Because of the steric hindrance between amino acids, many angle combinations are not possible neither they are many conformations of residues. By plotting the angle combinations of every residues, it is possible to determine the torsional angles permitted as well as insight into the protein structure.[118] Although, it was reported from the literature that the areas where dihedral angles fall into the good orientation, do not take into account of small areas that are not normally considered from this analysis but several strains were found

to populate normally not allowed areas of the plot.[119] In addition, the analysis of torsional angles is based only on two features of each amino acid residue and therefore it provides a partial picture of the conformational richness of the protein.[120]

1.12.4 Tunnel investigation in protein structures

Pathways present in biological systems connecting the interior of the structure with the surrounding environment are called tunnels and they form the areas to access macromolecules. Tunnels in proteins have many purposes, they can serve as sites for binding or piloting ions, small molecules, peptides, nucleic acids and water. The substrate specificity in enzymes can be regulated by interactions of the molecule at the binding site, and also by the selectivity towards the substrate of those pathways leading to the active site. Enzymes selectivity is then considered a function of the pathway geometry and chemical properties related to the interaction with the substrate. In biology, chemistry and rational drug design a crucial matter is the prediction and elucidation of channels in macromolecules. The main intent is to identify all possible channels in the structure, their dimensions (length and wideness), the narrowest radius (denominated bottleneck radius) and screening residues forming the pathways, considering the global geometry of the system. Pathways in proteins can thus impact the molecule reactivity as they control the nature and magnitude of the interaction that the protein establishes with the substrate.

Transmembrane proteins involved with ions and small molecules transport and signal transduction, peptide way out channels that ribosomes employ to release newly synthesised proteins are a few examples of systems whose function relies on channels. The comprehension of the channels location within the system structure and their characteristics assist in tackling issues in rational drug design, protein engineering and enzymology among the others. Over the years, a few algorithms and software tools have been developed for the identification and characterisation of tunnel networks. Software like HOLE,[121] MOLE,[122] CAVER[123] and MolAxis[124] are among the most popular software employed for the tunnel

search in protein structures. They differ from their ability to look just for pores on the systems surface or for channels in the whole biological structure, from the accuracy in identifying channels emanating from protein cavities or from a user-defined starting point inside (e.g. the active site). These methods make use of mathematical algorithms to scan the surface of the protein and investigate the tunnels formed in the structure and the amino acid residues composing tunnels. Their accuracy and performance depend by the algorithm employed.

1.12.5 Quantum chemistry

Quantum chemistry is a branch of chemistry that applies quantum mechanical principles and equations to study atoms and molecules in order to deeply understand matter at its most fundamental level.[125] Quantum chemical calculations make it possible to accurately predict the structure of molecules and their electrostatic properties. The electronic structure of molecules is the state of motion of electrons in the electrostatic field created by stationary nuclei. The electronic structure is represented by either the wave function and the energy associated to each electron composing the molecule. By solving quantum chemical calculations it is possible to obtain an approximation of the wave function that, can be used to calculate other properties of the system of interest such as potential energy, spectroscopic properties and redox potential.

In quantum mechanics, particles have wavelike properties and the Schrödinger equation, a particular wave equation, controls how these waves behave. From classical dynamics it is known that the energy of a classical, nonrelativistic particle with momentum $\vec{p}$ that is subject to a conservative force derived from a potential $V(\vec{r})$ is:

$$E = \frac{\vec{p}^2}{2m} + V(\vec{r}) \tag{12}$$

For simplicity, a constant potential $V(\vec{r}) = V_0 = const$ is considered. That is the force free case, whose dispersion relation between ω and $\vec{k}$ for waves describing the particle motion should

be:

$$\hbar\omega = \frac{\hbar^2 \vec{k}^2}{2m} + V_0 \qquad \textbf{(13)}$$

This represents the dispersion relation for waves in one dimension, where the wavenumber $k(\omega)$, was a function of frequency. For waves in three dimensions the frequency of the wave is a function of the three components of the wave vector. Each wave with a given wave vector $\vec{k}$ has the time dependence as described:

$$e^{j(\vec{k}\vec{r}-\omega t)}, \text{with} \ \ \omega = \frac{\hbar\vec{k}^2}{2m} + \frac{V_0}{\hbar} \qquad \textbf{(14)}$$

That represents an electromagnetic wave with phase fronts traveling to the right. This notation conforms with the physics oriented literature. The superposition of such waves in $\vec{k}$–space allows to construct wave packets in real space:

$$\psi(\vec{r},t) = \int \phi_\omega(\vec{k},\omega) \ e^{j(\vec{k}\vec{r}-\omega t)} d^3k \ d\omega \qquad \textbf{(15)}$$

The inverse transform of the above expression leads to:

$$\phi_\omega(\vec{k},\omega) = \frac{1}{(2\pi)^4} \int \psi(\vec{r},t) \ e^{-j(\vec{k}\vec{r}-\omega t)} d^3k \ dt \qquad \textbf{(16)}$$

with

$$\phi_\omega(\vec{k},\omega) = \phi(k) \ \delta\left(\omega - \frac{\hbar\vec{k}^2}{2m} - \frac{V_0}{\hbar}\right) \qquad \textbf{(17)}$$

Another way to rewrite the wave function in Eq. (15) is by carrying out the frequency integration over ω:

$$\psi(\vec{r}, t) = \int \phi(k) \, exp\left(j\left[\vec{k} \cdot \vec{r} - \left(\frac{\hbar \vec{k}^2}{2m} + \frac{V_0}{\hbar} \right) t \right] \right) d^3k \tag{18}$$

Due to the Fourier relationship between the wave function in space and time coordinates, and the wave function in wave vector and frequency coordinates, is given:

$$\phi_\omega \left(\vec{k}, \omega \right) \ \leftrightarrow \ \psi \left(\vec{r}, t \right) \tag{19}$$

Then there are:

$$\omega \, \phi_\omega \left(\vec{k}, \omega \right) \ \leftrightarrow \ j \frac{\partial \psi(\vec{r}, t)}{\partial t} \tag{20}$$

$$\vec{k} \, \phi_\omega \left(k, \omega \right) \ \leftrightarrow \ -j\nabla \, \psi(\vec{r}, t) \tag{21}$$

$$\vec{k}^2 \, \phi_\omega \left(k, \omega \right) \ \leftrightarrow \ -\Delta \, \psi \left(\vec{r}, t \right) \tag{22}$$

where is obtained:

$$\nabla = \ \vec{e}_x \frac{\partial}{\partial x} + \vec{e}_y \frac{\partial}{\partial y} + \vec{e}_z \frac{\partial}{\partial z} \tag{23}$$

$$\Delta = \nabla \cdot \nabla \equiv \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{24}$$

From the dispersion relation, it follows by multiplication with the wave function in the wave vector and frequency domain:

$$\hbar\omega \, \phi_\omega(k, \omega) \ = \ \frac{\hbar^2 k^2}{2m} \phi_\omega(k, \omega) + V_0 \phi_\omega(k, \omega) \tag{25}$$

The corresponding equation in space and time domain with the inverse transformation is:

$$j \hbar \frac{\partial \psi(\vec{r}, t)}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi(\vec{r}, t) + V_0 \psi(\vec{r}, t) \qquad \textbf{(26)}$$

The generalization of Eq.(26) for a constant potential to the instance of an arbitrary potential in space leads finally to the Schrödinger equation:

$$j \hbar \frac{\partial \psi(\vec{r}, t)}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi(\vec{r}, t) + V(\vec{r}) \psi(\vec{r}, t) \qquad \textbf{(27)}$$

The Schrödinger equation can not be derived from classical mechanics, but classical mechanics can be re-derived from the Schrödinger Equation in some limit. The success of Eq.(27) in describing the experimentally observed quantum mechanical phenomena correctly justifies this equation. Initially the square magnitude of the wave function $|\Psi(\vec{r}, t)|^2$ was explained as the particle density. The $|\Psi(\vec{r}, t)|^2 \, dV$ is the probability to find a particle in the volume dV at position $\vec{r}$, if a measurement of the position of the particle is carried out at time t. The term $\Psi(\vec{r}, t)$ itself is then considered to be the probability amplitude to find the particle at position $\vec{r}$ at time t. In contrast to classical mechanics where the state of a particle is precisely described by its position and momentum, in quantum theory the full information about a particle is represented by its wave function $\Psi(\vec{r}, t)$. The wave function $\Psi(\vec{r}, t)$ enables to calculate the outcome of a measurement of any possible observable related to the particle, like its position or momentum.

The real wave function of a system is far too complex to be found directly but it can be approximated by simpler wave functions. The self-consistent field method[126] is an iterative method that approximates the Schrödinger equation to retrieve a more accurate set of orbitals that are used to solve the Schrödinger equation again until the results converge. To solve the Schrödinger equation, or its approximation, several software have been developed over the years to calculate the wave function of an atomic or molecular system: Gaussian,[127] Jaguar,[128] ORCA[129] are among the most popular computational software. They make use of methods of different accuracy to best represent the electronic structure and a basis set that

represents a set of functions (basis functions) to best describe the electronic wave function of the system. From the total energy of a system, the potential energy surface is created where local energy minima are found as a function of the coordinates of all nuclei. The derivative of the energy with respect to all displacements of the nuclei is zero but a local minimum is a stationary point that leads to an increase in energy. The local minimum that describes the global minimum corresponds to the most stable geometry; this process of calculating stationary points is called geometry optimisation.

1.13 Aims of this work

The goal of this project was to develop a detailed understanding of the mechanism of oxygen tolerance in radical SAM enzymes, with a view to translating this knowledge in future to members of the superfamily that show biotechnological promise. Only two examples of naturally oxygen-resistant radical SAM enzymes had appeared in the literature, and to conduct the computational study, the two LAM enzymes were selected as the oxygen-sensitive variant was the only one for which the crystal structure had been solved at the commencement of the project. The key-milestone would be to design a mutant form of the clostridial enzyme lysine 2,3-aminomutase that could resist to the oxidative stress in presence of air.

The main questions addressed by this project are:

I. What is the origin of oxygen stability of the few members of the radical SAM enzymes, and what might be their differences in secondary structure and three-dimensional spatial conformation of the environment surrounding the [Fe$_4$-S$_4$] cluster in their active site? To start the computational investigation, a satisfactory structure of the unknown *Bs*LAM enzyme must be developed and the whole sequence of the enzyme needs to be modelled to have a realistic structure to work on. Validation of the model must follow to test its reliability in an MD simulations protocol to ensure meaningful results. These aspects are addressed in Chapter Three.

II. Is the tolerance to oxygen of bacillary LAM originating from particular bulky/scavenger residues that reside in close proximity of the O$_2$-sensitive to inhibition [Fe$_4$-S$_4$] cluster inside the protein active site that either obstruct the diffusion of oxygen or interact with it during its transport to the interior of the enzyme? The presence of bulky residues in *Bs*LAM that acts as either air-flow obstruction or sulfur-containing residues that can trap oxygen molecules could explain the unusual tolerance to oxygen. MD simulations on both the enzymes were carried out to look for differences in the residues close to the active site and are discussed in detail in Chapter Four, section one (4.1).

65

III. Is it possible to derive information about the $[Fe_4-S_4]$ cluster redox potential and the protein environment inside these metalloenzymes using computational techniques and predict *in silico* which mutant effects backbone modifications that will result in novel air-stable mutants? QM calculations were used to measure the relative midpoint potential of the $[Fe_4-S_4]$ cluster in the two LAM enzymes, including selected residues in the active site environment that will likely tune the redox potential. The results are discussed in Chapter Four section two (4.2).

IV. Can the oxygen-tolerant behaviour, be transferred to other SAM radical enzymes through site-directed mutagenesis, on inspiration of a previous work where it was possible to create an enzyme mutant with such a useful characteristic? The putative amino acid residues that impact the oxygen tolerance of *Bs*LAM can guide the transformation of the *Cs*LAM structure by mutagenesis. Subsequent MD simulations are able to be used, alongside other structural analyses to estimate which changes impact oxygen-sensitivity positively. These approaches are described in Chapter Four, section three (4.3).

2 Methodology section

This project focuses on atomistic molecular dynamics (MD) simulations using AMBER[130] and quantum mechanics (QM) calculations using Gaussian 09.[131] Molecular dynamics simulations of proteins are able to provide information about the enzyme dynamics and loop motions in solvent over time and under known conditions of temperature and pressure. These simulations are able to signpost the effects of intrinsic (proteins chemical structure) and external (solvent interaction) factors on protein conformations, loop movements, diffusion of molecular oxygen inside the protein and the protein environment surrounding the active site. Over the nanosecond time scale of MD simulations, a dynamic picture emerges of proteins in water, where structural changes can be visualised and their movement and behaviour in solution can be compared. The resulting data also provides information about the physical properties of biological systems, including structural movements and cavity formation within the structure. These motions are governed by the calculated atom trajectories and so provide an approximation of an enzyme's behaviour. QM calculations were used to obtain optimised structures of the cofactors such as the [Fe_4-S_4] cluster and the SAM molecule to take as starting-points for simulating the realistic behaviours of these cofactors inside the active site in absence of the substrate.

In order to perform a molecular dynamics simulation of a biomolecule, reliable starting structures are needed for LAM. The structure must be a three-dimensional atomic-resolution model of the protein revealing the arrangements of all the atoms it contains and is typically determined by techniques such as X-ray crystallography and NMR spectroscopy. If there is no solved crystal structure for the protein of interest – as in the case of *Bs*LAM -, another way to generate a model structure is to use the homology or comparative modelling. Homology modelling was therefore employed to create a model of the *Bs*LAM enzyme to initiate the investigation on the radical SAM enzymes lysine 2,3-aminomutases. After validation of the homology model, both the structures of the oxygen-sensitive *Cs*LAM and its *Bs*LAM variant

were subjected to MD simulations for at least 400 ns. The simulations were then analysed to establish differences in the dynamics and loop motions of the two enzymes in different conditions of cofactors present in the active site either with the [Fe$_4$-S$_4$] cluster or with the SAM molecule coordinated to the metal cluster. Identification of potential oxygen pathways formed during the MD simulations in the enzyme structures was carried out by the tunnel-search software CAVER[132] which provided information about the enzyme dynamics and loop fluctuations potentially related to the presence of different cofactors within the active site. The software also indicated the amino acid residues that were found along tunnels that could account for differences in size of these tunnels and either obstruct the flow or allow interaction of oxygen molecules with scavenger residues. The [Fe$_4$-S$_4$] cluster was then analysed using QM calculations to gain information about electrostatic structure, using the software Gaussian09. The electrostatic structure was studied in both the oxidation states of the metal centre to investigate the vertical electron affinity and the differences that could have arisen between the two LAM enzymes. Subsequently, the protein environment around the [Fe$_4$-S$_4$] cluster in the two enzymes was studied to identify redox-modulating residues that could have an impact on the electrostatic structure of the cluster and make the *Bs*LAM metal centre less susceptible to O$_2$ attack. A limited number of residues were identified in this regard, which allowed the creation of *Cs*LAM mutants carrying the mutation that may cause a modified midpoint potential in the enzymes. The dipole moments in all the LAM enzymes, obtained via MD simulations, were then calculated and compared to estimate which among the designed mutants could have similar electrostatic and redox potential to the oxygen-sensitive LAM enzyme and so to identify a potentially oxygen-tolerant *Cs*LAM mutant candidate.

2.1   Homology modelling of radical SAM enzymes

In the early stages of this project, the model of *Bs*LAM and the model used for method validation, the *Cc*THIC, were both created by homology modelling as no crystal structures of those proteins were available. Thus, structurally related crystal structures were used as templates to generate the model. The protein backbone and sidechains could be represented by a well-defined protein force field and the parameters for cluster and SAM molecule were taken from the literature.[77, 133] We needed, for *B. subtilis* LAM (UniProtKB: O34676, 471 amino acids), a high-quality model. The homology model was generated and its quality assessed to validate the model and ensure its reliability to be used in MD simulations.

The online protein structure and function prediction server I-TASSER (Iterative Threading ASSEmbly Refinement)[116] was chosen to predict the 3D-structure of the LAM and THIC enzymes. I-TASSER is a bioinformatic suite of programs for protein structure prediction. The webserver generates 3D atomic models using a combination of *ab initio* folding and threading methods (from the PDB databank), starting from the amino acid sequence of the biomolecule of interest. If there were possible threading unaligned regions due to missing templates, I-TASSER builds the target structure by using *ab initio* modelling that was reported generating good quality models of small proteins.[134] I-TASSER was one of the consistently top-ranked methods in CASP (Critical Assessment of Structure Prediction), a community-wide range experiment to find the best bioinformatic method for protein structure prediction.[135] I-TASSER generated five models with structures that corresponded to the largest structure clusters based on the pair-wise structure similarity and ranked by the C-score, which is calculated on the threading template alignments and the convergence parameters of the structure assembly simulation. Usually the first proposed model reports the highest C-score, hence the highest confidence. The I-TASSER modelling procedure starts from detecting and ranking structure templates, identified by a meta-server fold recognition approach containing multiple threading programs (LOMETS),[136] where each threading program can generate thousands of template alignments, and they are sorted by average performance in large-scale

benchmark test experiments. I-TASSER selects the template with the highest rank from each threading program.

I-TASSER ranks the so-predicted models by using the "C-score" which is a confidence score for estimating their quality. It is measured on the significance of the threading template alignments and the convergence parameters of the structure assembly simulations. The C-score usually is in the range of -5 to 2, where a C-score of higher value defines a model with a higher confidence and *vice-versa*.; a C-score > -1.5 indicates the model has a correct global topology. The TM-score is a scale I-TASSER employs for measuring the structural similarity between two structures; it is usually more reliable than RMSD because it is not sensitive to local errors. RMSD measures the average distance of all residue pairs in two structures, so a local error will raise the RMSD value even though the global topology is correct. TM-score is insensitive to the local modelling error because smaller distances are weighted more strongly than the larger distances. Thus a TM-score >0.5 indicates a model of correct topology whilst a TM-score <0.17 means a random similarity. The first ranked model is generally the highest quality one, reporting good values of both C- and TM-score but may possess a high RMSD that implies a bad model. As has been previously indicated,[137] this is due to the low correlation between RMSD and the other quality criteria employed by I-TASSER, describing the case of the good quality of the whole topology of the model and a large local modelling error that results in high values of RMSD.

A docking method was employed to fit the SAM molecule into the active site and orient that to face the [Fe_4–S_4] cluster and promote the accurate coordination. Four SAM molecules were placed into the tetrameric structure of *Bs*LAM using Maestro 11 visualisation package implemented in the simulation software Schrödinger-Maestro (v. 7.0)[138,139] and docked using the Glide software package.[140,141]

2.2   Molecular dynamics (MD) simulation protocol

The molecular dynamic trajectories were produced in this project by all-atom molecular dynamics (MD) simulations using the AMBER14 and 16 software,[142] which used the AMBER ff14SB force field[143] along with the generalised AMBER force field (GAFF)[144,145] for non-standard residues (Appendix 6.1). The force field parameters for SAM are based on electrostatic reparameterised force field parameters from Saez and Vöringher-Martinez.[146] The parameters for the $[Fe_4-S_4]$ cluster are based on the parametrization of biological relevant iron−sulfur clusters by Carvalho and Swart.[77] All simulations were carried out using either central processing unit (CPU) or graphics processing unit (GPU) acceleration[147] installed on the high performance computing (HPC) facilities of the University of Nottingham. Initial simulations (relating to the oxygen sensitive simulated systems) have been performed through the CPU acceleration, whilst the latter ones have been produced with the benefit of the faster calculations allowed by the graphic card installed on the HPC.

Every enzyme structure was embedded in a truncated octahedron water box using a 10 Å cut-off and solvated with simple point charge (SPC/E)[148] water model and then the whole system was neutralised with Cl^- or Na^+. All systems were simulated using periodic boundary conditions. The particle-mesh Ewald (PME)[149] algorithm was used to calculate long-range electrostatics and the other non-bonded interactions have been calculated with a cut-off of 12 Å to not slow down the calculation and still include significant inter atomic interactions. Equilibration and production steps were carried out at a constant temperature of 300 K and constant pressure of 1 atm. The temperature was adjusted using Langevin dynamics[150] with a collision frequency of 1 ps^{-1} and isotropic scaling was used to maintain the pressure with a relaxation time of 2 ps.

All-atom molecular dynamics (MD) simulations were performed on either the crystal structure of LAM (pdb: 2A5H), the homology model of *Bs*LAM, the engineered *Cs*LAM mutants (based on pdb: 2A5H) and the THIC enzyme structures used in the homology model validation (pdb: 3EPN). Each bio-system undergoing simulation (beside those involved in the validation

73

protocol) contained either the [Fe$_4$-S$_4$] cluster, coordinated to the cysteine-bridging residues in the catalytic pockets or the SAM molecule coordinated to the [Fe$_4$-S$_4$] cluster in each catalytic pocket (figure 12). Although LAM binds the cofactor PLP, this molecule was not included in the binding studies due to the fact that the project was not focused on the whole LAM catalysis mechanism but rather on the comparison between the two LAM enzymes to elucidate potential structural and dynamics differences. The apo-form of the enzyme was only used in the homology model simulation of the THIC crystal structure. The MatchMaker tool in UCSF Chimera[151] software allowed the assembly of the multimeric forms of the SAM enzymes employed (tetrameric in the case of LAMs and dimeric structure in the case of THIC) using the relative crystal structure as the reference structure.

All X-ray crystal structures that were retrieved from the PDB needed some modification of the files to make them suitable for MD simulations. These files contain coordinates of each atom composing the macromolecule, which represents the necessary part of the file to feed the MD software to start the simulation.

The H++ webserver 3.2 version[152,153] assuming a pH of 7.0 and dielectric constant of 10.0 and

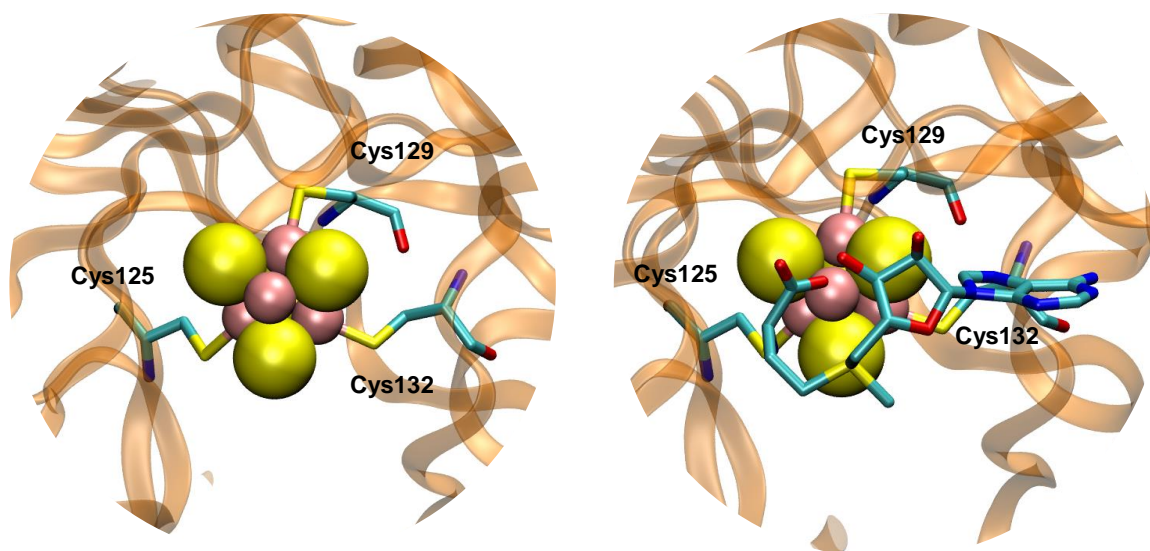**Figure 12. Representation of the active site chosen for each simulated bio-system. The selected enzymes were put into the MD protocol carrying just the [Fe$_4$-S$_4$] cluster (left) or SAM coordinated to the metal cluster (right). The FeS cluster is depicted in *Van der Waals* and the SAM molecule is in *ball and stick* representation. Cysteine ligands are shown with labels taken from *Cs*LAM.**

other defaults applied, was used to check the protonation states of those enzymatic residues that could exist with different protonation states. The server makes a screen of the submitted PDB file containing the structural information of the biomolecule then it calculates the local pKa based on neighbouring residues applying implicit water solvation. From that it can estimate the most likely protonation state at a certain pH for those specific residues.

There are amino acids that might exist in the protonated or un-protonated form whilst other residues can be protonated at different positions and it mainly depends on where in the protein they are found, either on the surface or internally in the structure. Acid residues are more likely to be un-protonated if they are on the protein surface and *vice versa* for basic residues; whilst the protonation state of histidine, as it contains two nitrogen atoms on its imidazole ring, is able to specify on the δ-, ε-nitrogen or both histidine ring positions. All these different forms are expressed with a different 3-letter code residue name in the PDB file just like any other residue names, so the specific protonation state is transferrable between software. In a PDB file, acronyms like GLH and ASH, denotes the charged form of glutamate and aspartate residues while HID, HIE and HIP indicate which N-atom(s) on the histidine ring harbours a proton, respectively the δ, ε or both the nitrogen positions. Hydrogens might be mistakenly added where they generate clashes, especially if a cofactor or other molecules is present in the PDB file, so the file is often stripped of hydrogen coordinates before feeding into *xleap*.

The *xleap* program is contained in the AMBER[130] package and reads in the coordinate information, topology and force field, outputting the starting files that will be used for the trajectory calculations. *Xleap* applies the selected force field to parameterise every atom and custom residues. The topology and coordinate files for each structural complex were generated using the *xleap* package. This initial file was used to run the two-step energy minimisation, using the steepest descent followed by conjugate gradient minimization where the whole system was relaxed to a lower energy *minimum*. The two minimisations were run using 10000 iterations on the first step with applied restraints to the protein and cofactor of

30 kcal/mol; the second step was used a maximum iteration number of 8000 and restraints were applied only to cofactor molecules. This step is fundamental as the initial geometry may not be at its lowest energy level and this initial step will take the system to an overall relative minimum level of energy to avoid steric clashes during the simulation. The newly optimised geometry was used in the subsequent heating step where the low energy system was slowly (1 ns simulation length) heated up from 0 to 300K in an NVT ensemble, at constant volume. An equilibration step at constant temperature and pressure (NPT ensemble, 2 nanoseconds long) was run, after which the MD production step and simulation was performed for at least 400 nanoseconds. Trajectory production was performed over multiple steps of 10 nanoseconds each until the overall simulation time scale of at least 400 ns was collected. We decided to run one long MD simulation each molecular structure because we have had to take into account the large amount of atoms composing the enzymes and the presence of homology modelled structures. In literature there is an ongoing debate if more short simulations are better than several long MD simulations to describe the dynamic behaviour of biological systems. Many and short MD simulations can be more accurate because it allows the sampling of more data and accurate results and are relatively less computationally expensive.[154] On the other hand, longer timescale MD simulations can capture biological processes, which majority of them happen on the nano- and micro-seconds scale.[155] In our case we decided to run prolonged MD simulations instead of multiple short simulations because homology models often needs longer equilibration and the behaviour of this equilibration can give evidence of good or bad homology models. Particularly, the C-terminal loop of the model structure that represents the area of the enzyme with minor accuracy might need more time to equilibrate. In this study the homology model was used to investigate dynamic water (and Oxygen) channels that also need longer sampling to describe changes over time. These factors affected the stabilisation of the molecular systems during performance of MD simulations and they might take longer to reach a low-energy structural conformation that could represent the enzymes as close to reality as possible. As we wanted to investigate oxygen channels

formation in our enzyme structures, longer MD timescale was also beneficial in the observation of such channels and their development over time to better understand their properties. Also, MD run performance time and CPU resource consumption that simulations used to require at the time this project started drove us in performing one long simulation. Nonetheless, we did recognise that more simulation runs to compare average each other, would have provided more accurate outcomes.

When the simulation had been performed for enough time without any visible instabilities in the system, all the trajectory files were collected and merged together, setting the offset for reading in trajectory files to 50 to generate 4000 snapshots of enzyme simulation. Finally, trajectories were reimaged and stripped of solvent and counterions to reduce the file to a reasonable byte size (for data storage reasons), using *cpptraj*,[156] then the final outcome visualised through VMD.[157] *Cpptraj* was used to calculate RMSD, RMSF, and the vector-dipole moment interaction in each simulation. RMSD measures the average distance between the atoms of the simulation and a reference structure (the backbone atoms and the PDB structure initially created by *xleap* were chosen unless otherwise reported). RMSF reports the atomic positional fluctuation of an atom compared to the first reference position (the first structure at the start of the simulation) or to the average structure in the MD over time. Backbone and alpha-carbon atoms were chosen to represent the fluctuation of each residue in RMSD and RMSF respectively. Low values reported from the measurement describe small fluctuations and ordered loop motions, with high values describing large fluctuation. All the tools and software employed are based on Unix shell or command-line interfaces (CLI).

2.3   Homology model validation

The Ramachandran plot is a widely employed technique for the checking of the stereochemical quality of enzyme structures and it was generated through the web server PROCHECK[158,159] on either crystal structures and homology models. Before studying the

dynamical differences of LAM enzymes, a different model validation technique that could assess the reliability of the model when submitted to the MD simulation protocol was necessary. In addition, because of the relatively short sequence identity of ~60% between the two LAM enzymes and the lack of other radical SAM enzymes structurally solved to use as templates, it was decided to use a further technique to employ in the homology model validation.

To better understand and compare the course of every simulation and identify the nature of the different molecular motions and conformational states of each enzyme and their monomers, principal component analysis (PCA)[160,161,162] was performed on all the produced molecular dynamics trajectories. The principal components analysis is a statistical technique used to find patterns in high-dimensional data and it is often used in data analysis to reveal the internal data structure to best explain the data variance. As trajectory files produced by MD simulations provide a substantial quantity of information and data, the dimensional space of the data needs to be reduced to easily identify a configurational space that contains a few degrees of freedom where anharmonic motion occurs. The analysis of the principal components extracts the significant dominant modes in the motion of the biomolecule from the trajectory, reducing the dimensionality and quantity of data, without losing important information so that key trends can be extracted.[163] The important movements or loop fluctuations of the enzyme dynamics can be represented by an ensemble of eigenvectors retrieved from the diagonalisation of the covariance matrix of the Cα-atoms to which eigenvalues are correlated. The analysis of such eigenvectors elucidates the internal movements of a biomolecule and their structural conformations that have been sampled during the MD simulation.

The Ramachandran plot is a widely used technique to investigate potential structural instabilities in proteins because of the simplification it offers in representing their complex structure. It has become a common practice to use the Ramachandran plot to observe the dihedral angles distribution of the model structure to validate. In many cases, the majority of

the residues that fall in the allowed region of the plot is enough to consider the model of a good quality. However, by reducing the representation of residue conformers to only two specific geometrical parameters, some of the conformational features that characterise each residue composing the structure is lost, furthermore, this analysis do not consider weak interactions. In the case of LAM enzymes though, the Ramachandran plot was used only to make an initial assessment of the generated models. Considering the relatively low sequence identity of our pair of LAM enzymes (~60%) the analysis of the torsional angles distribution could not be sufficient to validate the reliability of the models. Both the crystal structures used as templates were solved at a relatively low resolution (angstrom) that in the process of the homology model generation, means that structural imperfections could be transferred to the model.

Principal component analysis (PCA) was used to investigate the energetically favoured overall enzyme structure in each LAM simulation to observe how the presence of SAM in the active site impacts the dynamic conformations. PCA was also used in the homology modelling validation to quantify the difference of the dynamic fluctuation in the MD simulation of the crystal structure and the homology model of LAM-related SAM enzyme. If the simulations of the crystal structure and its homology model, built from different templates than the original structure, would converge, then the *Bs*LAM homology model can be considered reliable for the dynamics aspects of the MD simulations.[164]

PCAs were performed using the pyPcazip software package[165] and the principal component plots were obtained by using the web-based interactive computational environment Jupyter notebook.[166] PCA was performed for each MD trajectory over the whole length of the simulation (4000 frames) by selecting Cα-atoms of each residue present in the system to analyse the whole multimeric structure. Alpha-carbon atoms were employed to describe protein motion, as they are representative at the residue level.

MD simulations require high quality model inputs to generate reliable and meaningful results. The process of homology model validation makes checks to ensure the generated 3D

79

structure is well predicted, without inner steric clashes or bad torsion angles. In order to validate the quality of the generated *B. subtilis* LAM model and establish its reliability for MD simulation, another crystallographically-characterised radical SAM enzyme with similar structural properties to the lysine 2,3-aminomutase was identified from the protein data bank. The validation method aimed specifically to identify a pair of enzymes with both their crystal structures available, which shared a TIM-barrel conserved motif and structural identity of around 60%, just like the *Cs*LAM-*Bs*LAM pair. I-TASSER was then used to generate the homology model of the known enzyme by excluding the existing crystal structure in the template search, so that the second template (with 60% of sequence identity) would be used. Once the homology model was generated, it was used to perform MD simulation together with the crystal structure of the chosen enzyme thus to formally have two MD simulations of the same enzyme: the crystal structure and the homology model structure built up by using a different template that shared about 60% of sequence identity. The MD simulations were run for 400 ns and analysed to check if both the THIC enzymes had dynamically covered a similar conformational space that was detected by the PCA[149] and could account for similar explored conformational states.

If this was the case, this homology model approach could be considered reliable for further dynamics. The adopted strategy allows then for comparison of the template structure with that obtained crystallographically and identification of specific modelling artefacts that might arise and which might therefore be expected in the similarly modelled *Bs*LAM. A reliable model of another LAM-related enzyme could be considered as providing a good case that the *Bs*LAM model generated under similar conditions of sequence identity, and consistent protocols, should also be reliable.

This added validation is not a direct validation of the *Bs*LAM structure but an indirect assessment, by demonstrating that the homology of a structurally closely related enzyme is of good quality. This approach offers a wider view of the dynamical aspects of the model structure and it could give a more accurate estimation about the reliability of the homology

model structure when submitted to an MD simulations protocol. That could give more confidence on a structure modelled on the few radical SAM enzymes structure available.

2.4   Tunnel-cluster analysis

In order to identify potential tunnels that might be used by molecular oxygen to diffuse into the active site of lysine 2,3-aminomutase, the software CAVER 3.0[167] was used. This software makes use of a Voronoi diagram to represent the framework of tunnels inside a protein structure during a simulation. A Voronoi diagram partitions the simulation space into polygons that enable a localisation of channels and bottleneck residues. This algorithm also has the benefit of both reducing the error relative to the previously used the finite-grid approximation[132] and requiring less processor time. The calculation and clustering of pathways provides specific characteristics of each transport pathway and the evolution of these pathways to identify changes over the simulation time. Automatic clustering allows the analysis of channels identified throughout the simulation and not just one static structure, as has historically been the case.[168] Invisible pathways in enzyme structures can thus be dynamically investigated.

Tunnel-cluster analysis was performed on the solvent-stripped lysine 2,3-aminomutase MD simulations with a selection of 40 enzyme structures (1 structure extracted every 10 ns of the whole simulation), choosing the $[Fe_4\text{-}S_4]$ cluster contained in monomer A as the target of the calculation. Every tunnel identified was able to connect the buried FeS cluster to the enzyme surface, and they were compared to determine the accessibility of oxygen to the deep catalytic pocket of each different LAM.

Characteristics about the tunnel-clusters identified in the CAVER output included: the number of snapshots in which the cluster has been identified to see how frequently the tunnels that compose the cluster are spread over the MD snapshots and the average and maximum bottleneck radius, which represents the narrowest part of the recognised tunnel to

identify access to water molecules and the tunnel length. A priority value describes the relevance of a given tunnel-cluster and is calculated by averaging the sum of a tunnel's throughputs. The throughput in each tunnel-cluster is defined as the energy cost of each identified tunnel averaged over all the starting conformations. This average therefore represents the contribution of each tunnel to the cluster over all snapshots so that the given cluster of tunnels can be ranked through the cost function from the most energetically favourable to the least one. The cost function of a path is defined by Eq.28:

$$c(r, L) = \frac{L}{r^n} \tag{28}$$

The cost function is dependent on the ratio between the tunnel length and its radius so shorter and wider tunnels will be considered more likely to occur in the enzyme structure rather than long and narrow tunnels which they can be formed basically everywhere throughout the enzyme structure. A value of zero is adopted for those enzyme conformations where no tunnels of the mentioned cluster are found and the highest value is used in conformations that contain many tunnels with high throughput. These criteria means that frequently appearing, short, wide and energetically-convenient tunnels drive the rank of the clusters of tunnels more highly.

2.5 Quantum mechanical (QM) calculations and protein environmental effect on the FeS cluster

The starting atomic system was represented by those atoms composing the $[Fe_4\text{-}S_4](SCH_3)_3$ system and structures were taken from MD snapshots at 10, 70, 110, 150, 170, 280, 300, 320, 380 and 400 ns of the *Bs*LAM and snapshots at 10, 70, 150, 190, 210, 250, 290, 310, 350 and 400 ns of the *Cs*LAM. Subsequently, the $[Fe_4\text{-}S_4](SCH_3)_3$ systems were extracted from these snapshots by retrieving atom-coordinates of the FeS cluster and the three coordinated

cysteinyl ligands, including just the sulfur and Cβ atom, to represent the metal cluster-coordinating sulfur atoms and the covalently bound methyl moiety, respectively. The Gaussian09[169] package is a program for doing *ab initio* and semiempirical calculations and was used to run QM calculations of the sub-systems *in vacuo*. The a*b initio* method use calculations based purely on the principle of QM without experimental data, whilst semi-empirical calculations make use of parameters based on experimental data to correct the approximation of the QM principles.

The QM workflow was based on the initial geometry optimisation of the groups of atoms to achieve their lowest structural energy. At the first step the *ab initio* method was employed to start the optimisation on the MD-retrieved molecular structure and the optimised geometry was used for the subsequent calculations. The geometry optimisation of the oxidised state of the FeS cluster consisted of 3-steps: (1) A preliminary HF/STO-3G optimisation[170] (2) A further optimisation with the hybrid functional B3LYP employing the 6-311 +G(d) basis set[171,172,173] and (3) a final geometry optimisation using a custom hybrid functional at 5% Hartree-Fock and 95% Becke88 exchange,[174] and 100% Perdew86 correlation functional[175] with the 6-311 +G(d) basis set as per Harris and Szilagyj.[176] Convergence was confirmed at each step. The optimisation was initiated from a low level of theory, so the simple *ab initio* method was used and the Hartree-Fock (HF) correlation functional with the 6-311 +G(d) basis set was the starting of the protocol. The minimised structure was used as the starting point for the second optimisation that was performed using an improved accuracy DFT method like B3LYP hybrid functional with the 6-311 +G(d) basis set. The last optimisation was then run by adopting the custom hybrid density functional at 5% Hartree-Fock and 95% Becke88 exchange and 100% Perdew86 correlation functional with the 6-311 +G(d) basis set, as suggested by Harris and Szilagiy[176] to calculate the FeS cluster vertical electron affinity. The validation of this level of theory to use for FeS clusters is widely reported.[177,178] By using a protocol that goes from a less accurate to a higher theoretical method accuracy, some common issues related to achieving convergence are less likely. In addition, methods such as

HF are less time-demanding allowing an easier completion of the optimisation. The final geometry was eventually used to perform a single-point energy calculation, with the B(5%HF)P86 density functional, on both the oxidised and reduced state. The charge/multiplicity pair assumed for both CsLAM and BsLAM were -1/+1 in the oxidised state and -2/+2 in the reduced one as they represent the two states of the LAM FeS cubane cluster ([Fe_4-S_4]$^{2+}$ and [Fe_4-S_4]$^{1+}$). The final energies were used to calculate the vertical electron affinity (VEA) of the QM-optimised systems as it defines the energetic difference between the oxidised and reduced electronic state of the FeS cluster and were used to check if they were different between the two LAMs. The VEA, which is defined as an electronic structural property affected by electrostatic interactions, allows estimation of the relative redox potential and thus can identify changes in reactivity, and is defined by Eq.29:

$$VEA = E_{ox}(oxidised\ geometry) - E_{red}(oxidised\ geometry) \qquad \textbf{(29)}$$

where $E_{ox}(oxidised\ geometry)$ is the final energy of the geometrically optimised system in the [Fe_4-S_4]$^{2+}$ oxidation state while $E_{red}(oxidised\ geometry)$ is the energy of the metal cluster in its reduced (+1) state and it provides a value related to the electronic structure of the system. If there is a difference in the energy between the FeS clusters in the two enzymes, the capacity of molecular oxygen to oxidise the cluster will be different, demonstrating a contribution of the protein environment to LAM air-stability.

The tendency to be oxidised is explained by the definition of reduction potential (E°) of the metal cluster, which represents the likelihood for the FeS cluster to gain electrons so that to be reduced to the catalytically active oxidation state[179] and is reported by Eq.30:

$$M^m + ne^- \rightarrow M^{m-n} \qquad \textbf{(30)}$$

More positive values of E° increase the ability of the metal centre to acquire electrons, which makes it less prone to oxidation and similarly, centres with less positive values of E° are more

favourably oxidised. With a higher redox potential, the centre is less susceptible to oxidative damage from reactive oxygen species and molecular oxygen. Another phenomenon that affects the midpoint potential is the electrostatic environment around the cluster made up by the hydrogen bond network established from either the amide backbone or active site residue sidechains to the sulfur atoms of the ligands or the bridging sulfides (formally NH··S hydrogen-bond).[180] The electrostatic field generated by residues in the protein environment is thus a likely candidate to focus on instead of studying every single residues around the prosthetic group and their single contribution to the electrostatics. The search of the relevant residues is not a trivial task because many are the amino acids in the active site and studying each of them would be a time-consuming task.

A way to logically choose the right amino acids was found in a study conducted by Harris and Szilagyj[176] focused on the protein environmental effects over the electrostatic potential of FeS clusters. They developed a method that facilitates the choice of those residues in the protein environment around the FeS cluster that are likely to affect its electronic structure and, in turn, tune its redox potential. Based on QM calculations over the contribution to the electrostatic field by different amino acids, they provided a set of rules for the identification of redox modulating residues in proteins. Their work focused on the protein environment of a $[Fe_2-S_2]$ cluster contained in a FeFe-hydrogenase, but they demonstrated in the same work that the set of rules is applicable to the prosthetic group of any metalloenzyme. For this reason this guide was applied to the $[Fe_4-S_4]$ cluster of LAM enzymes. The rules are aimed at identifying amino acid residues that are likely to hydrogen bond to the metal cluster, by mapping four coordination spheres where such residues are present. The criteria employed to choose and include residues in coordination spheres concentrated on the most important functional groups that are likely to hydrogen bond to the $[Fe_4-S_4]$ cluster within the LAM active site, namely:

- Coordinated cysteinyl ligands, including the neighbouring residue to their carbonyl group peptide bond

85

- Residues that hydrogen-bond to sulfur atoms to which Fe ions are bonded to (S either in the cluster or in the ligand)

- Carbonyl moieties in the amino acid sidechain or the carbonyl group from their backbone within 4 Å of the S atoms to which Fe is coordinated to

- Residues with carbonyl moieties where the O⋯Fe distance is less than 10 Å

- Buried charged groups within 7 Å of Fe ions

Such rules were applied to build up the redox-modulating residues map of each LAM enzyme to use a model for creating *in-silico* LAM mutants carrying the differences in the active site electrostatic environment.

The use made of the set of rules only provided a qualitative analysis of the protein environment experienced by the FeS cluster in the two LAM enzymes and the electrostatics generated. A complete investigation of the electrostatics and the redox potential would have included in the calculation other factors known to affect these features such as solvent accessibility,[181] the protein matrix,[182] the reorganisation energy of the cluster[183] and the configuration of main-chain amide groups.[184] The purpose of this project was limited to the identification of those residues that could possibly have a major impact on the FeS cluster midpoint potential without considering other the factors above mentioned. This approximation was considered sensible as the two LAM enzymes might share several common features at the active site level, but a detailed calculation would generate a wider picture of the case by highlighting other contributions.

2.6   Other computational details

The MD trajectories were produced by all-atom molecular dynamics (MD) simulations packages AMBER14 and 16,[185] which are suites of biomolecular simulation programs and allow the subsequent analysis with the tool *cpptraj* included in the suite. All the simulations

were carried out using either multiple central processing unit (CPU) or graphics processing unit (GPU)[186,187,188] acceleration installed on the high performance computing cluster (HPC) Minerva of the University of Nottingham. With 166 compute nodes, 10 of which with 128GB RAM, 500GB SATA hard drive and 135TB usable storage capacity, the HPC facility allowed the expansion of data processing and computation, helping to perform faster large-scale simulations and their analyses relative to the performance on a single machine. Due to the timing of graphic card procurement by the group, the early simulations (relating to the oxygen sensitive simulated systems) have been performed through the CPU acceleration, whilst the latter ones have been produced with the benefit of the faster calculations allowed by the NVIDIA graphic card installed on the HPC. NVIDIA GPU graphic cards can largely accelerate the PMEMD program in Amber14 in performing simulations.

All structural diagrams, pictures, and dynamic visualisation of trajectories were obtained through the VMD visualisation software.[189] The molecular graphics and the building of the multimeric form of each of the simulated systems was performed using the UCSF Chimera package[190] and all the graphs and plotted values were displayed and rendered using Gnuplot program.[191] The Schrödinger suite of programs was used to perform molecular docking of the SAM molecule into the homology model using Maestro software.[192,193]

3 Homology Modelling and MD simulations of radical SAM enzymes

3.1 Homology modelling of THIC enzymes and homology model validation

By comparing the MD simulations of a related pair of enzyme structures it is possible to validate if the homology model shows the same structural dynamic behaviour as the simulation from the crystal structure, which is a good indication for a reliable homology model. The suitable pair of enzymes identified were the *S*-adenosylmethionine-dependent radical enzyme thiamine pyrimidine synthase, THIC, from *Caulobacter crescentus* (PDB code: 3EPM, 3EPN, UniProtKB: Q9A6Q5, 612 amino acids)[150] and the *Arabidopsis thaliana* THIC (PDB code: 4N7Q, UniProtKB: O82392, 644 amino acids)[151] that share 55% sequence identity from their sequence alignment. Both enzymes contain the characteristic TIM barrel of radical SAM proteins in their homodimeric structure[152] and a [Fe_4-S_4] cluster and bound SAM molecule, used as a co-substrate. The cluster-binding cysteine residues (CYS568, CYS571 and CYS576) of *Cc*THIC are located in the C-terminus part of the protein, in contrast to the majority of the radical SAM enzymes, but this section is missing from every THIC crystal structure available. A few residues were absent in the *At*THIC crystal structure 4N7Q: namely residues 71 – 84 and residues 557 – 652, with a different *At*THIC crystal structure (PDB code: 4S25) having different coverage and missing residues at 69 – 78 and 592 – 644. Modelling of the missing residues was performed by the *ab initio* method included in I-TASSER due to the lack of templates, consistent with the methodology used to prepare the *Bs*LAM homology model. However, the radical SAM enzyme TIM barrel structure was entirely conserved, so the active site and most part of the protein was used to compare their dynamics and check how similar they would have been.

I-TASSER was then used to generate the homology model of *Cc*THIC and template search restraints were applied to drive the search of template structures by excluding the *Cc*THIC

crystal structure from the search template, by excluding its PDB file. LOMETS found the best templates as deriving from the *At*THIC crystal structure (primarily 4N7Q.pdb, with a single hit from 4S25.pdb), confirming to be the best matching structure after the three-dimensional *Cc*THIC one. The model with the highest C-score of 0.05 and TM-score of 0.72 ± 0.11 was chosen and the RMSD was estimated by I-TASSER to be 7.6 ± 4.3 Å. As has been highlighted in the previous section, I-TASSER RMSD predicted values may not be as reliable as the other criteria to describe the quality of the model topology. We therefore measured through VMD the RMSD of Cα atoms between the *Cc*THIC monomeric model retrieved from I-TASSER and the Chimera-edited *Cc*THIC crystal structure monomer after addition of missing residues (as explained in the next section) obtaining an overall RMSD of 1.29 Å. These results were consistent with the values obtained for the *Bs*LAM homology model that reported similar values (C-score 0.04, TM-score 0.72±0.11 and estimated RMSD 7.1±4.1 Å) as it is shown later in the chapter. The generated *Cc*THIC model was consistent with its own crystal structure as it is shown by their superposed structures in Figure 13a (the X-ray structure is represented in green and the homology model in purple). The stereochemical quality of the model structure was assessed by the analysis of the overall conformation and residue-by-residue geometries by estimating steric clashes and plotting the distribution of phi and psi torsional angles of each residue composing the protein structure. By looking at the disposition of the φ and ψ backbone dihedral angles in the plot, amino acid residues with either strained or unusual torsion angles are identified from the model geometry and in this way torsional clashes can be indicated. The Ramachandran plot of the *Cc*THIC homology model (fig. 13b) reported the backbone torsion angle distributions to be 75.7% in the most favoured region, 18.4% in the allowed region, 3.6% in the generously allowed region and 2.3% in the disallowed region. The torsion angles distribution was consistent with the gaps highlighted during the homology model generation that, because of a lack of structural templates, residues 91 – 127 and 155 – 175 were predicted by the I-TASSER *ab-initio* method. As shown in figure 14b, most of the residues lying in the disallowed region correlate with those within

the gaps.

The Ramachandran plot of the model was compared with that one generated for the crystal structure (Figure 13c) after addition of the missing residues by Chimera as reported in the next section. The distribution of the dihedral angles was reported to be 88.1% in the most favoured region and 11% in the allowed region whilst 0.2 and 0.7% in the generously allowed and disallowed regions respectively. Most of the amino acid residues in the disallowed region (Phe93, His97 and Ala98) belonged to the residues lacking in the initial structure and so were added by Chimera; thus they might show some imperfection. The two distributions did not
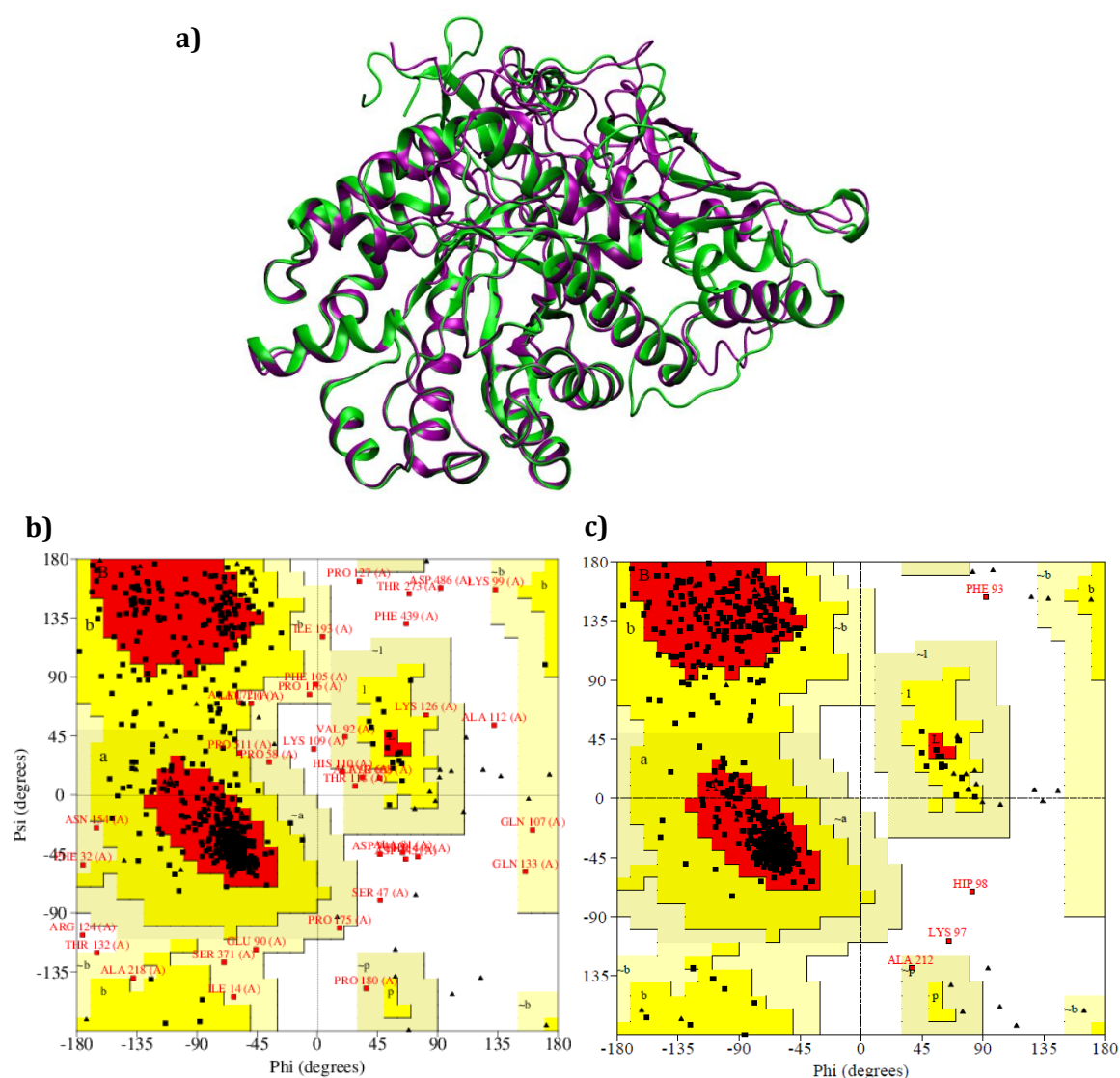

Figure 13. a) Overall structural superposition of the *Cc*THIC crystal structure (green) and the generated homology model (purple). b) Ramachandran plot of the *Cc*THIC homology model and c) the crystal structure after refinement.

91

report remarkable differences, likely due to the absence of local templates in the homology model generation and the missing residues in the crystal structure. Despite the limitations that such a structure possesses due to possible imperfections that the homology model might have reported, MD simulations of the THIC homology model compared to the crystal structure were still relevant. Furthermore, as the crystal structure of the enzyme for which the homology model was generated was available, we newly retrieved the *Cc*THIC model structure by removing restrictions to I-TASSER so it could generate the model by using the *Cc*THIC crystal structure as a template. The RMSD of the Cα atoms between the new *Cc*THIC model and the *Cc*THIC crystal structure reported a deviation of 0.56 Å. Such a low value added a further estimation that a good quality model of the *Cc*THIC enzyme had been achieved.

3.2   Thiamine biosynthesis enzyme THIC MD simulations

3.2.1   The crystal structure and its modifications

MD simulations of the THIC crystal structure and its homology model were carried out in one run for at least 400 nanoseconds in order to provide evidence that the homology model generated by I-TASSER with relatively low sequence identity could be considered reliable for a deeper computational study. The simulations were performed in an orthorhombic box of water for the homodimeric structures without [Fe_4S_4] cluster, SAM and substrate contained. The crystal structure PDB file (PDB code: 3EPM) was stripped of water, cofactors and substrate, including removal of definitions for the HMH, SO_4^{2-} and Zn^{2+} cofactors with consequent solvation of the left gaps. A few residues were missing from the crystal structure and were added by the Refine Loops tool in UCSF Chimera where missing segments are built by *de novo* techniques. Residues 98 – 111 and 223 – 227 (and residue 97 in monomer B of the crystal structure) were added by using the Discrete Optimised Protein Energy (DOPE) tool.[153] By using Lennard-Jones potentials, DOPE derives an atomic distance-dependent statistical potential from a sample of the native structure, it is based on an improved reference state that

corresponds to non-interacting atoms in a homogeneous sphere with the radius dependent on a sample native structure. The quality of the newly built segments was assessed with a normalised zDOPE score where negative values indicate a better model. Loops rated with the lowest score were therefore chosen. The best allowed modelled loops by the Chimera software afforded different values for each of the two monomers composing the missing loops in the THIC dimer. The final best zDOPE score was -1.50 for monomer A in the THIC crystal structure and -1.46 for monomer B of the dimer and these were incorporated to generate the final *Cc*THIC structure. The PDB structure was submitted to the H++ webserver for analysis of the predicted protonation states for charged residues. The protonation state of the histidine residue on the surface was estimated to be in the fully protonated form HIP whilst the majority of histidines were suggested to be protonated on their nitrogen atom at the ε–position (HIE). Once the structural PDB file had been prepared, it was fed into *xleap* to create the system topology, parameter and coordinate files used for the production of simulated MD trajectories.

3.2.2 The assembly of the homology model dimeric structure

The generated homology model monomer structure was used to build up the homo dimeric structure of THIC, consistent with the crystallographic data, using the visualisation software UCSF Chimera. The Chimera MatchMaker tool was used to fit both the monomers to the *Cc*THIC crystal structure by superposition achieved through RMS-fitting. A TER flag between the two enzyme chains was added to the retrieved PDB file to prevent the software reading it as one unique long monomeric chain. The protonation state of the charged residues was checked and refined with the H++ webserver, suggesting the majority of the histidine residues were to be assigned the mono-protonated state (HIE) compared to the doubly-protonated states (HIP) found on the enzyme surface. The subsequently generated PDB file was used as the input for *xleap* to generate the files for the submission run.

3.2.3 MD simulation analysis and the homology model validation protocol

All-atom molecular dynamics (MD) simulations were performed on the *Cc*THIC crystal structure and the *Cc*THIC homology model. Only the apo-form of the enzymes were used for production of MD trajectories of the THIC structures for which the same protocol used for MD simulations of LAM were applied (see section 3.4). The final trajectories have been reimaged and stripped of solvent and counterions by using *cpptraj* and observed through VMD. *Cpptraj* was used to calculate the RMSD of backbone atoms and RMSF of alpha-carbon atoms of residues composing the two enzymes compared to the average structure retrieved from their respective THIC MD simulation. The RMSD of the two THIC simulations compared to their starting structure are shown in figure 14a and 14b for the crystal structure and the homology model respectively. In both the enzyme MD simulations a stationary phase was reached after
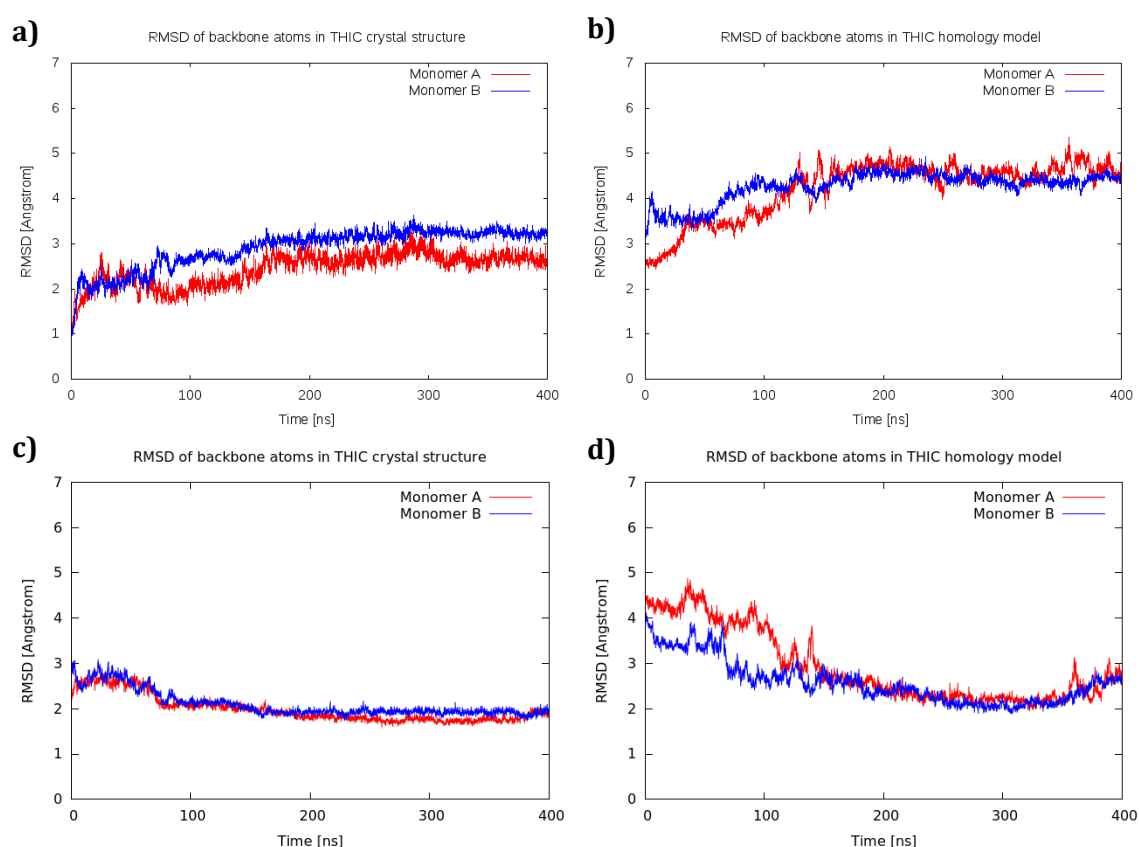

Figure 14. RMSD of 400ns MD simulation of crystal structure THIC (a, c) and the homology model (b, d) compared to the starting structure and to the average structure respectively. In red and blue, monomer A and B are coloured.

~100 ns. The backbone atoms of the homology model (figure 14b) showed a ~1-2 Å higher deviation from the reference structure than in the case of the crystal structure THIC but it did not seem to account for different conformational structures. A further RMSD analysis was performed to the backbone atoms in the two enzymes compared to the average structure retrieved between 200 – 400 ns of the *Cc*THIC crystal structure (figure 14c) and the THIC homology model (figure 14d). In the case of the refined crystal structure the stable spatial atomic distribution was reached after ~100 ns whilst the homology model took ~50 ns more and in both cases the average RMSD values was around 2 Å for each monomer composing the enzymes. The relatively higher trend in the homology model showed in figure 14b was due to the fluctuation of loops and less defined structural areas as showed by the analysis of the RMSF of Cα atoms compared to their average structure as for above (figure 15). The homology modelled structure (Fig. 15b) showed more significant fluctuation overall, especially around residues 50 – 100 where loop reconstruction had been needed. This sequence corresponded to the peripheral loop on top of the enzyme surface that was likely to have some more free mobility. Figure 15a reported that the *ab initio* reconstructed missing loops of the crystal structure (residues 97/98 to 111 and 223 to 227) looked more oscillating than the rest of the structure and it was expected to be a less ordered area. In both the crystal structure and homology model a structural refinement on their N-terminus was necessary because some
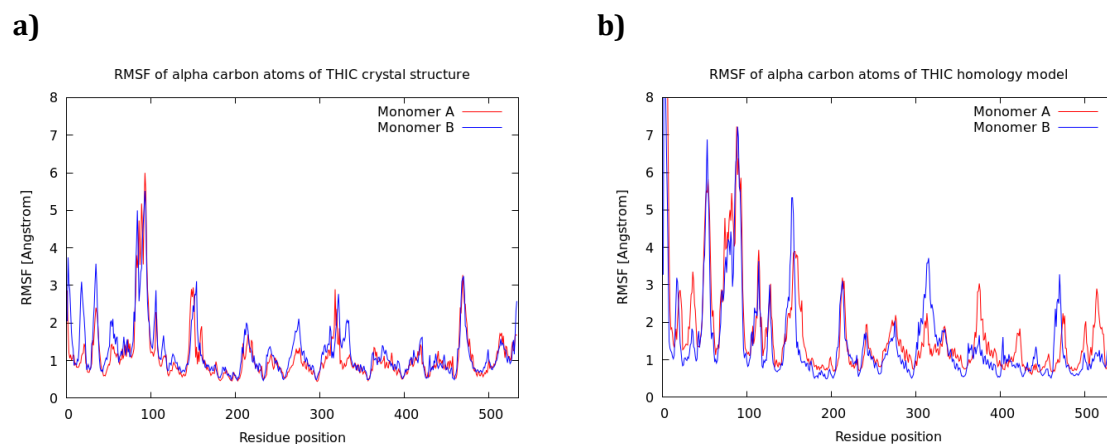
a)   b)

Figure 15. RMSF of Cα atoms compared to the average structure in the crystal structure THIC (a) and the homology model (b).

loops were missing from either the template employed to generate the model or from the published enzyme 3D-structure. The overall enzyme dynamics was very similar between the two enzymes suggesting a consistent fluctuation of the THIC enzyme in its apo-form without cofactors and substrate present. Principal component analysis of Cα atoms provided a different view of the motion of the THIC models during the MD simulation as shown in Figure 16. The analysis of the principal components of the dynamics in the THIC structures highlighted some differences between the two simulations. From a visual inspection, it was possible to identify the dominant dynamics observed during the simulations by sampling the two eigenvectors reported to be in common between the two MD simulations and projecting the trajectories along each eigenvector. A big portion of the overall fluctuation of the THIC enzymes could be recorded by low-frequency eigenvectors that are generally described by the first principal components. The analysis of the two principal components was performed using a common subspace defined by PC1 and PC2 (Proj 1 and Proj 2 respectively, Figure 16) to look at the conformational sampling in the two MD simulations. Both the structures were
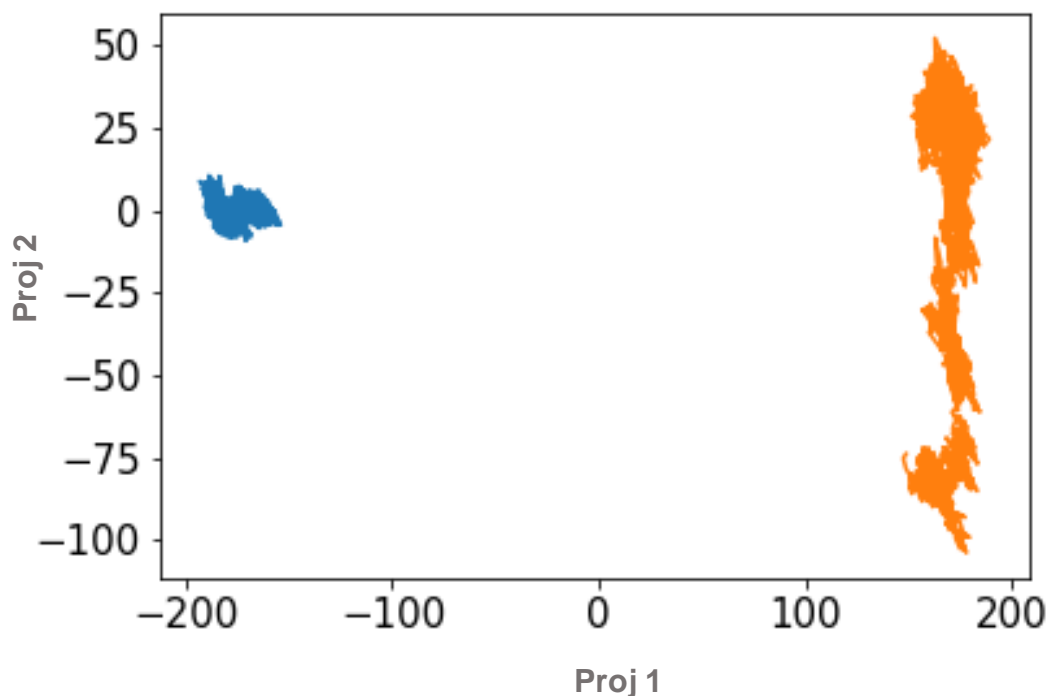

Figure 16. 2D maps of the projections of first and second PC in both THIC simulations. The sampled conformational states differed along the PC1 identifying the crystal structure as a basin on the left (in blue) and the homology model on the right side of the plot (in orange).

populated by one main conformational state reporting some differences between the two as shown by the two energy basins in the plot , representing the crystal structure depicted in blue and the homology model in orange (figure 16). The two energy minima described by the first two PCs, showed differences along the PC1 (by ~320 kcal/mol) compared to PC2 (by ~30 kcal/mol). The first *eigenvector* in both simulations accounts for the largest and slowest motion in the THICs dynamics and it sampled two protein conformations that are energetically more different rather than those sampled by the second principal component.[194] The dynamics in the two enzymes were defined by one conformational state although in their dynamics there might have been some other fluctuations that took part in the overall dynamics and could not be sampled by the first two principal components. The visualisation of the first PC helped with the identification of those areas reporting the major variation along the THIC structures. The trajectory corresponding to the first PC is represented in figure 17 in tube representation showing the large-scale motion of the protein. As it could be observed,
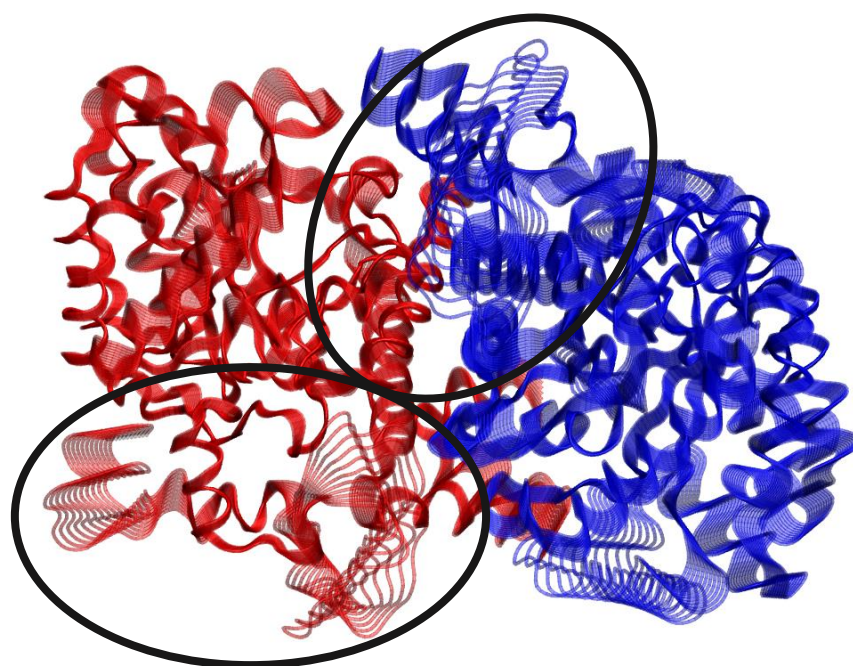

Figure 17. Visualisation of the first principal component sampled from the combined trajectories of THIC crystal structure and homology model MD simulations. The structures are represented in tube, monomer A is shown in red whilst monomer B in blue. The ovals enclose the N-termini of both monomers as the areas of major variation.

the areas of wider fluctuation between the two enzymes involved those residues found in the N-terminus loop of the two monomers (highlighted by the dark grey ovals in fig. 17). This data was consistent with the RMSF confirming that the disordered motion was likely to be due to ~110 residues positioned in the N-terminal loop in both the enzyme structures. The dynamic contribution of these wide oscillations were sampled by the PCA resulting in different contributions to their conformational state that made the two basins appearing distant from each other. The PCA described two structurally different sampled conformations between the two THIC enzymes and that might be due primarily to the flexible loops that were not defined in the *Cc*THIC crystal structure. A more concentrated PC analysis provides a wider picture with more information about their motion.

To confirm the similarity of the protein dynamics between the two THIC MD simulations when the residues in the N-terminal loop were exempted, PCA was run again by excluding residues 1 – 110 of both monomers from the sampling (figure 18). This separated the loop
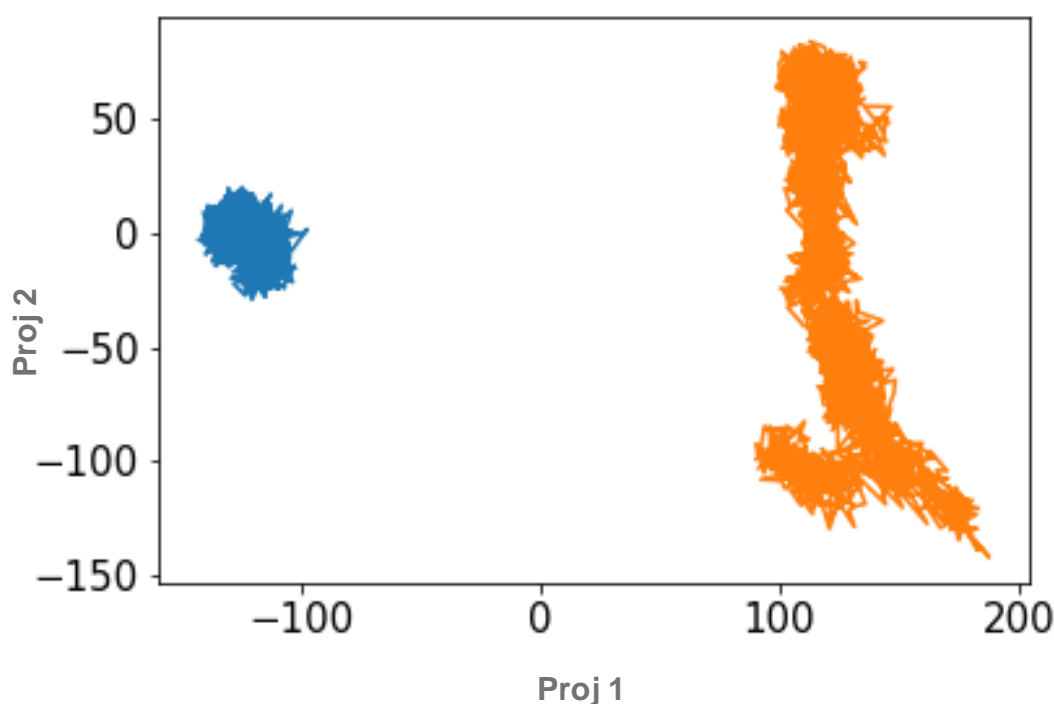

Figure 18. New combined PCA of THIC MD simulations with removed N-termini from the dimeric structure. Projections of PC1 and PC2 identified two different basins again, for the crystal structure (on the left) and the homology model (on the right).

that showed the most disordered motions from the other residues composing the 3D structure of both the THIC enzymes. In Figure 18, the new PC analysis run on both simulations combined, collected in total 35 eigenvectors instead of the only two ones that were present in the previous analysis suggesting the sampling of more collective modes between the two THIC simulations. Once again, the first two principal component, displayed the presence of one largely populated conformational state in each simulation. The two large basins described the sampled conformational states in the two THIC structures that were driven by their eigenvectors and they looked more similar and closer to each other. The new RMSD and RMSF of the THIC MD simulations compared to their average structure with the removed N-terminus from both monomers reported very similar trends as shown in Appendix 6.3. The RMSD of both the crystal structure (a) and the homology model (b) looked stable at the value of 2 Å for both monomers in THICs and for the whole MD simulation, consistent with a decreased disordered contribution to the overall enzyme dynamics. The RMSF reported a fluctuation peak of ~1.5 Å around residue 150 in the homology model compared to the crystal structure. That did not represent an issue as the fluctuation was also noticed in the previous RMSF analysis whilst the rest of the homology model structure kept similar oscillation amplitude as before. A small contribution to protein dynamics, although, still brings a significant separation along PC1 between the two energy minima as to suggest the two simulations converged on the PC2 rather than PC1. Therefore, we looked at the next sampled principal component, PC3, in the subspace overlap together with the projections of PC2, Figure 19. What is reported, is one large basin representing the superposition of the two conformational states sampled by PC2 (Proj 2) and PC3 (Proj 3) in the two MD simulations. The two energy minima nicely overlaid due to the closeness of the sampled collective modes between the two MD simulations suggesting the high degree of similarity between the two enzyme dynamics. Without considering the contribution to protein dynamics arisen from the fluctuating loops at the N-terminus, the two energy minima looked equal enough to reach the superposition and matching of the sampled eigenvectors. In Figure 20 are shown the second

and third eigenvector of the combined trajectories for both THIC MD simulations as representatives of the relevant principal components that drove the protein dynamics in both the crystal structure and the homology model. Porcupine plots indicate the motion of Cα atoms in principal components, with "spikes" pointing in the direction of the motion of the atom which length describes the amplitude of the motion. Porcupine plots represented the loop fluctuations sampled within the favourite conformational state driven by the PC2 (Fig. 20a) and PC3 (Fig. 20b). The enzyme dynamics was visually more ordered without prominent oscillation of any area of the enzymes and it was observed driving the fluctuation of the TIM barrel, specifically the peripheral α-helices around it. Few residues could be observed having a wider fluctuation compared to the rest of the structure but that did not affect the overall conformational state. The energy difference along the PC3 was likely to be originated by remaining peripheral loops where the RMSD was less similar in the two enzymes as shown by their structural alignment in Appendix 6.5.
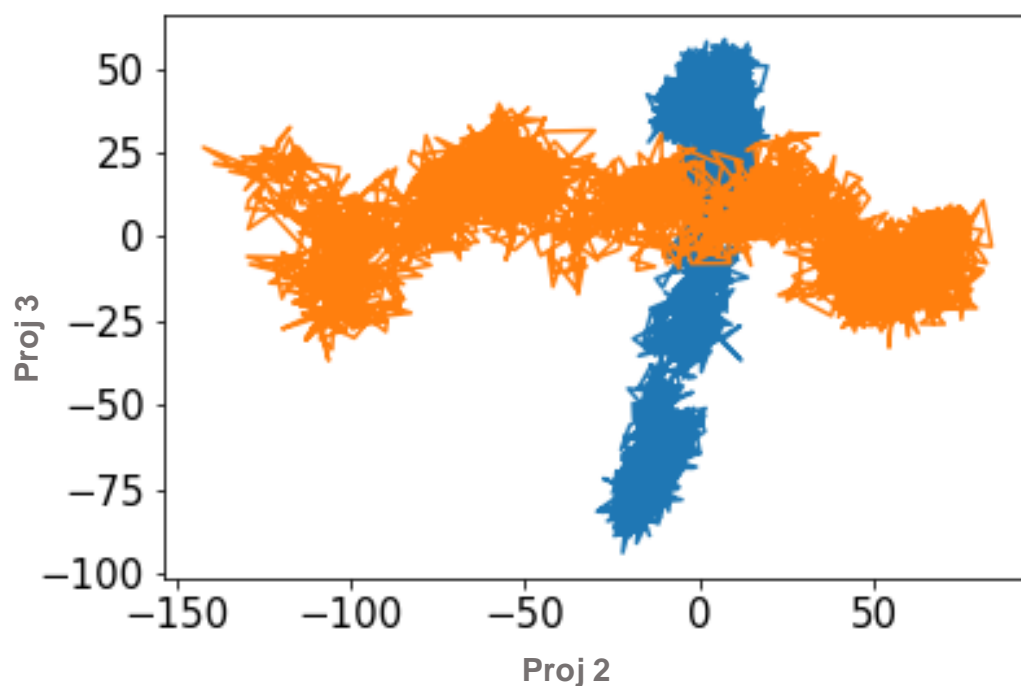

Figure 19. PCA of combined THIC trajectories over the PC2 and PC3. THIC MD simulations with removed N-termini from the dimeric structure. Projections of PC2 and PC3 sampled the superposition of the eigenvectors sampled in the two THIC MD simulations. The crystal structure is represented in blue and the homology model in orange

By calculating the dot product matrix between the eigenvectors identified by the PC analysis performed on the two MD simulations individually, it was possible to look at the collective modes data files of the two enzyme structures. The subspace overlap between the two THIC enzymes could estimate the correlated motions of the two MD simulations individually, reporting an overlap of 38.3% for the whole enzyme systems whilst it reached 51.3% when the N-termini were removed from the sampling. It has been previously reported that the subspace overlap between two MD simulations accounting for approximately 50% means
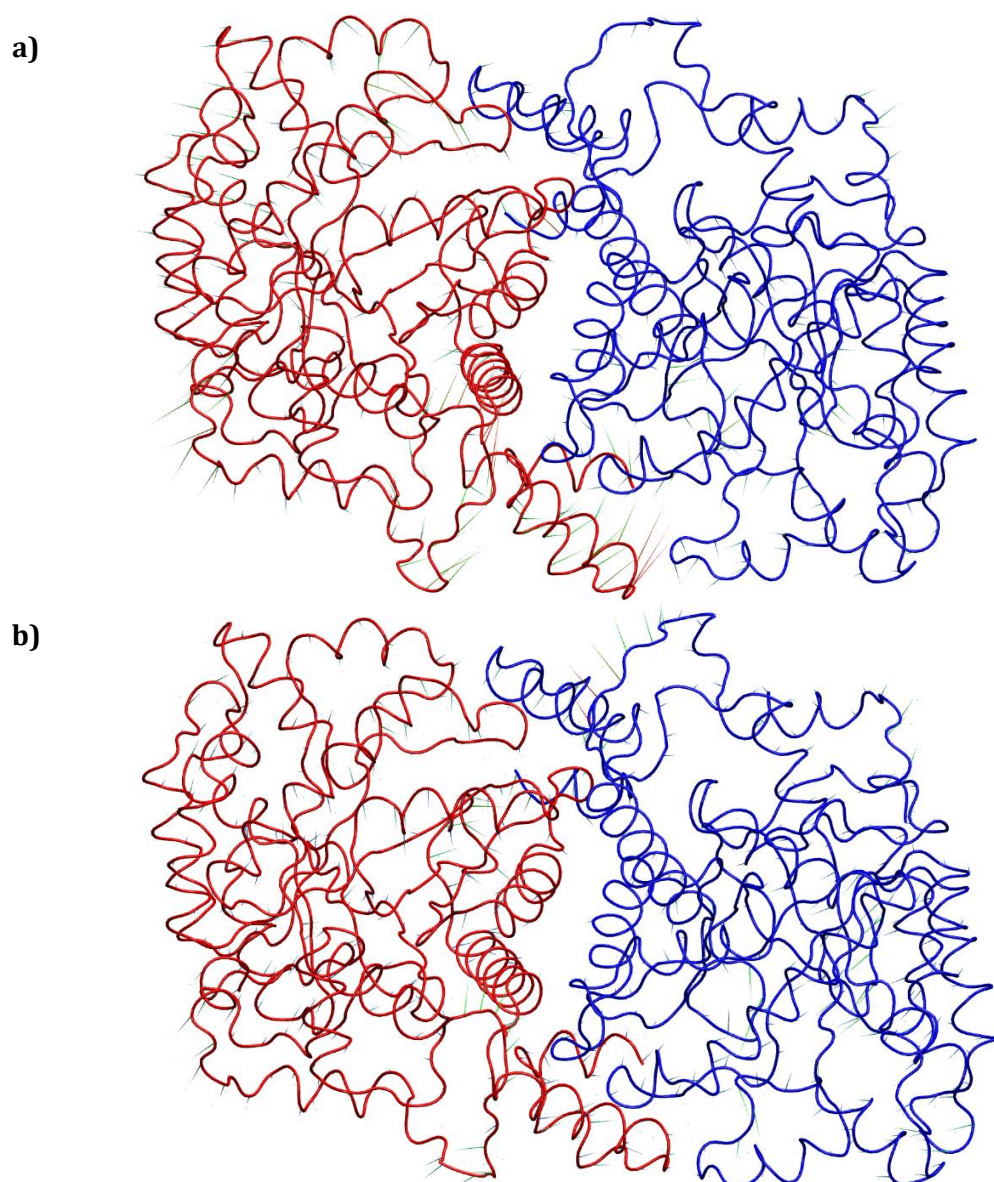
a)

b)

Figure 20. Porcupine plots of sampled PC2 and PC3, a) and b) respectively, of combined THIC crystal structure and homology model MD trajectories with N-termini removed from the sampling. The main fluctuation was localised in the peripheral helices around the TIM-barrel like structure.

that similarities of the subspaces of the individual trajectories are significant and that confirms the quality of the model.[195,196] The analysis of the principal components sampled both the MD simulations in the same area of the common subspace composed by the second and the third projections of the PC subspace. The matching *eigenvectors* described similar fluctuations sampled in the conformational space explored by the two MD simulations, which could account for equivalent dynamics in the THIC structures. The N-terminus was the part of the enzyme containing the area with missing templates or *de novo* generated loops that was observed to be the most disordered and fluctuating area of both the THIC enzymes. Although, it needs to be considered that possible *ab-initio* generated enzyme loops during the homology model generation by the I-TASSER webserver, might be the least reliable part of the whole protein structure in MD simulations. Despite the low sequence identity of 55% between the THIC crystal structure and the template adopted, that was lower than that between the two LAMs (60%), we could estimate that the model was of good quality with respect to the dynamics. Dynamically speaking, the homology model built with this level of similarity to its template could be considered good enough to conduct an MD simulation study on other enzymes, as in the case of LAM enzymes.

The assessment of the homology model showed the model to have overall a good quality despite the relatively low sequence identity (55%) with the best identified template and the decent C-score (0.05). Furthermore, the Ramachandran plot analysis reported a dihedral angles distribution to be slightly acceptable as residues identified in allowed regions were <80%. Conversely, from the comparison of the fluctuations between the two MD simulations of the THIC structures it was reported a general similarity, but that was not enough to validate the model structure. The initial PC analysis evidenced quite distant relatedness in the conformational states explored during the MD simulations between the two THIC structures. A closer analysis of the PCs revealed that the main differences in the two structures were originated from flexible loops that had no reliable templates to be modelled with; however, those loops were not part of the TIM-barrel structure and their removal from the analysis did

not significantly affect the outcomes of the investigation. The overall THIC core structures, excluding those fluctuating loops, were reported to behave similarly over the time length of 400ns, particularly the TIM-barrel structure that accounts for the active site. The matching *eigenvectors* that was evidenced from the subspace described by the second and third PCs, confirmed the similar behaviour in the two structures, hence the good quality of the generated model.

3.3   *Bs*LAM homology modelling and quality assessment

The *Bs*LAM amino acid sequence was used as the target sequence in I-TASSER to predict its 3D structure. Each threading program employed by the webserver identified the *Clostridium subterminale* SB4 LAM[197] in complex with *S*-adenosylmethionine at a resolution of 2.1 Å (PDB code: 2A5H, UniProtKB: Q9XBQ8, 416 amino acids) as the best template. The sequence identity of 60% was considered sufficiently high to build a reliable model of *B. subtilis* LAM. The template was aligned up to residue 425 of the target sequence, and the remaining 46 residues were built by I-TASSER using *ab initio* structural modelling methods due to a lack of templates. The C-score reported by I-TASSER for the proposed *Bs*LAM model was 0.04 with TM-score of 0.72±0.11 and an estimated RMSD of 7.1±4.1 Å. The model was estimated to be of high confidence and with a TM-score in the range of the correct topology, despite the relatively high RMSD values that have been illustrated previously might happen even in the case of good C- and TM-scores.[198]

The structural overlay between the predicted *Bs*LAM model and the 2A5H crystal structure template, along with the overlay of the key residues required for [Fe_4-S_4]-binding (inset) is shown in Figure 21. The helices and strands are well superposed and the upper red helix (highlighted by the arch in grey in the figure) represents the proposed conformation of the ~60 residues C-terminus extension typical of the *Bs*LAM, relative to the air-sensitive *Cs*LAM. The correct fitting of the model to the template is a good starting point in confirming a reasonable structure, but not considered reliable enough for further molecular dynamics without validation being carried out. The overall structural alignment looks correct but as it is shown in detail, the cluster-binding loop is not perfectly fit. This discrepancy might due to the difference in sequence alignment in that area of the enzyme. Even though the validation protocol estimated the reliability of the model created from I-TASSER, its quality was further assessed to look in detail whether poorer defined enzyme areas were present. Local RMSD values were measured for different regions of both the template and model structures to make a more detailed comparison with the RMSD estimated from I-TASSER.

The RMSD calculator tool included in the VMD software was used to calculate the root mean square deviation of specific areas in the two enzyme structures of *Bs*LAM homology model and *Cs*LAM crystal structure. The overall enzyme RMSD was measured taking into consideration the last amino acid residue in *Cs*LAM as the *Bs*LAM C-terminus cannot be overlaid for lack of structure in *Cs*LAM. The RMSD values across specific sections are reported in Table 1. This preliminary analysis was meant to look at the similarity of the generated

The longer *C*-terminus loop of *Bs*LAM
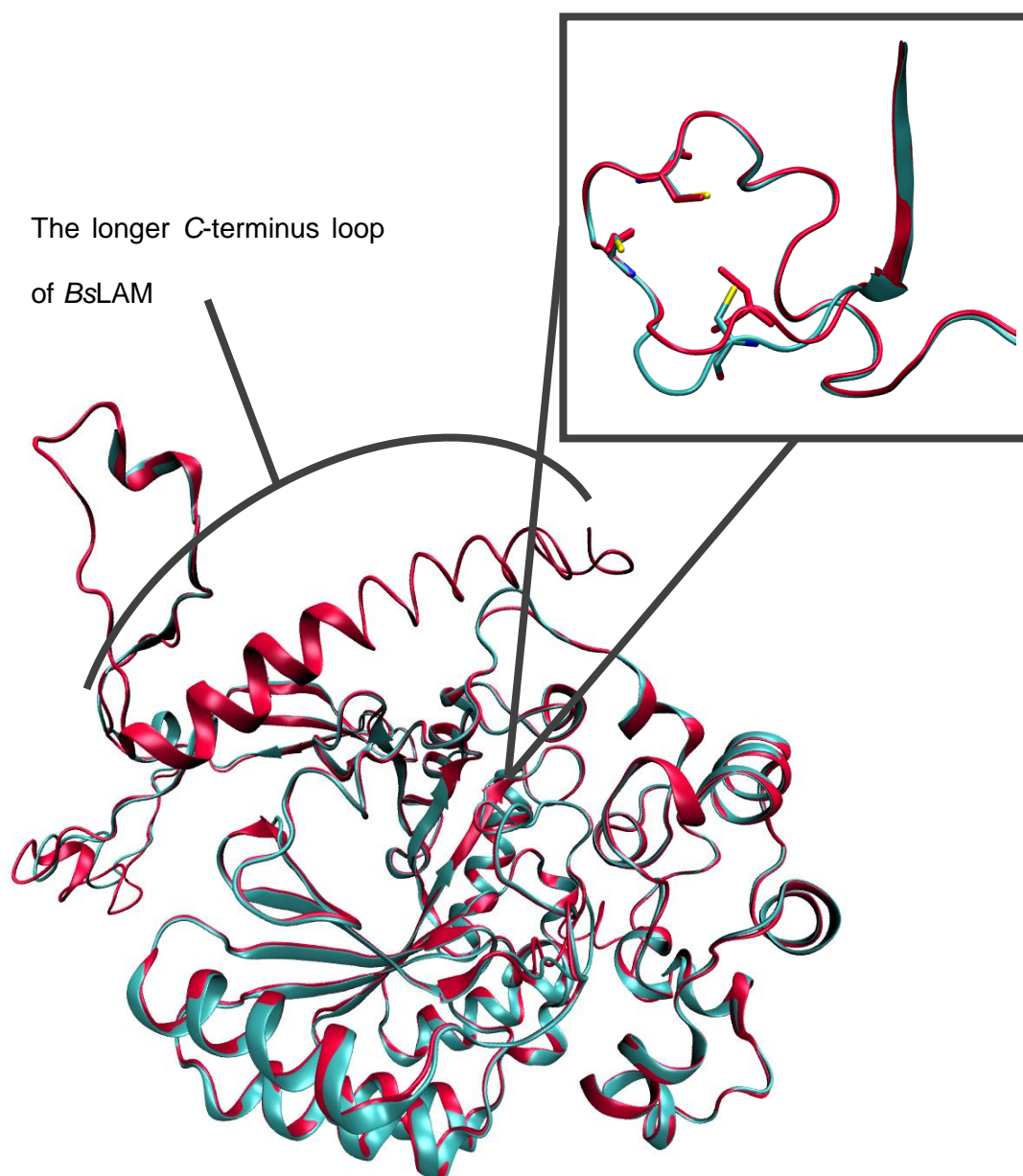
Figure 21. Structural overlay between the crystal structure *Cs*LAM (in cyan) and the *Bs*LAM model (purple). In the square, detail from the insight of the TIM barrel with both the enzymes cluster-binding cysteine residues. The long *Bs*LAM C-terminus helix is highlighted from the dark arch.

homology model compared to the crystal structure so to estimate the local similarity and the poorly defined areas of the model. The N-terminus loop (**I**) has a low RMSD value of 0.5 Å, indicating close similarity with the crystal structure in this initial part of the enzyme whilst in the pyridinium-coordinating loop (**II**) has a deviation of 0.9 Å, slightly higher but still consistent with a good structural similarity. The cluster-containing loop (**III**) also had a low deviation of 0.9 Å, despite the structural alignment being visually different as previously reported in the detail shown in figure 21. The areas containing the $\beta4$ and $\beta5$ loop (**V** and **VI**), involved in SAM coordination and orientation, reported the lowest values of 0.3 Å whilst the PGGGGK motif region (**VII**) showed a low RMSD of 0.6 Å as well as the whole enzyme superposition (**IX**, without including the *Bs*LAM C-terminus extension) that gives a good RMSD of 0.6 (**IX**). The area with the most substantial deviation between the two enzymes is (**IV**) with a RMSD of 1.2 Å, however this section does not have an established role in the enzyme catalysis. Although the total RMSD appeared good for the majority of the model structure, the non-linear values of the local RMSD indicate that the model is imperfect and we may expect similar fluctuations between **II**, **III**, **IV**, and **VIII** loops of *Bs*LAM as they reported

	I	II	III	IV	V	VI	VII	VIII	IX
LAM residues	52-111	112-132	125-133	134-144	172-202	220-250	330-350	374-389	
BsLAM Model residues	61-120	121-141	134-142	143-153	181-211	229-259	339-359	387-402	1-409
RMSD	0.5	0.9	0.9	1.2	0.3	0.3	0.6	0.9	0.6

Table 1. Calculated RMSD values for different parts of both LAM enzyme structure. The template sequence and bacillary model, along with the RMSD are shown in the rows divided by different sections of the structures (numbered from I to IX).

similar values of accuracy. The stereochemical quality of the model structure was assessed by the analysis of the overall conformation and the residue-by-residue geometries by estimating steric clashes and plotting the distribution of phi and psi angles of each residue in the protein structure. Ramachandran plot of *Bs*LAM model (Figure 22a) was generated by the PROCHECK suite of programs for protein structure validation. By looking at the disposition of the φ and ψ backbone dihedral angles in the plot, amino acid residues with either strained or unusual torsion angles were identified from the model geometry and in this way torsional clashes could be indicated.

The Ramachandran plot of the proposed *Bs*LAM homology model indicated the population of backbone residues to be: 71.0% in favoured, 23.7% in allowed, 3.2% in generously allowed, and 2.2% in disallowed regions of torsion angles. The amino acid residues that were found in the disallowed regions, did not belong to important areas of the enzyme as those highlighted in the sequence alignment in figure 11. Among the residues reported in either generously allowed or disallowed regions, three of them belonged to the very first N-terminal part whilst two of them were part of the *Bs*LAM C-terminal loop. The areas were built by the *ab-initio* method of I-TASSER for lack of templates and it was somehow expected to show some imperfections. The remaining amino acids that fell in the unfavourable areas of the Ramachandran plot were part of solvent-exposed loops and surface areas of the enzyme structure of minor relevance. A PROCHECK analysis of the *Cs*LAM template used to create the model was run for comparison of the dihedral angles distribution. The analysis of the crystal structure of *Cs*LAM (Figure 22b) highlighted the backbone residues distributed at 72.5% in the favoured regions of torsion angles, 21.3% in the allowed, 4.3% in the generously allowed and at 1.9% in the disallowed ones. The two distributions were consistent between the crystal structure and the homology model as it was expected, due to the model creation from its template as it minimised the differences. Nevertheless, the stereochemical analysis reported the presence of a few amino acids with an unfavourable conformation. These values might indicate a fair amount of steric clashes in the enzyme structure but it might be the

consequence of the low resolution of the published *Cs*LAM crystal structure as it could lead to worse dihedral scores.[199] That might not be an issue under the molecular dynamic aspect as the properties of the *Bs*LAM homology model were consistent with the method adopted for its generation and demonstrated previously in section 3.3. After these starting considerations we could establish the similarity between the *Cs*LAM template and the *Bs*LAM homology model. The model looked reliable for most part of the enzyme structure and its relevant catalytic areas but it had to be considered that potential concerns might have come from the bacillary LAM C-terminus due to its uncertainty. Even though to the *Bs*LAM C-terminus has not been implicated in important roles for either enzyme stability or functionality, a PROCHECK was run for the last 72 residues of the retrieved model. The results showed a distribution of 50.0% in favoured, 46.9% in allowed, 1.6% in generously allowed and 1.6% in
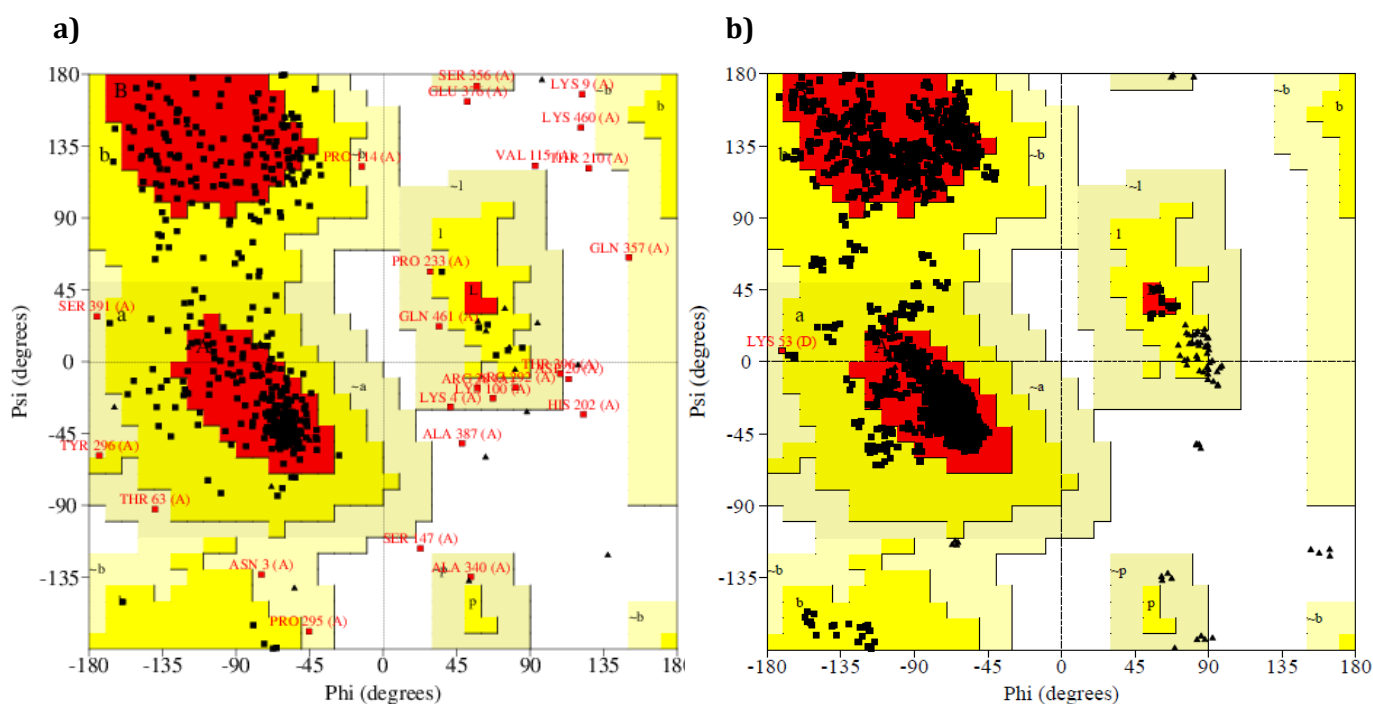
Figure 22. a) Ramachandran plot of backbone torsional angles (φ and ψ) contained in the *B. subtilis* LAM model retrieved from PROCHECK reported 71.0% in favoured, 23.7% in allowed, 3.2% in generously allowed and 2.2% in disallowed area of the plot. b) Ramachandran plot of *Cs*LAM crystal structure reported torsion angles to be for 89% in favoured, 10.9% in allowed and 0.1% in the generously allowed regions of the plot.

disallowed regions of torsion angles (see Appendix 6.6). Half of the residues considered were in the favoured region, whilst just less than half were in allowed region as it mostly looked like an α-helix. Thus, the newly created *Bs*LAM-extension could not be considered the full source of imperfection and this means that the model inherited the steric clashes of the crystal structure. The Ramachandran plot of the homology model indicated a limited number of residues as having steric clashes. Those residues did not belong to specific parts of the monomeric chain nor they was attributed to particular enzyme function. These clashes were consistent with the model template, therefore they were not considered as artefacts of the modelling process. The model created by using the *Cs*LAM crystal structure as a template had a less reliable C-terminus due to the lack of templates but the overall TIM-barrel structure and the cluster-coordinating loop were consistent in the two LAM structures. That was needed to undergo structural investigations and MD simulations of the two enzymes to figure out their differences.

This analysis reported that the static *Bs*LAM model structure was affected by few imperfections that could be mostly attributed to the *Cs*LAM crystal structure and the *ab initio* generated C-terminal loop as a consequence for the lack of suitable templates. On the other hand, the THIC homology model validation examined in the previous section, under similar sequence identity conditions as for the LAM enzyme pair, provided reliable MD simulations. These combined studies gave confidence that the *Bs*LAM homology model was good enough to carry on with the investigation of the differences between the active site and loop movements in both LAM enzymes.

The homology model validation protocol employed in this study, established that is possible to generate a reliable protein structure with conditions of relatively low sequence identity of 55 – 60%. The widely used methods reported in the literature to validate homology models are PROCHECK[158] and other analysis that can be retrieved from the Ramachandran plot (orientation of Cβ).[200] Normally the good quality of the model is assessed by the dihedral angles distribution of around 90% in the favoured positions,[201] whilst in the case of LAM it

was reported to be at 71%, so a PROCHECK run was not enough to validate the model. In other cases measurement of the z-score can provide a measure for statistical significance between matched structures has been used for validation[202,203] The RMSD can be used for validation in case the X-ray structure of the model was solved for comparison[134] but that does not normally validate a homology model. In this case it was even more difficult, because the sequence identity between *Cs*LAM and *Bs*LAM was lower compared to other previous works about homology modelling[204] and the PROCHECK analysis reported lower values than those recommended in the Ramachandran plot analysis. Furthermore, the long C-terminal loop in *Bs*LAM that was generated without a template, added more uncertainties on the quality of the model. For these reasons the validation of the model was chosen to be carried on by analysing the MD simulations of a related pair of enzymes. There is no previous reported approach that we are aware of that create the homology model of a related pair of enzymes and run MD simulation to validate the model. The approach employed could provide hints about the dynamical conformational reliability of the homology model so created, giving confidence that despite the low sequence identity, the model generated possesses a reasonably high level of accuracy to be used for further investigation.

3.4 Molecular Dynamics (MD) simulations of LAM enzymes

3.4.1 Oxygen sensitive lysine 2,3-aminomutase *Cs*LAM MD simulation settings

The X-ray crystal structure of the oxygen sensitive LAM from *Clostridium subterminale* is available from the protein data bank website (https://www.rcsb.org/structure/2a5h). To be able to use the 2A5H.pdb file in Amber, methionine residues containing selenium atom named as SE were renamed in SD. Cysteine residues at positions 268, 375, 377, 380 that coordinate zinc atoms that help keep the LAM structure together[92] were renamed as CYM to make them bind the Zn^{2+} ions. The cysteine residues involved in the $[Fe_4\text{-}S_4]$ cluster coordination were renamed from CYS into CYF, due to their later differing treatment because of their

involvement in bonding the Fe atoms of the metal cluster. For these CYF residues, a custom force field modification file was used to correctly parameterise cluster-coordinating cysteine residues. Once these major edits were completed, hydrogen atoms were removed from the file and the H++ webserver was used to suggest the protonation states for other charged residues. The only acidic residue to be predicted as having a neutral charge was GLU313, renamed GLH, and every histidine was estimated to be in the fully protonated form HIP except for HID230 and HID318. The crystal structure of *Cs*LAM and its mutants was furtherly customised by including four different combinations of the protonation states for His131 and His230 (two amino acids spatially closed to the FeS cluster), with each of the four monomers representing one of the possible states (Figure 23). The protonation variance on these important histidine residues allows assessment of the most likely natural protonation states
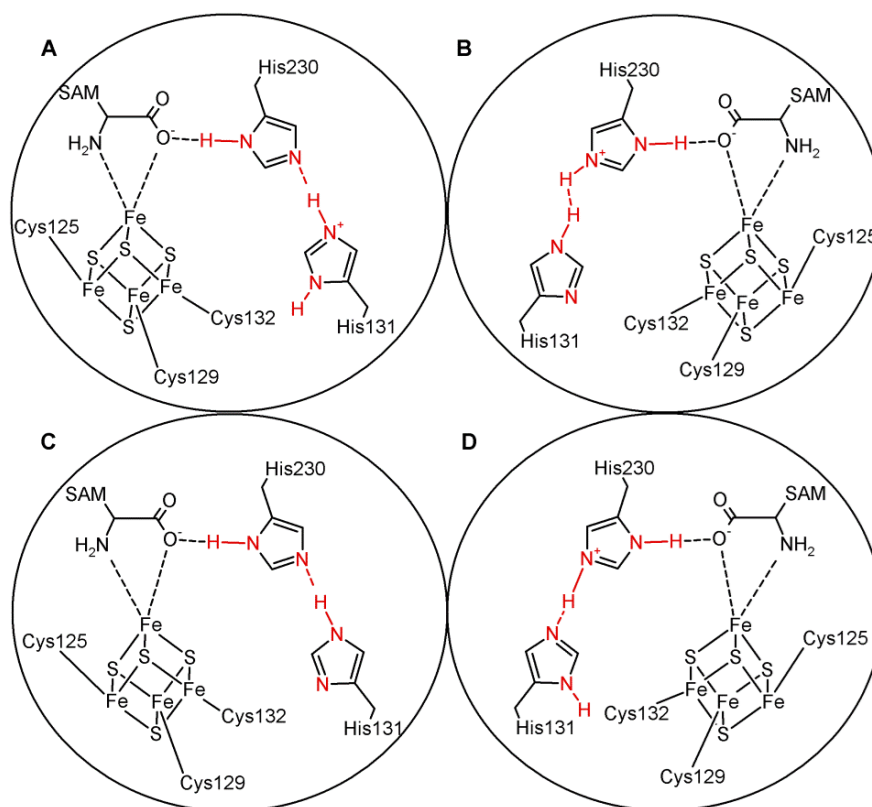
Figure 23. Customised protonation states for His131 and His230 residues that are spatially closed to the [Fe$_4$-S$_4$] cluster. Combinations were customised as HIP131/HID230 in monomer A; HIE131/HIP230 in monomer B; HIE131/HID230 in monomer C and HID131/HIP230 in monomer D. Combination A was suggested by H++.

in the absence of experimental data, where incorrect protonation could influence the dynamics of the [Fe$_4$S$_4$] cluster or SAM cofactor when present in the active site. As no information was available about the likelihood of the protonation state of the imidazole ring for these histidine residues, we chose different combinations of protonated histidine residues. This was done to ensure the representation of each protonation possibility for these residues and to test whether these combinations might have a relevant influence on the overall enzyme dynamics or the SAM-coordination. The combinations of each histidine in the active site throughout the four monomers are represented in Figure 23: HIP131 and HID230 in monomer A; HIE131 and HIP230 in monomer B; HIE131 and HID230 in monomer C and HID131 and HIP230 in monomer D. The monomer in Fig. 23a carried the protonation suggested by H++; the monomer in Fig. 23b might be the least favourable among all the monomers with the protons from each histidine close to each other; and the monomer in Figure 23d had the mirrored protonation state as 23a. This set up allowed observations of whether there are protonation-dependent influences either in the flexibility of the SAM molecule, the fluctuation of the monomers in the protein individually, or collective fluctuations that might influence the flexibility of the whole enzyme structure.

Distance restraints were applied to the [Fe$_4$-S$_4$] cluster to keep its cuboidal structure together and to prevent any possible distortion or simulation artefact to the cluster. The restraints were created by measuring bond distances in the crystal structure, between Fe and S atoms within the cluster and between Fe and S$_{cys}$ (the sulfur atom present in the cluster-coordinating cysteine residues) of each binding cysteine. Such distances were used to create restraints that applied a mild force of 5 Kcal·mol^{-1}Å^{-2} for the whole simulation and avoid collapse of the FeS cluster.

3.4.2   Analysis of the oxygen-sensitive *Cs*LAM MD simulations

A detailed study of the oxygen sensitive LAM starting from its crystal structure was conducted. Simulations of 400 ns were obtained for the tetrameric structure of *Cs*LAM

containing either a [Fe$_4$-S$_4$] cluster coordinated by three cysteine residues in each active site, or both [Fe$_4$-S$_4$] cluster and SAM molecules coordinated in each monomer for a total a four cofactors per LAM enzyme.

The root-mean square deviation (RMSD) and root-mean square fluctuations (RMSF) were calculated for each individual monomer of the LAM tetrameric structure compared to their initial structure and the average one. RMSD compared to the crystal structure before the MD simulations was used to assess the simulations and whether they reached convergence during the 400 ns time scale and they are shown in Figure 24. Plots a) and b) showed the RMSD compared to the starting structure of *Cs*LAM without SAM and with SAM included in the active site respectively. The RMSD indicated the MD simulation when SAM was present to be ~1 Å lower than the average value of that without SAM, until the last half of the simulation where
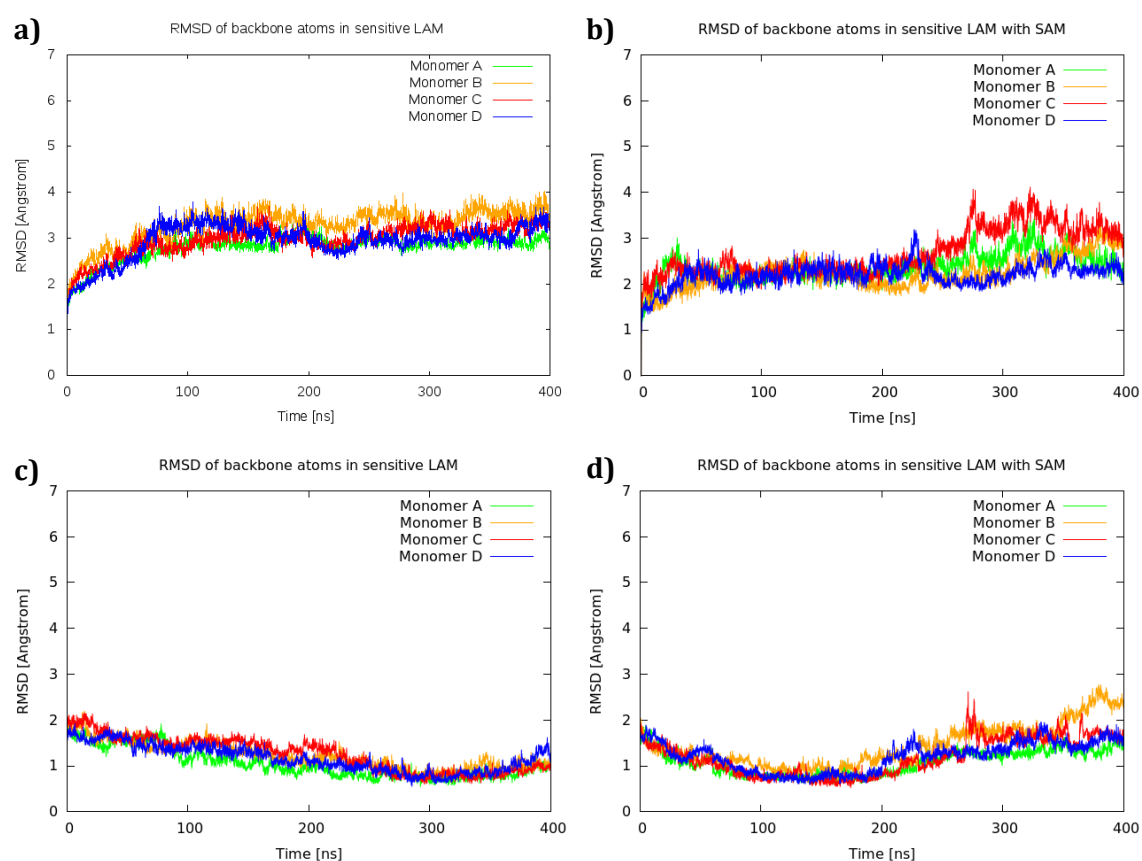
Figure 24. RMSD compared to the starting structure of the produced 400 ns MD simulation of *Cs*LAM without SAM and with SAM, a) and b) respectively. Plots c) and d) reported the RMSD compared to the average structure without SAM (300 – 400 ns) and with SAM (100 – 200 ns) in the catalytic pocket.

113

monomer C (plot 24b) increased its deviation up to 4 Å (~1 Å higher than the other monomers). Plots 24c reported the RMSD compared to the average structure (300 – 400 ns) of the sensitive LAM without the SAM molecule contained whilst 24d the RMSD compared to the average structure between 100 – 200 ns when SAM was in. It was possible to notice again that the SAM-containing system reached RMSD values slightly lower than the SAM-free structure except for the last 50 ns of the MD simulation where monomer B assumed a deviation of 0.5 Å higher than the rest of the enzyme structure. No remarkable structural instabilities were reported in both the enzyme systems.

In figure 25 the RMSF showed the fluctuation of the 409 residues composing the monomeric structures of the SAM-free (25a) and the SAM-containing (25b) enzyme compared to their respective average structure. The RMSF showed results that were consistent with what RMSD reported. RMSF of monomers B and C (Fig. 25a) outlined the highest fluctuation (up to 4 Å) on residues around positions 100, 140, 320 and 340. The same trend could be observed for monomer B when SAM was included in the LAM (Fig. 25b), with positions 100 and 140 showing a high fluctuation, while the C-terminus looked more stable under these conditions. In both cases the cluster-bridging loop (residues 123 – 130) and the SAM-coordinating loops (residues 195 – 201, 225 – 230) did not experience significant fluctuation; the loop containing Lys337, which is involved in the PLP coordination, showed to have similar variation in both
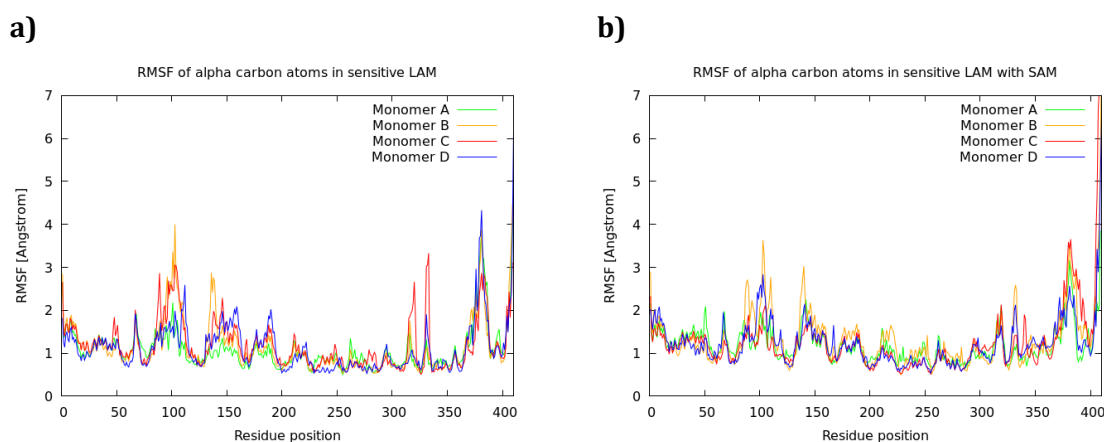
a)   **b)**

Figure 25. RMSF compared to the average structure of *Cs*LAM without SAM (a) and with SAM (b), in each pocket. The colours adopted for each monomer are shown in the legend in the top right corner.

LAM systems, reporting monomer C as the outlier of the unbound-SAM system and monomer B with the highest fluctuation. The loop that contained the cysteine residues involved in the zinc coordination (residues 375 – 380) did not show remarkable fluctuations either with or without SAM, as an indication that the area of the enzyme seems not to be affected by the conditions of the active site.

Differences between the two simulations are evident depending on whether the SAM molecule is either absent in the catalytic pocket (Figures 24a, c and 25a) or present (Figures 24b, d and 25b), and vary for each monomer. A general slight instability was observed when the SAM cofactor was not present in the cavity to coordinate the metal cluster. This suggests that the SAM-containing structures are more similar to those observed in the crystal structure and that SAM plays a role in helping to maintain this conformation. Previous works from the literature reported the motion of the helix interacting with the carboxylate moiety of homocysteine upon binding in the active site of methionine synthase[205] and that is not propagated to the neighbouring domains. Different conformations were also observed in the sactionine bond-forming enzyme CteB depending on the presence of SAM in the active site containing the $[Fe_4$-$S_4]$ cluster.[206] The Bateman module in Cystathionine β-synthase[207] was observed undergoing a migration on top of the catalytic site upon binding of the SAM molecule and other important conformational changes were observed in the threonine synthase.[208] In the transcription factor B protein that is homologous to ribosomal RNA methyltransferases, small structural differences were observed between the *apo*-form and the SAM-bound enzymes.[209] In the peptide-modifying radical SAM enzyme SuiB it was reported a low residue fluctuation in the SAM-bound protein that led to loops fluctuation upon substrate binding.[210] The residues in the enzyme without SAM fluctuate more than with SAM, consistent with both the RMSD and RMSF results. The residues just before and after the cluster-bridging loop are the most fluctuating when SAM is not in the catalytic pocket. The areas between residues 90 to 110 and 140 to 160 described two different groups of loops and peripheral helices that directly face the active site and SAM, and the fluctuation decrease on SAM binding suggesting

that those loops may be involved in the SAM-cofactor stabilisation. That is consistent with the residues involved in the SAM coordination as it was shown previously in Figure 11. Particularly, the fluctuation of those loops that are close to the N-terminus of the enzyme (residues 90 – 110) resemble the loop motion that was observed in the enzyme SuiB where the N-terminal adopted an archetypal RRE domain.[210] In the case of the SuiB enzyme, the RRE domain is composed of three-stranded antiparallel β-sheets that sits next to a four helical bundle that forms a sort of wHTH motif that protrudes from the catalytic core (Figure 26a). The *Cs*LAM MD simulations indicated a wider fluctuation over the N-terminal extension just above the TIM barrel when SAM is not bound. The loop formed by residues 90 – 110 (Figure 26b, in red) in *Cs*LAM oscillates just above the $[Fe_4\text{-}S_4]$ cluster, over the TIM barrel of the enzyme. This fluctuation is reduced when SAM is bound to the FeS cluster. Such a coil that is linked to a four-helical bundle might play a role in the substrate recognition of the lysine 2,3-aminomutase. Residue fluctuation is more noticeable for those moieties that lie at the C-terminal part of each monomer (residues 370 – 409) and again at the supposed PLP-
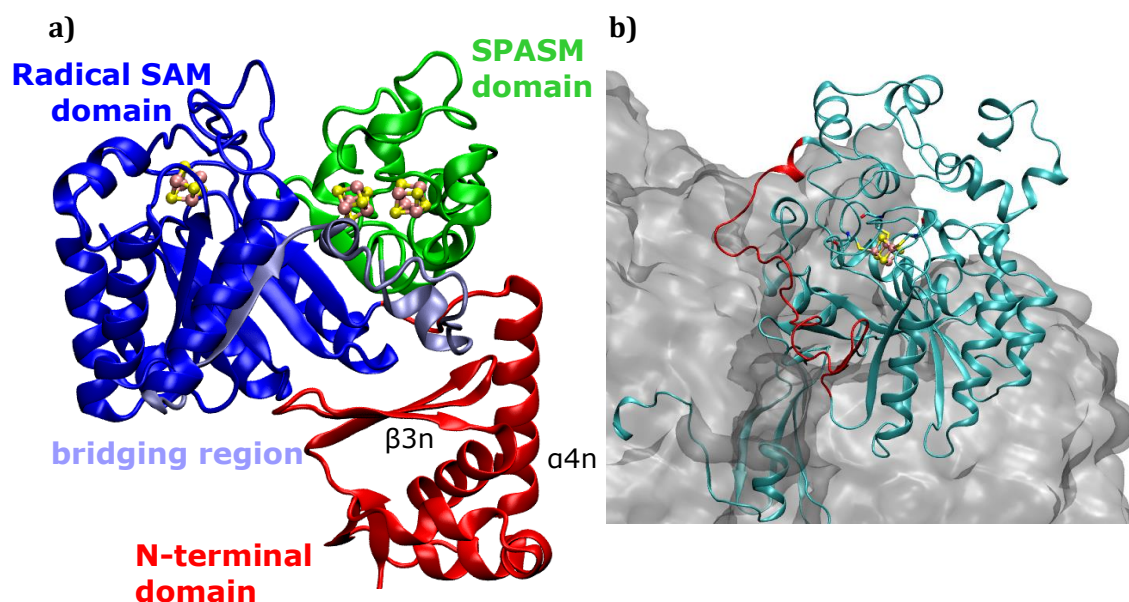
Figure 26. The structure of radical SAM enzyme SuiB with highlighted N-terminal domain and random coil associated to the substrate recognition domain (a), image adapted from Davis et al[209] Monomer A of unbound-SAM CsLAM enzyme (cyan) with highlighted wobbling coil (red) that stands between the TIM barrel and the four-helical N-terminal loop which resembled the SuiB structure involved in the substrate recognition (b).

coordinating loop (residues 341 – 346), as it was shown for the SAM-free simulation (Figure 25a). Residues 123 – 130 appear to be more stable when SAM is bound and generally just monomer B seems to have residues with the slightly higher mobility (Figure 25b). The RMSF plots showed an increased flexibility when SAM was absent from the enzyme structure, mainly in those regions of high fluctuation of the SAM-bound system.

Fluctuations in the absence of SAM raise the possibility that this increased mobility is present to facilitate the access of the cofactors to their binding sites. The consistently least stable monomer (monomer B) in both SAM-absent and SAM-present simulations indicates that the protonation states chosen for these histidine residues (HIE131 and HIP230) may not be ideal and are unlikely to be the native state, as they might create an overall destabilisation of the structure. By Looking at the measured distances between the N-epsilon atoms of His131 and His230 and distance between the N-epsilon atom of His230 and the C-zeta atom of Arg200 we could exclude any issues with stability of the binding site arising from different histidine protonation states (see Appendix 6.10). A different visualisation of the motions driving the conformational dynamics in the two tetrameric structures is given by the PCA (Figure 27, the unbound-SAM system is depicted in blue whilst the bound-SAM is in orange). The first two collective modes of both simulations were projected on a common subspace and it gives the opportunity to look at how similar the two simulations were. The first two PC on a common subspace show the main conformational states of the enzymes that were driven by the most frequent structure motions detected in the MD simulations. Figure 27 shows the most important and slowest motions that described one mainly populated conformational state for each simulation. The whole tetramer was analysed and, consistent with previous data, the overall fluctuation is quite similar between the two simulations with bound and unbound SAM. The conformational MD sampling of the lysine 2,3-aminomutase with SAM bound to the FeS cluster is shown in Figure 27 (in orange). The basin in the lower left-hand side of the map describes instead the main, highly-populated state when SAM was absent from the system (in blue). The simulation without SAM sampled an area of the PC1/PC2 space that was not

sampled in the SAM-bound simulation as expected from two differently cofactor-containing simulations. The blue basin represented the major conformational state sampled during the simulation without SAM in the active site and it is visible the bigger area it described compared to the SAM-bound PC sampling. That was consistent with the general larger fluctuation observed in the MD simulation of *Cs*LAM without SAM in the active site that might have caused a wider loop fluctuation. This variation might describe the LAM structure without SAM as fluctuating between two conformational states but by looking at the same PCs over time, we could exclude the presence of different conformations sampled but only one largely fluctuating instead (Appendix 6.11). Although, it might describe the behaviour of LAM when only the FeS cluster is present and it expects to bind the SAM cofactor from the environment, hence the more substantial fluctuation. This analysis confirms the higher loop fluctuation that the absence of SAM could have contributed to the less ordered motion,
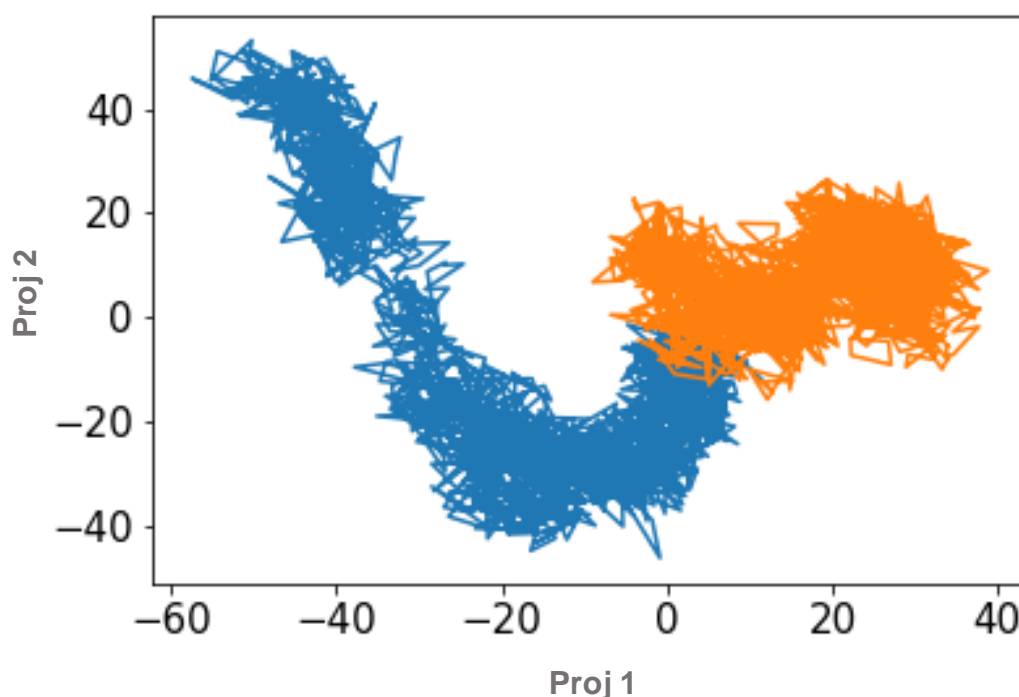
Figure 27. 2D map of the overall projections of the first and second principal components (Proj1 and Proj2 respectively) in both *Cs*LAM simulated systems. The basins relative to both the MD simulations (bound-SAM in orange, unbound-SAM in blue) defined two different parts of the PC1-PC2 subspace. The blue basin covered a bigger area, consistent with the larger general fluctuation observed in the unbound-SAM MD simulation.

consistent with the outcome of the RMSD/RMSF analysis. The motions represented by PC1 and PC2 were represented by porcupine plots in Appendix 6.12. Monomers were coloured following the colour code showed for RMSD and RMSF and each arrow described the direction of alpha-carbon atoms composing the LAM structures that was sampled by PCA. Once again it was visible the larger fluctuation affected the C-termini in both MD simulations with larger arrows, i.e. fluctuation, in case of the unbound-SAM LAM system.

For the oxygen sensitive LAM simulation containing the SAM cofactor, the correct orientation and coordination of the SAM molecule toward the FeS cluster over the MD time scale were checked for consistency with coordination distances measured in the crystal structure of the lysine 2,3-aminomutase (Figure 28). The distances on the starting structures were visualised
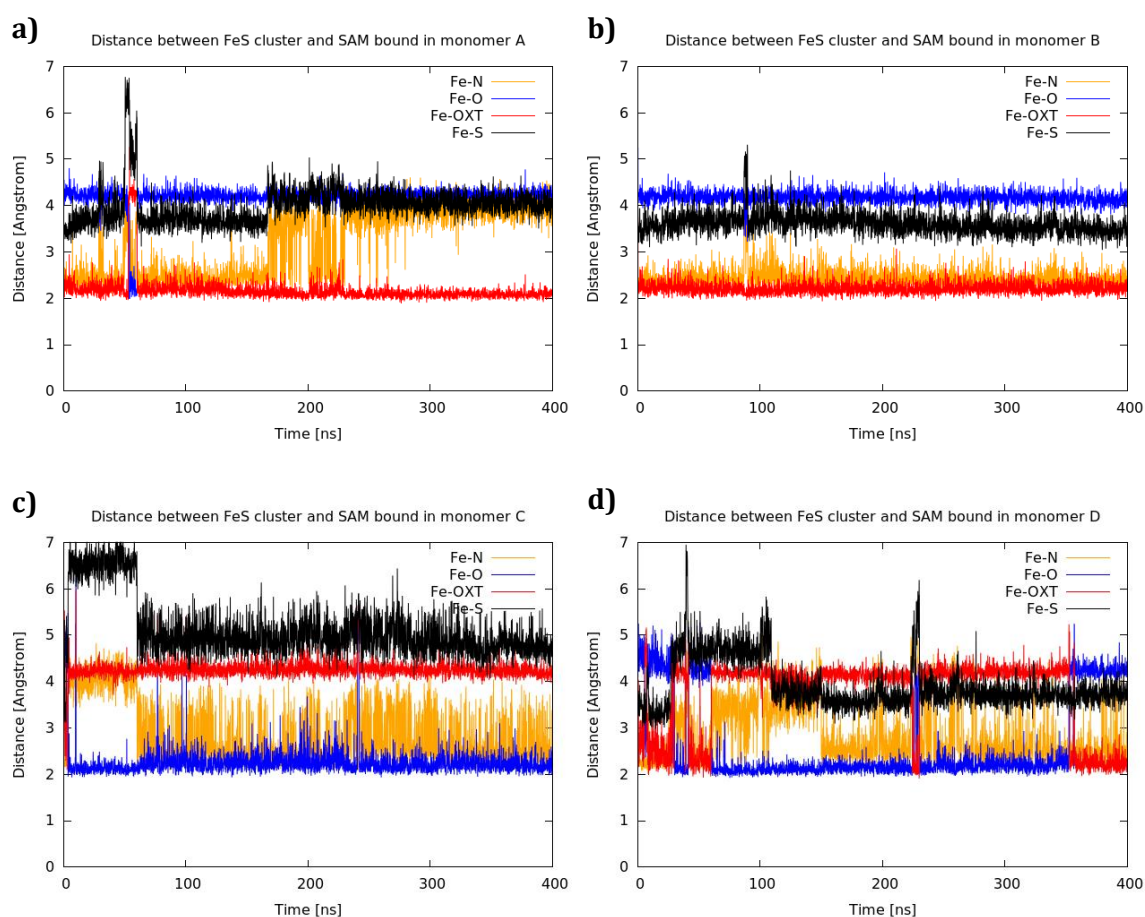
Figure 28. Average distances between key atoms in SAM and unique Fe atoms in the crystal structure. Measured distances from unique Fe and amine nitrogen (orange) and carboxylate oxygen atoms (red and blue) in the four monomers of LAM.

119

and measured with VMD and were taken between the unique cluster iron (Fe) and the carboxylate oxygen atoms (O, OXT), the amine nitrogen (N) and the sulfur atom (S) of the methionine moiety of SAM throughout the four monomers. An average bond distance of ~2.38 Å was measured for the coordination to the unique Fe atom from the amine (N) and the closest-to-cluster carboxylate oxygen (O), consistent with other results reported.[95] The second carboxylate oxygen atom (OXT) is further from the unique iron by ~4 Å as it does not coordinate to the [Fe_4S_4] cluster and the S atom binds to the metal cluster at a distance of 3.25 Å consistent with previous results.[92] The overall binding distances were measured during the length of the MD simulations and were consistent with those from the crystal structure, although monomers C and D (figure 28c and 28d respectively) showed that a rotation between the two oxygen atoms may occur and the sulfur atom in monomer C may be found at some distance from the metal cluster. Monomer A (fig. 28a) showed stable coordination of SAM despite the small shift of the N atom from the FeS cluster by ~2 Å (at ~190ns), which drove the SAM molecule a bit further from the metal cluster. Monomer B (fig. 28b), on the other hand, showed a stable coordination of the SAM cofactor for the whole simulation. Perhaps the fluctuation of the different monomers influenced the SAM coordination or the slightly higher RMSF of monomer B may indicate that, after the correct SAM coordination, the enzyme was ready to bind substrate and cofactor by an increased loop fluctuation that perhaps created cavities to let these molecules in.

3.4.3 Oxygen tolerant lysine 2,3-aminomutase *Bs*LAM MD simulation settings

3.4.3.1 The build-up of the tetramer of the homology model and cluster insertion

I-TASSER provided the homology model for an individual monomer of *Bs*LAM, and this was used to build up the tetrameric structure using the visualisation software UCSF Chimera. An RMS-fit of each monomer to the crystal structure template assured the correct superposition between monomers providing the final tetrameric structure of the *Bs*LAM. Once the final

tetrameric form was achieved, it was possible to fit the [Fe$_4$-S$_4$] clusters in each cavity of the homology model based on the crystal structure template, with a minimal adjustment of their orientation toward the cluster-binding cysteine residues. The structure was exported and submitted to the same checks of the *Cs*LAM structure; namely a check of the protonation states by uploading it to H++ and refinement of the final structure. *Bs*LAM did not contain the residue His131 close to the [Fe$_4$-S$_4$] cluster as seen in *Cs*LAM but Tyr131 instead, therefore no customised protonation states were applied and every monomer was considered equivalent. The final structure was edited accordingly to what the H++ analysis reported. Residues ASP276 and GLU378 were considered in their neutral form and HIS11, 97, 120, 202 and 436 were all set up in their double charged HIP form, while all the other histidine residues were set in the HIE form. CYS134, 138 and 141 were the [Fe$_4$-S$_4$]-coordinating cysteine amino acids that were edited to CYF residues to apply the correct MD force field.

Mild distance restraints were applied to the FeS cluster inner bonds and the coordinated cysteine sulfur atoms, in line with the procedure applied to *Cs*LAM. The bond distances were checked to be consistent with the 2A5H crystal structure cluster bonds and restraints were applied for the whole simulation at a mild constant force of 5 Kcal/molÅ.

3.4.3.2 Molecular docking of SAM cofactor in the homology model

Superposition of the SAM from the crystal structure to afford equivalent SAM-bound *Bs*LAM, as per the procedure for the FeS clusters, produced a poor quality model that revealed having several issues during MD, thus SAM was inserted through molecular docking instead. The tetrameric starting structure of the *Bs*LAM containing the [Fe$_4$-S$_4$] clusters was extracted from the SAM-free *Bs*LAM MD simulation after 100 ns to ensure a relaxed starting structure. Previous attempts of docking the SAM cofactor from the original starting structure lead to many stabilisation issues over the MD simulation. The tetrameric structure of the homology model *Bs*LAM was gathered together starting from the monomeric structure, then overlaid to the *Cs*LAM crystal structure. The final built-up tetramer of the homology model needed longer

to stabilise and the presence of SAM in the active site must have contributed to the minimisation issues, thus to necessarily rely on a different and relaxed starting structure.

All water molecules and ions were deleted, with formal charges and bond orders already assigned from the MD simulation structure. The structure of the SAM molecules was extracted from the crystal structure of *Cs*LAM (2A5H) and it was placed into the relaxed *Bs*LAM structure containing the $[Fe_4-S_4]$ clusters. The homology model and SAM molecules were treated by the Ligand and Protein Preparation Wizard workflow[211] prior to docking and hydrogen atoms were added to SAM. An overall charge of 0 was used for the SAM zwitterion, consisting of a positive charge on the sulfonium moiety and a negative charge on the carboxylate. A rigid Glide Standard Precision (SP)[212] ligand docking was then performed and the best structures were selected for each cavity, based on both the consistent orientation of SAM relative to other radical SAM crystal structures and Glide scores. Receptor grids were calculated to look at the possible interactions of SAM in the different host catalytic pockets. An OPLS3 force field[213] was applied with default parameters for the Van der Waals scaling factor of 1.00 and a charge cutoff of 0.25 for the protein environment, whilst for ligand atoms the default parameters were set at 0.80 and 0.15, respectively. A cubic box centred on the Fe atom that SAM binds to, was generated for each receptor pocket. The Fe atom of the cluster that faces and coordinates to SAM was used to create a hydrogen-bond/metal constraint to keep SAM oriented in front of the cluster in the right position. Each of the 40 SAM energy-minimised poses per pocket were scored on energy-minimised structures and those with the lowest Glide score value was selected as the best docked position. The energy values for the SAM molecule in each monomer varied from -7.2 kcal·mol^{-1} (monomer A) to -5.1 kcal·mol^{-1} (monomer D), despite highly similar coordination to the FeS. This variation was attributed to small differences in the conformations of the active site. When all the SAM molecules were docked into the enzyme structure and the overall energy minimised, the resultant PDB file was processed by *xleap* to create the necessary parameters and structural files to start the MD simulation of *Bs*LAM containing FeS clusters and SAM.

3.4.4   Analysis of the oxygen-tolerant *Bs*LAM MD simulations

Simulations of *Bs*LAM were performed for 400 ns on [Fe$_4$-S$_4$]-containing structures both with and without SAM bound. The simulations were run using the same protocol applied to the crystal structure MD simulation of *Cs*LAM. In general, the enzyme systems seemed to have reached higher deviations from their starting structures (Figures 29a and 29b) compared to the previous oxygen-sensitive MD simulations. The SAM-containing simulation did not deviate from the reference structure as much as the SAM-free enzyme system and they showed an average difference in RMSD by ~1 Å between the two, consistent with the outcome arisen from *Cs*LAM simulations. It could be noticed a sudden higher deviation by ~2 Å in monomer C of the system containing SAM in the last 100 ns MD simulation. The RMSD values
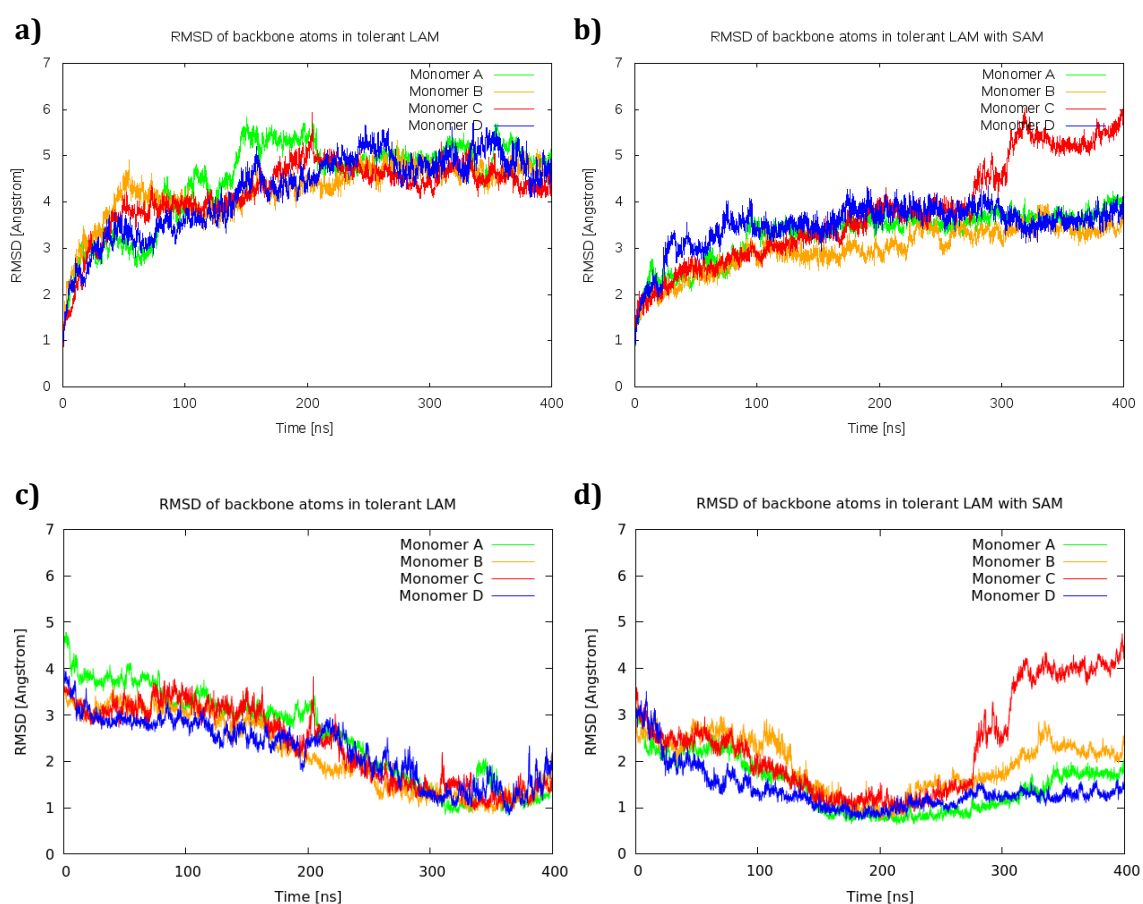
Figure 29. RMSD compared to the starting structure of 400 ns MD simulation of *Bs*LAM without SAM and with SAM, a) and b) respectively. Plots c) and d) reported the RMSD compared to the average structure without SAM taken between 300 – 400 ns and with SAM between 200 – 300 ns.

were a bit higher than those observed in the case of *Cs*LAM but they were <6 Å so the structure was not considered a non-native one.[214] Furthermore, the tolerant LAM structure is a homology model and, despite being built on the *Cs*LAM template, it was predictable to see a higher distance from the template as the model is technically a different structure.[164] The variance between *Cs*LAM and *Bs*LAM RMSD is consistent with that recorded between THIC crystal structure and the homology model, as shown previously in the chapter. The RMSD was also measured compared to the average structure taken between 300 – 400 ns in case of the SAM-free system and between 200 – 300 ns for the SAM-containing enzyme, figures 29c and 29d respectively. Once again, the SAM-free system reached RMSD values higher by an average of ~1 Å than the SAM-bound structure but no particular instabilities were reported. Figure 29c showed that the rising in RMSD of monomer A that was visible in figure 29a did not represent a significant deviation from the average structure and therefore it was not considered an issue. Figure 29d reported instead the same deviation that was observed in figure 29b in the last ~80 ns of monomer C that assumed 2 Å higher values than the rest of the enzyme structure.

A view of the RMS-fluctuation of each C$^\alpha$ atom composing the *Bs*LAM enzymes is showed in Figure 30, compared to their own average structure. Consistent with the RMSD results, the

a)   b)
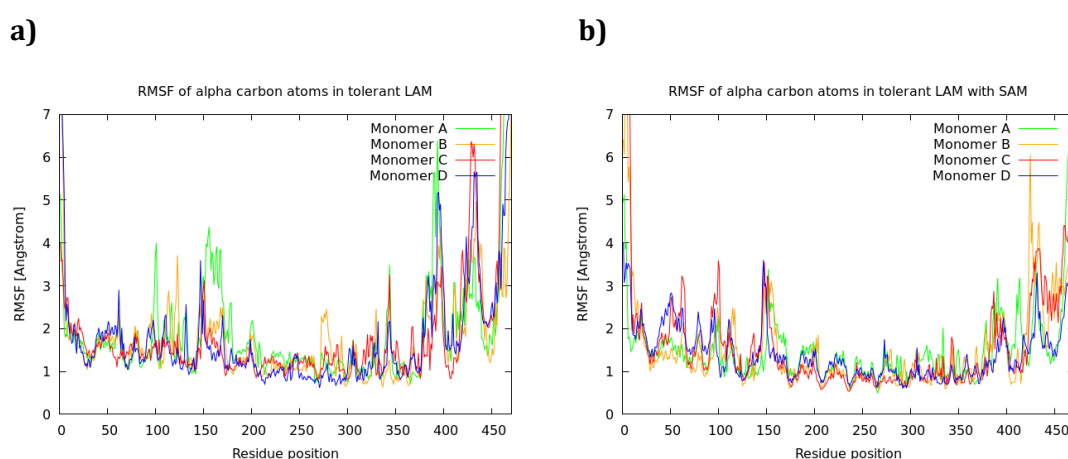
Figure 30. RMSF compared to the average structure of the oxygen-tolerant *Bs*LAM without SAM (a) and with SAM (b) in each pocket. The colours adopted for each monomer are shown in the legend on the top right corner.

fluctuation of the SAM-bound system looked moderate compared to the SAM-free structure that showed peaks at positions 120, 160, 370 and 430. The RMSD of monomer A in the unbound-SAM system showed in Figure 29a outlined a deviation between 100 – 200 ns, that seemed to be consistent with the fluctuation of those residues present just after the cluster-binding loop (residues 150 – 170). Analysis of RMSF between 0 – 100 ns and 200 – 400 ns were conducted using the starting structure as reference to have a better look at the time intervals where the average structure would lead to unreliable outcomes. As showed in Appendix 6.13, plots a) and b) highlighted a decreased fluctuation of the area, confirming the high RMSF values were due to the perturbation between 150 – 200 ns. The binding-cysteine loop residues at position 134, 138, 141 showed the lowest fluctuation (<1.5 Å) of the area of interest of the enzyme, as well as the SAM-coordinating residues on the β4 (residues 204 – 209) and β5 loop (residues 234 – 239), which did not fluctuate over 1 Å consistent with the results observed for *Cs*LAM. The β4 and β5 loops contain the residues that interact with SAM when is bound to the active site. Fluctuation in the absence of SAM was ~1 Å, slightly higher than in the case of the SAM-bound simulation. Moving toward the C-terminus of the monomeric chains, less stable areas of the enzyme could be observed in both simulations. *Bs*LAM, in contrast to the clostridial counter enzyme, lacks the zinc-coordinating cysteine residues, and a high fluctuation of the terminal area between residues 380 – 400 could be observed. The C-terminus ~60 residues-long loop is only present in the oxygen-tolerant *Bs*LAM therefore it could not be created through the homology modelling technique due to the lack of templates. I-TASSER did create the terminal loop from scratch using the *ab initio* methods, thus some fluctuation is expected in this region.

By looking at the RMSD of the *Bs*LAM MD simulations without considering the last 110 amino acid residues (the C-terminus) it was possible to notice a general decrease in their deviation by ~2 Å in the case of the SAM-free structure and by 1 Å when SAM was in (figures (c) and (d) of Appendix 6.13). Interestingly the monomer C (in red) of the SAM-containing enzyme structure still showed a pronounced deviation in the last 80 ns of the MD simulation: that was

due to another part of the enzyme that contributed to the increased deviation. By the RMSF in figure 30 it was noticed that the first residues contained in the N-terminus showed a high fluctuation and new RMSD and RMSF were measured for the SAM-bound MD simulation without considering the first 20 residues (plots e) and f) of Appendix 6.13). The RMSD of the SAM-bound *Bs*LAM looked stable without increased variations in the final part of the MD simulation confirming a contribution to the enzyme motion by few residues at the N-terminus. Since those residues did not cover important roles in the enzyme, their contribution to the whole enzyme dynamics was not considered relevant. The newly measured RMSF also showed a general decrease in fluctuation by ~0.5 A, consistent with the removed contribution of the disordered N-loop. In Figure 31 is reported in two snapshots retrieved from the SAM-bound *Bs*LAM MD simulation the fluctuation of the N-terminus in the C monomer of the enzyme. On the left of the figure is detailed the loop in a position that is close to the enzyme structure and such a position was kept for the majority of the simulation. On the right of Fig.
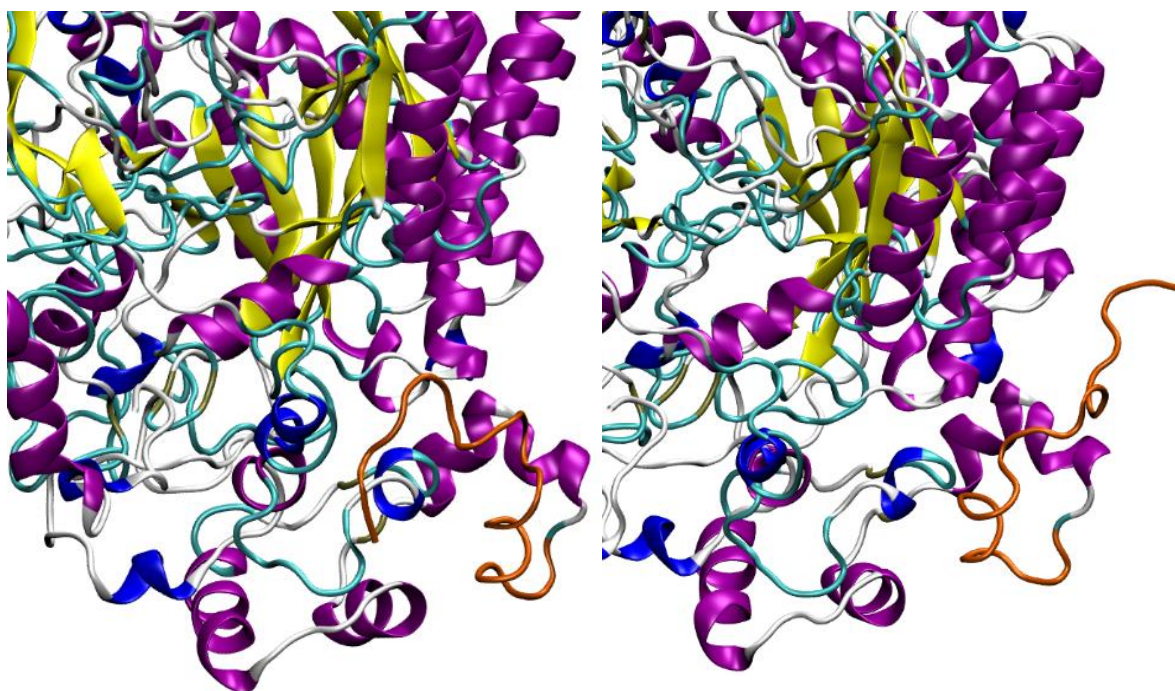
Figure 31. Fluctuation of the 20 N-terminal residues of monomer C of SAM-bound *Bs*LAM MD simulation, highlighted in orange. On the left is reported the normal position of the loop whilst on the right, the loop in the ending part of the simulation whilst it assumed the larger deviation.

31 the loop got further away from the protein and that appeared as an increased RMSD in the last 100 – 80 ns of the MD simulation.

All the monomers composing the oxygen-stable LAM showed an ordered dynamics and low fluctuation at the FeS-cluster loop level (0.9 Å), as well as the other SAM-interacting β-sheets showed a general low fluctuation (0.6 – 0.9 Å). That was consistent with the good coordination of the SAM molecule thus leading to a more ordered dynamics. The last C-terminal part of the monomers in the SAM-containing enzyme showed less fluctuation the MD simulation without SAM in the active site (Figures 29a and 29c). Nonetheless, although the terminal loop was generated by using *ab-initio* method in I-TASSER, it showed a cleaner dynamics when SAM was present in the enzyme, a behaviour that was already observed in the MD simulations of the oxygen-sensitive LAM (section 3.4.2). Perhaps the motion of the C-terminal loop in the lysine 2,3-aminomutase is involved with the recognition of the substrate and cofactor that may be present in the environment and the higher fluctuation may be due to the recognition and insertion of molecules into the active site. These fluctuations might also be involved with the detection and recognition of the reducing protein that activates LAM upon the FeS-cluster reduction by the donated electron of another enzyme *in vivo*.

The deviation of the small PGGGK loop that contains the amino acid residue Lys346 involved in binding to the PLP cofactor when the substrate is not present in the active site,[85] might be due to the absence of the PLP molecule within the enzyme structure. The deviation was noticed in the *Cs*LAM and it is consistent with the lack of coordination of such a cofactor by the enzymes.

The simulation with SAM docked into the enzyme structure looked more stable than the SAM-free system. By their RMSD and RMSF we could state that the SAM-bound simulation was generally less fluctuating with the C-terminus as the most mobile area of the enzyme. Monomer C also received a contribution to its dynamics by the first few residues in the N-terminus but they did not account for crucial amino acids. The unbound-SAM simulation showed a higher fluctuation over the same areas of the SAM-docked structure, and the

deviation from its starting structure was also due to the perturbation between ~130 – 200 ns but remarkable structural instabilities were not noticed.

Principal component analysis gave a different insight about the dynamics of the whole tetrameric structure in the two systems (Figure 32). The projections of the first two collective modes on a common subspace was considered as well as it was done for *Cs*LAM. The left side of the free-energy landscape (figure 32, in blue) described a large area that represented the most important conformational state accessed during the MD simulation, the dynamics of LAM without SAM in its cavities. That means that the enzyme fluctuation during the simulation have driven the dynamics of the protein, describing one large and quite disordered conformational state, compared to the orange area that described the energy landscape of the SAM-bound system. That means that as well as it was observed for *Cs*LAM, the two simulations sampled a different area of the PC1/PC2 space, i.e., they have not converged to the same microstate, consistent with simulations containing different cofactors, that describe
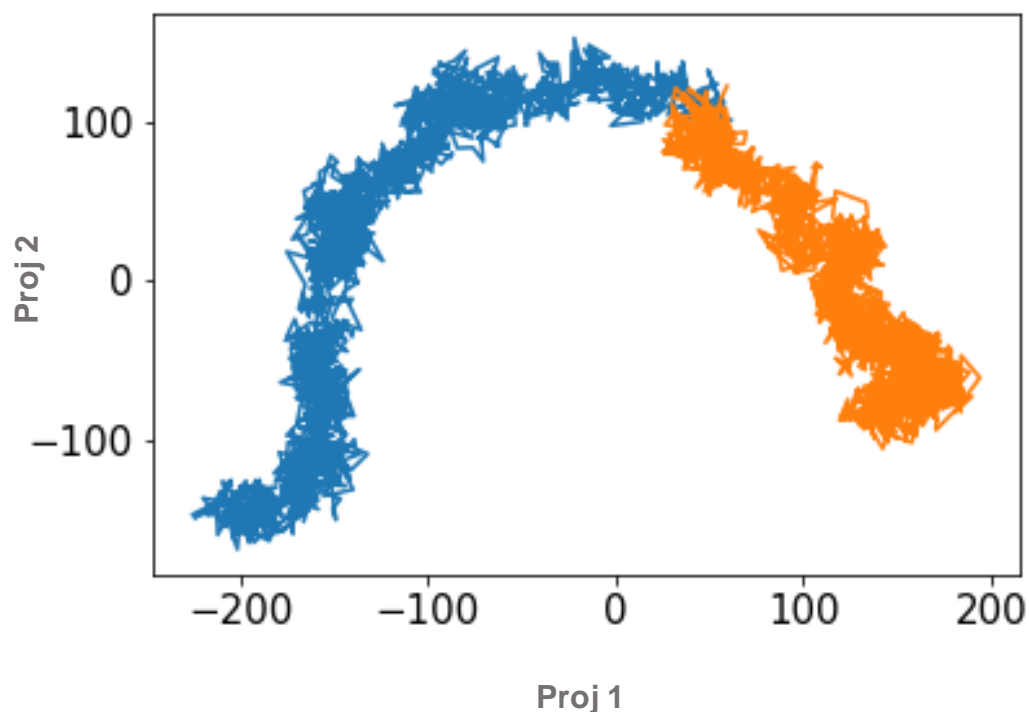
Figure 32. 2D map of the overall projections of first and second principle components (PC) in both the simulated systems. *Bs*LAM with SAM is on the right while on the left the two signals are from the unbound-SAM simulations.

formally two different MD simulations. Just like in *Cs*LAM, the main conformational state of both the simulations are closer along PC2 than along PC1, consistent with results reported previously for *Cs*LAM.

On the right part of the plot in figure 32, in orange, a significant region representing the PC of the *Bs*LAM with docked SAM could be sampled. As it was already observed by its RMSD, the MD simulation showed a flat trend in the deviation of its monomers describing a more stable and less fluctuating enzyme dynamics. The large basin represented the main conformational state populated by the enzyme during the simulation. The smaller area of the basin might be related with the sudden deviation of monomer C at the end of the simulation that could describe a different motion. The lower fluctuation of residues resulted in the inner motions of the enzyme being similar in all monomers, thus resulting in one mainly populated conformational state. It seemed like the LAM structure without SAM, fluctuated between more conformational states, as it was previously observed in *Cs*LAM, but by looking at the same PC subspace over time (Appendix 6.14), we could exclude the presence of other conformational states accessed by the enzyme, consistent with results we observed previously in the case of *Cs*LAM. Once again, that might have described the behaviour of LAM when the [Fe_4–S_4] cluster is the only cofactor present in the active site. This analysis confirms the higher loop fluctuation that the absence of SAM could have contributed to the less ordered motion, consistent with the outcome of the RMSD/RMSF analysis showed previously in the paragraph**.** The motions represented by PC1 and PC2 were represented by porcupine plots in Appendix 6.15. Monomers were coloured following the colour code showed for RMSD and RMSF and each arrow described the direction of alpha-carbon atoms composing the LAM structures that was sampled by PCA. Once again it was visible the larger fluctuation affected the C-termini in both MD simulations with larger arrows, i.e. fluctuation, particularly in case of the unbound-SAM *Bs*LAM system. The orientation and coordination of the SAM molecule toward the FeS cluster in the SAM-bound *Bs*LAM simulation was checked by measuring and comparing with coordination distances measured in the crystal structure of LAM. The

distances between the unique cluster Iron atom (Fe) and the carboxylate oxygen atoms (O, OXT), the amine nitrogen (N) and the sulfur atom (S) of the methionine moiety of SAM in the four monomers were measured for the entire simulation (Figure 33). A difference in the coordination throughout the monomeric structures was observed, with monomer A showing the loosest coordination with long bond distances. Here it seems as if SAM had dissociated from the FeS cluster during the simulation. The RMSF showed previously that the A monomer was one of the highest fluctuating in the SAM-interacting loops but one of the most stable from the RMSD. This discrepancy indicated that* a less reliable coordination between SAM and FeS cluster was established in this monomer (Figure 33a). The distances between SAM and the metal cluster were as twice as those detected in the crystal structure MD simulation. Another unfixed coordination was observed in the C monomer (Figure 33c), where an initial
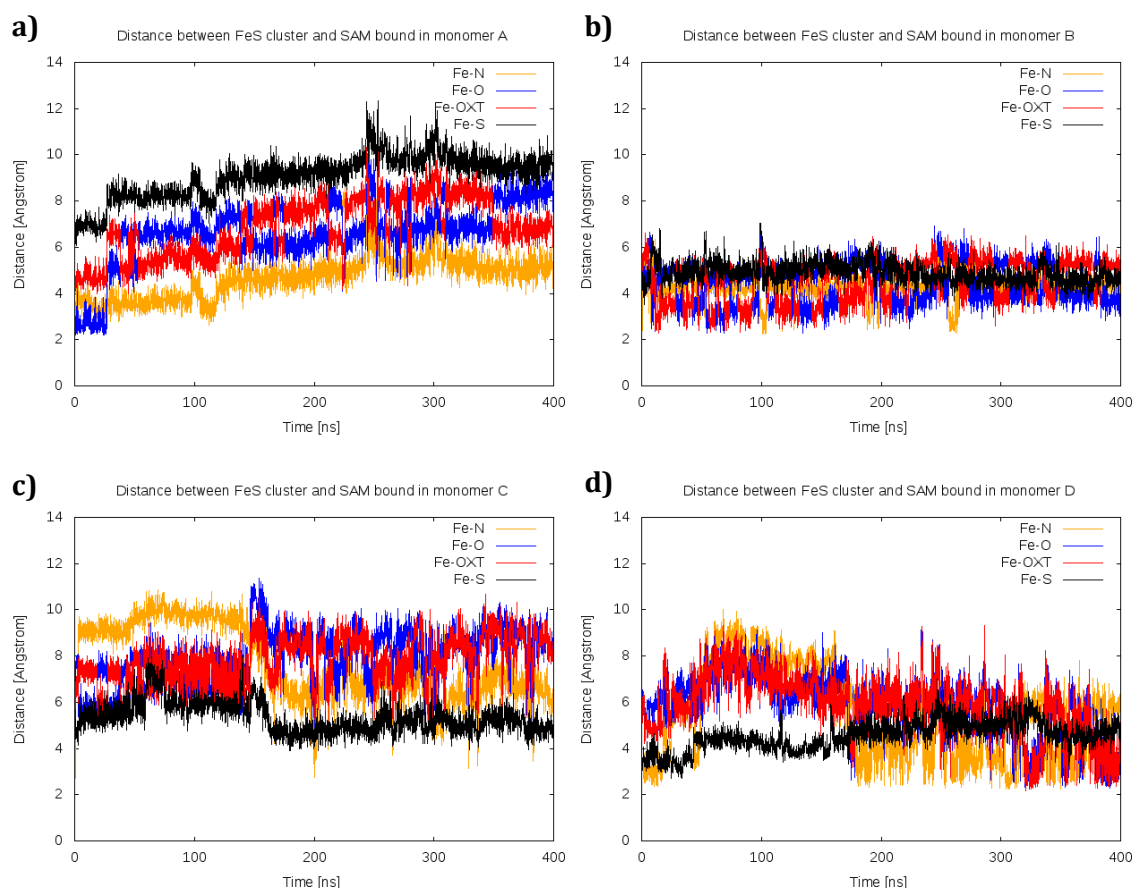
Figure 33. Coordination distances between SAM and [Fe$_4$-S$_4$] cluster atoms in the four monomers in docked SAM MD simulation over the four monomers (a - d). Colours are as previously shown in section 3.4.2.

rotation had swapped the amide and the carboxylate moiety keeping the molecule of SAM at a distance 2 Å further than it was in the crystal structure MD simulation. In the B and D monomers (Figures 33b and 33d, respectively), the SAM coordination was more stable without sudden, dramatic shifts of the SAM molecule during the length of the simulation and with the average coordination distance in the range of those from the crystal structure. In every monomer, the distance between the amine nitrogen (N) and the Fe atom was always in the range of 2 – 6 Å consistent with *Cs*LAM bond lengths previously observed and literature precedents.[92]

By comparing the MD simulation of the crystallographically-derived *Cs*LAM with the homology-modelled *Bs*LAM, the absence of SAM leads to destabilisation in the individual monomers, consistent with visible fluctuation of those loops that may be involved in the access of the cofactor and substrate to the active site. The SAM coordination in front of the FeS cluster was similar in the two LAM enzymes, although docking of SAM appeared to lead to a slightly further coordination-distance from the [Fe$_4$-S$_4$] cluster. Regardless, the SAM molecule did not dissociate over the timeframe of the simulation nor was any remarkable destabilisation detected. The dynamics of the homology model simulations were consistent with the crystal structure simulations, where both showed destabilisation when SAM was not included in the active site as well as similar fluctuating residues in both enzymes. When SAM is not in the enzymes the structure explored two conformational states, due to less structural stability compared to when SAM was not included. The SAM coordinated structures were consistent and although the SAM was further from the FeS cluster in the docked *Bs*LAM compared to the *Cs*LAM structure, it remained in the active site and correctly coordinated. The homology model of *Bs*LAM could then be considered reliable on the TIM-barrel structure with both SAM- and [Fe$_4$S$_4$]-coordinating residues that made an accurate enzyme active site, consistent with the crystal structure. Less accurate was reported to be the SAM-[Fe$_4$S$_4$] coordination since only monomer B and D showed no remarkable disruption for most of the MD time length, however, SAM molecules were still coordinating the cuboidal cluster and did

not leave the active site. Structural issues arose in the C-terminus area of the enzyme that was predicted by the computational approach adopted by the webserver during the homology model creation for lack of templates. As it was expected, the final part of the monomers composing the homology model of *Bs*LAM showed a high fluctuation compared to the rest of the structure however that did not form an important part of the enzyme for the purpose of our study. Nonetheless, further studies on the lysine 2,3-aminomutase family may help ruling out the function of the longer C-terminal part that in the *Cs*LAM is missing.

From the consistency between the simulations in terms of general behaviour, and the previous structural comparisons showing homology model production with as good quality of parameters as extant crystal structures, the approach outlined is considered sufficient to produce a model from which simulations could be considered valid. On this basis, the MD simulations of the two enzymes were used to investigate trends and make qualitative observations on the key features that might confer oxygen stability in LAM. Particularly, it seems from the MD simulations, that monomers **B** and **D** were the most reliable areas of the *Bs*LAM enzyme and monomer **A** that of the *Cs*LAM and they were used to evaluate further experiments.

4 Structural features accounting for differences in oxygen resistance

4.1 Oxygen accessibility to the [Fe$_4$-S$_4$] cluster

4.1.1 Tunnel formation in LAM enzymes

Oxygen-labile metal cluster-containing proteins have been shown to resist oxidative damage to their cofactor by preventing the interaction of molecular oxygen with the sensitive catalytic centre. Specific residues have evolved in strategically positions, typically near the metal centre. Effective amino acids have a bigger, bulkier structure than their anaerobic protein variant to sterically shield the catalytic site (typically Ile, Phe, Trp, Arg, Tyr)[215,62,58,56] or they contain reactive centres (like Cys, Met)[57,216] that can interact with radical oxygen species by trapping and consume their harmful potential. The oxidation of methionine was reported generating methionine sulfoxide, neutralising the oxidising action of ROS species.[217]

The presence of such strategically-located residues could account for the unusual oxygen-resistance in *Bacillus subtilis* LAM and perhaps in other oxygen-tolerant SAM enzymes, and would also provide a clear mechanism for engineering more tolerant enzymes. If the *Bs*LAM oxygen-tolerance is governed by the presence of key residues in the structure that are not present in the *Cs*LAM, the engineering of new air-manageable biocatalysts of the SAM enzyme superfamily could be achieved by the substitution of amino acids that are either structurally larger or have different reaction characteristics, as appropriate.

Oxygen species are normally dissolved in water and cellular media and, as such, amino acid residues that can control these species are more likely to be located along tunnels and pathways that link the protein surface to the deeply buried [Fe$_4$-S$_4$] cluster. Previous experiments reported the effective improvement of O$_2$ resistance in the cyanobacterial [NiFe]-hydrogenase by replacement of a single amino acid with a methionine residue.[216] It was reported that at the O$_2$ concentration of 10 μM, the activity of the [NiFe]-hydrogenase

wild-type was inhibited by 95% whilst the mutant I64M could retain 50% of its activity in the cellular extract. Other measurements were performed at the enzyme level reporting the activity of the wild-type enzyme to be more affected than the I64M mutant at similar O_2 concentration, consistent with the increase of oxygen tolerance in the mutated [NiFe]-hydrogenase.

Oxygen pathways that link the protein surface to the active site are formed inside the enzyme structure because of protein dynamics. Conformational changes of the protein 3D structure can therefore generate more or less stable cavities that allow the solvent and other molecules access to the enzyme active site and interact with the catalytic cluster. If the cluster is in its reduced oxidation state, degradation by oxygen can occur. Therefore comparative analysis of tunnels formed in both *Cs*LAM and *Bs*LAM may highlight the presence of fewer and narrower tunnels to limit the presence of oxygen-scavenging residues in the gas pathways if this is the origin of oxygen tolerance. Two key aspects were examined: the size and persistence of tunnels formed in the enzyme structure due to loop movements and the nature of the specific amino acid residues lining these tunnels that may act to trap dioxygen and species otherwise harmful for the buried cluster. Pathways that were formed during the MD simulations of both LAM enzymes either with or without SAM were analysed.

4.1.2   Tunnel search calculation settings

To identify those amino acid differences in the two LAM enzymes that could interact with dioxygen entering the enzyme cavities, the evolution over time of dynamic gas pathways, investigated and compared between the two LAM enzymes. Identification of dynamic tunnels connecting the internal buried active site of the enzymes with the surrounding environment was achieved using the software CAVER 3.0.[123] CAVER allows the extraction of transient tunnels that are formed in static protein structures, and provides analysis and visualisation tools for tunnels in protein structures. Tunnels connecting the LAM buried active site with the protein surface were identified and, in CAVER, the shape of the enzymatic tunnel was

approximated as a pipeline with cross section of variable width. Such an approximation can estimate how wide the tunnel is for accessing the deepest site in the pocket and is able to highlight bottlenecks in the tunnel and the amino acid residues that compose those tunnels. The final identified clusters of tunnels are visualised as a set of intersecting balls of equal radii going through the enzyme structure to show the pathways present in the protein systems and the tunnel-lining residues. This visualisation greatly helps in the detection of bottleneck residues that are present in the active site close to the catalytic centre and provides a deeper view of the differences between possible bottleneck residues in the two LAM enzymes. The tunnel search calculation was set up in order to look for tunnels in those LAM molecular systems and identified pathways big enough to let molecular oxygen and other ROS access the enzyme in its deepest cavity.

The tunnel search calculation was performed on both *Cs*LAM and *Bs*LAM MD simulations either containing the SAM molecule in their active site, or with this cofactor absent. The PDB input structures that were needed to feed into CAVER to search for tunnels, were retrieved from every MD simulation of LAM by using a custom *cpptraj* script applied to the produced MD trajectories. The structures were extracted every 10 nanoseconds from whole 400ns MD simulations after been processed (reimaged and stripped of solvent and ions) providing a total of 40 PDB structures. These structures sampled the enzyme's overall conformations and it provided a general description of tunnels that would have been identified among the diverse enzymes. The tunnel-search computation was set up with starting point of the calculations, i.e. the bottom of each tunnel, at the [Fe$_4$-S$_4$] cluster residue contained in monomer A of either the clostridial or bacillary LAM. The FeS cluster residue was the starting point of the calculation as it is the primary target of oxidative damage with the aim to look for the most favourable and shortest pathways that connect the buried active site with the protein surface and identify possible airflow-blocking residues. Among the monomers of both the enzymes, it was chosen the least fluctuating and with the best SAM-[Fe$_4$-S$_4$] coordination so to have a stable monomer in order to minimise any source of error in the calculation.

Monomer A was selected because it was among the least fluctuating monomers of both LAMs with best SAM-coordination during the whole length of the MD simulation of both LAM enzymes. To perform an efficient and less computer-demanding tunnel computation, some of the options of choice were adapted to the different molecular systems which the computation would have been run on whilst other options were left as default unless otherwise stated. The probe radius, that indicates the minimum radius a tunnel must have to be identified in the structures, was set to the minimum of 0.9 Å, which was a balanced value in respect to computer memory usage and number of tunnel-clusters identified. The probe radius value was taken from the work by Collazo *et al.*,[218] where CAVER was employed to find oxygen channels in the soybean lipoxygenase-1 protein structure. The clustering threshold was optimised to 7.0 to restrict the number of redundant tunnels found. As similar tunnels are divided into different groups, adjusting the clustering procedure allows manageable results to be obtained.

4.1.3   Analysis of the identified clusters of tunnels

The CAVER software was able to identify tunnels in an ensemble of protein structures and enabled the automatic clustering of identified tunnels in order to assign a correspondence between the tunnels found in different snapshots. The results of clustering were given as the overall number of tunnels that belong to an identified cluster of tunnels that have been found in each of the 40 snapshots used to feed CAVER. CAVER reported the outcome of the calculations as the overall number of tunnels found in each structure that belong to different clusters; each structure may therefore experience the formation of more tunnels that did not belong to the same clusters. By setting the threshold level CAVER avoids tunnel redundancy during the calculation in identifying each tunnel and clustering. The most representative clusters of tunnels were identified in most of the structures submitted to CAVER and they were observed to be in common either in the clostridial or in the bacillary LAM simulation. Other minor clusters that only appeared sporadically over the PDB structures, generally

because they are too long or energetically unfavourable, have not been analysed.

The tunnel-search calculation was performed on the *Cs*LAM and *Bs*LAM MD simulations with SAM molecule coordinated to the FeS cluster. The outcomes were analysed and compared between the two molecular systems to observe whether the main representative tunnel-clusters were equal in number and importance in both the enzymes and to look for pathways that are hindered in *Bs*LAM and are not in the *Cs*LAM. In total 56 clusters of tunnels were detected in the *Cs*LAM enzyme and 84 clusters were found in the *Bs*LAM. Snapshots with no detected tunnels of the specific cluster, indicated those tunnels were formed with a smaller radius than the set *probe radius* parameter. Structures and other cavities with narrower radius may be present, but would not be relevant for oxygen passage through the enzyme. The tunnel clusters were chosen based on their priority and on the maximum radius detected for the tunnel, as they were considered large enough and energetically favourable for oxygen species to flow through it.

The priority of the tunnel-clusters was different in the two enzymes. A lower number of frequently detected tunnels was found in the sensitive *Cs*LAM with the highest priority of 55.4%, relative to a larger number of tunnels with slightly lower frequency found in the bacillary LAM with top priority of 33.6% (Table 2).

Among all the clusters ranked, the most relevant ones were chosen by frequency (N° snapshots), the narrowness to let oxygen flow through (maximum bottleneck radius), and energetic importance (priority), so that three tunnel-clusters were analysed in both LAM enzymes. All the other clusters of tunnels were too narrow, long or energetically unfavourable to be potential oxygen pathways. More detected tunnels were expected in *Cs*LAM if there were common pathways between the two enzymes which in *Bs*LAM should have been narrower in radius. In addition, three tunnel-clusters were identified in both LAMs to be the most relevant ones, as if those were the only tunnels of importance for LAM. The larger number of less frequent pathways found in *Bs*LAM suggested that the higher fluctuation of the amino acid residues in the tolerant enzyme relative to the *Cs*LAM did generate more, longer and

Table 2. Tunnel-cluster characteristics in SAM-bound LAM enzymes. For each LAM enzyme were reported Tunnel cluster ID, number of snapshots, the average bottleneck radius, the maximum bottleneck radius and the tunnel priority.

	T. cluster ID	N° snapshots	Average bottleneck radius (Å)	Maximum bottleneck radius (Å)	Priority
*Cs*LAM	1	37	1.224 ± 0.127	1.43	55.4%
	2	36	1.215 ± 0.119	1.40	39.3%
	3	29	1.116 ± 0.153	1.40	25.5%
*Bs*LAM	1	37	1.090 ± 0.139	1.42	33.6%
	2	26	1.109 ± 0.132	1.41	26.5%
	3	29	1.067 ± 0.137	1.40	21.8%

occasional cavities; those did not describe important pathways that linked the protein surface with the active site. Furthermore, the clusters analysed were the most relevant ones among all the clusters and each of them showed a similar bottleneck radius, which seems to be consistent with the absence of bulkier residues in the active site of the oxygen-tolerant enzyme.

The same tunnel-search calculation was also performed on both the unbound-SAM MD simulations. The 40 snapshots without the SAM cofactor coordinated provided 83 clusters of tunnels for the *Cs*LAM structures and 171 clusters for the *Bs*LAM. This is twice of the tunnel clusters that were identified in the SAM-bound *Bs*LAM. *Cs*LAM also formed more tunnels when SAM was unbound. As a consequence, the priority of tunnel-clusters was lower than the

previous tunnel search calculation and equal between the two LAM enzymes. This resulted in the top priority of 37.7% in the clostridial LAM and 36.1% for *Bs*LAM (see Appendix 6.12). More and narrower tunnels were then found in both the LAM enzymes, consistent with the absence of SAM from the active site that lead to a consequent increase of loop fluctuations in the enzyme structure, creating more occasional pathways and slightly wider tunnels in both LAM enzymes. This widening decreased the overall tunnel-cluster priorities with respect to the SAM-bound LAM outcomes and is consistent with the presence of SAM in the active site conferring stability to the whole enzyme structure, in terms of conformational motions in the LAM structures. The tunnels found in *Bs*LAM with unbound-SAM were reported to be larger than the correspondent bound-SAM structure and such a difference was more remarkable than in the bound/unbound-SAM *Cs*LAM enzymes. That was likely due to the wider fluctuation of the enzyme monomers in absence of SAM from the active site, which was consistent with the outcomes of the *Bs*LAM MD simulations reported in the previous chapter. The absence of SAM had an impact on the loop motions in both LAM enzymes but the homology model structure of *Bs*LAM was more affected.

The selected clusters with the highest priority were composed by as the same amino acid residues that formed the tunnels in both the SAM-bound/unbound enzyme structures and the number of tunnel-clusters was significantly larger in those structures without SAM. For this reason we considered the SAM-bound tunnel-clusters results to analyse in detail as considered more significant.

4.1.4 Analysis of the tunnel-lining amino acid residues

Both the steric and chemical properties of the tunnel-lining and bottleneck residues were examined for their potential influence on the [Fe_4-S_4] cluster oxygen sensitivity regulation. The effect of the sulphur-containing residues placed close to the active site could protect the metal cluster from harmful oxidation by interacting and trapping dioxygen molecule and ROS species, so destroying their harmful potential. Sulfur-containing amino acid residues are

prone to oxidation by oxygen and ROS species allowing them to act as an antioxidant, a form of protection towards the harmful oxidative stress, when in proximity to the [Fe$_4$-S$_4$] cluster. in particular, the oxidation of methionine residues generates methionine sulfoxide, neutralising the oxidising action of the ROS.[217]

In order to figure out if the flow of oxygen and ROS species was either chemically or sterically obstructed by residues in the two enzymes, the common amino acid residues in the main tunnel-clusters were analysed. The bottleneck residues create the narrowest part of the tunnel, thus creating a point where the maximal interaction between that amino acid and oxygen, on the way to the active site, is more likely. As the tunnel-clusters identified, had their average radius close to that of dioxygen, all the amino acid residues involved in the formation of tunnels were assessed, as they all can in principle interact with molecular oxygen.

The sequence alignments from both LAM enzymes are shown in Figure 34 and all the residues found composing the three tunnel-clusters in the bound-SAM structures are highlighted. The three clusters are composed of residues belonging to different areas of the enzyme and, due to the similarity of the residues involved in the two enzymes we could match the position of the tunnel-cluster and the residues sitting along each cluster. Interestingly, the clusters were equally ranked between the two enzymes so that the same cluster had the same importance in both LAMs. The presence of methionine residues was first examined, as they are able to scavenge harmful oxygen species and there might be the possibility that some methionine residues were strategically placed among the tunnel-lining residues of the *Bs*LAM enzyme structure identified by CAVER. The overlaid sequence alignments between the two LAM enzymes (Fig. 34) shows the amino acid residues identified along the three clusters of tunnel for *Cs*LAM in purple and for the *Bs*LAM in cyan. The residues along the tunnels reside in similar structural areas or loops of the two LAM enzymes and are made up of common sets of amino acids residues. Tunnel-cluster 1 (Figure 34a) was focused on residues 95 – 120 in *Cs*LAM (105 – 130 *Bs*LAM) which covers one of the main fluctuating enzyme areas that was discussed in the previous chapter (90 – 120 and 140 – 160 in *Cs*LAM and residues 100 – 130

a)
```
sp|Q9XBQ8|KAMA_CLOSU    KEEEEGVAQCVKSLRMAITPYYLSLIDPNDPNDPVRKQAIPTALELNKAA 91
sp|O34676|KAMA_BACSU    EDEEEGVRISTKTIPLNITPYYASLMDPDNPRCPVRMQSVPLSEEMHKTK 100
                         :* ***     .:: : ***** **:* :.  . *:* *::*   *::::

sp|Q9XBQ8|KAMA_CLOSU    ADLEDPLHEDTDSPVPGLTHRYPDRVLLLITDMCSMYCRHCTRRRFAGQS 141
sp|O34676|KAMA_BACSU    YDLEDPLHEDEDSPVPGLTHRYPDRVLFLVTNQCSMYCRYCTRRRFSGQI 150
                         *  ****** *****************:*:*: ******:******:*.

sp|Q9XBQ8|KAMA_CLOSU    DDSMPMERIDKAIDYIRNTPQVRDVLLSGGDALLVSDETLEYIIAKLREI 191
sp|O34676|KAMA_BACSU    GMGVPKKQLDAAIAYIRETPEIRDCLISGGDGLLINDQILEYILKELRSI 200
                        . .:  ::* ** ** :**::** *:****.**:.*: **  : .** *

sp|Q9XBQ8|KAMA_CLOSU    PHVEIVRIGSRTPVVLPQRITPELVNMLKKYHPVWLNTHFNHPNEITEES 241
sp|O34676|KAMA_BACSU    PHLEVIRIGTRAPVVFPQRITDHLCEILKKYHPVWLNTHFNTSIEMTEES 250
                        **:*::***:*.***:***** .*  .:******:******  . *:* *:

sp|Q9XBQ8|KAMA_CLOSU    TRACQLLADAGVPLGNQSVLLRGVNDCVHVMKELVNKLVKIRVRPYYIYQ 291
sp|O34676|KAMA_BACSU    VEACEKLVNAGVPVGNQAVVLAGINDSVPIMKKLMHDLVKIRVRPYYIYQ 300
                         .**: *.:**:*:***:*:* *:**.: :**.*::.** :*********

sp|Q9XBQ8|KAMA_CLOSU    CDLSLGLEHFRTPVSKGIEIIEGLRGHTSGYCVPTFVVDAPGGGGKTPVM 341
sp|O34676|KAMA_BACSU    CDLSEGIGHFRAPVSKGLEIIEGLRGHTSGYAVPTFVVDAPGGGGKIALQ 350
                        **** *: ***.**:**.**************.************** .:
```

b)
```
sp|Q9XBQ8|KAMA_CLOSU    KEEEEGVAQCVKSLRMAITPYYLSLIDPNDPNDPVRKQAIPTALELNKAA 91
sp|O34676|KAMA_BACSU    EDEEEGVRISTKTIPLNITPYYASLMDPDNPRCPVRMQSVPLSEEMHKTK 100
                         :* ***     .:: : ***** **:* :.  . *:* *::*   *::::

sp|Q9XBQ8|KAMA_CLOSU    ADLEDPLHEDTDSPVPGLTHRYPDRVLLLITDMCSMYCRHCTRRRFAGQS 141
sp|O34676|KAMA_BACSU    YDLEDPLHEDEDSPVPGLTHRYPDRVLFLVTNQCSMYCRYCTRRRFSGQI 150
                         *  ****** *****************:*:*: ******:******:*.

sp|Q9XBQ8|KAMA_CLOSU    DDSMPMERIDKAIDYIRNTPQVRDVLLSGGDALLVSDETLEYIIAKLREI 191
sp|O34676|KAMA_BACSU    GMGVPKKQLDAAIAYIRETPEIRDCLISGGDGLLINDQILEYILKELRSI 200
                        . .:  ::* ** ** :**::** *:****.**:.*: **  : .** *

sp|Q9XBQ8|KAMA_CLOSU    PHVEIVRIGSRTPVVLPQRITPELVNMLKKYHPVWLNTHFNHPNEITEES 241
sp|O34676|KAMA_BACSU    PHLEVIRIGTRAPVVFPQRITDHLCEILKKYHPVWLNTHFNTSIEMTEES 250
                        **:*::***:*.***:***** .*  .:******:******  . *:* *:

sp|Q9XBQ8|KAMA_CLOSU    TRACQLLADAGVPLGNQSVLLRGVNDCVHVMKELVNKLVKIRVRPYYIYQ 291
sp|O34676|KAMA_BACSU    VEACEKLVNAGVPVGNQAVVLAGINDSVPIMKKLMHDLVKIRVRPYYIYQ 300
                         .**: *.:**:*:***:*:* *:**.: :**.*::.** :*********

sp|Q9XBQ8|KAMA_CLOSU    CDLSLGLEHFRTPVSKGIEIIEGLRGHTSGYCVPTFVVDAPGGGGKTPVM 341
sp|O34676|KAMA_BACSU    CDLSEGIGHFRAPVSKGLEIIEGLRGHTSGYAVPTFVVDAPGGGGKIALQ 350
                        **** *: ***.**:**.**************.************** .:
```

c)
```
sp|Q9XBQ8|KAMA_CLOSU    KEEEEGVAQCVKSLRMAITPYYLSLIDPNDPNDPVRKQAIPTALELNKAA 91
sp|O34676|KAMA_BACSU    EDEEEGVRISTKTIPLNITPYYASLMDPDNPRCPVRMQSVPLSEEMHKTK 100
                         :* ***     .:: : ***** **:* :.  . *:* *::*   *::::

sp|Q9XBQ8|KAMA_CLOSU    ADLEDPLHEDTDSPVPGLTHRYPDRVLLLITDMCSMYCRHCTRRRFAGQS 141
sp|O34676|KAMA_BACSU    YDLEDPLHEDEDSPVPGLTHRYPDRVLFLVTNQCSMYCRYCTRRRFSGQI 150
                         *  ****** *****************:*:*: ******:******:*.

sp|Q9XBQ8|KAMA_CLOSU    DDSMPMERIDKAIDYIRNTPQVRDVLLSGGDALLVSDETLEYIIAKLREI 191
sp|O34676|KAMA_BACSU    GMGVPKKQLDAAIAYIRETPEIRDCLISGGDGLLINDQILEYILKELRSI 200
                        . .:  ::* ** ** :**::** *:****.**:.*: **  : .** *

sp|Q9XBQ8|KAMA_CLOSU    PHVEIVRIGSRTPVVLPQRITPELVNMLKKYHPVWLNTHFNHPNEITEES 241
sp|O34676|KAMA_BACSU    PHLEVIRIGTRAPVVFPQRITDHLCEILKKYHPVWLNTHFNTSIEMTEES 250
                        **:*::***:*.***:***** .*  .:******:******  . *:* *:

sp|Q9XBQ8|KAMA_CLOSU    TRACQLLADAGVPLGNQSVLLRGVNDCVHVMKELVNKLVKIRVRPYYIYQ 291
sp|O34676|KAMA_BACSU    VEACEKLVNAGVPVGNQAVVLAGINDSVPIMKKLMHDLVKIRVRPYYIYQ 300
                         .**: *.:**:*:***:*:* *:**.: :**.*::.** :*********

sp|Q9XBQ8|KAMA_CLOSU    CDLSLGLEHFRTPVSKGIEIIEGLRGHTSGYCVPTFVVDAPGGGGKTPVM 341
sp|O34676|KAMA_BACSU    CDLSEGIGHFRAPVSKGLEIIEGLRGHTSGYAVPTFVVDAPGGGGKIALQ 350
                        **** *: ***.**:**.**************.************** .:
```

Figure 34. Sequence alignment of CsLAM and BsLAM. The residues along the three tunnel-clusters that were identified by CAVER are highlighted in purple for CsLAM and in cyan BsLAM enzyme in the SAM-bound simulations.

and 150 – 170 in *Bs*LAM). The tunnel-lining amino acid residues highlighted in Fig. 34 seemed not to present any remarkable difference between the two enzymes in terms of bulkiness or oxygen trapping ability. Tunnel-cluster 2 (Figure 34b) appeared in a different area of the enzyme structure because residues 138 – 150 (147 – 160 in *Bs*LAM) belonged to the protein loop that started at the end of the [Fe$_4$-S$_4$] cluster-binding loop and it ended up in a helix of the TIM-barrel structure that was observed to fluctuate. That caused an opening in the enzyme structure that lead to the formation of tunnels that involved residues 95 – 120 (or 105 – 130) that were identified to fluctuate as well as it was observed for cluster 1. The third tunnel-cluster (Figure 34c) involved several residues from the section of structure that covers the loop 95 – 120 in *Cs*LAM (105 – 130 in *Bs*LAM) but are also involved in the binding of the PLP cofactor when no substrate is present into LAM. The tunnels of this cluster include residues 330 – 340 in *Cs*LAM (340 – 350 in *Bs*LAM) and contained the PGGGGK loop involved in binding the PLP-cofactor and they might form a pathway for the access of the molecule but deeper studies are needed to confirm the mechanism. The tunnel-lining residues showed residues of consistent size no bigger, bulkier amino acids present in *Bs*LAM to sterically hinder the oxygen and ROS species flow that account for its oxygen-resistance. The sulfur-containing residues with potential oxygen-scavenging activity were equally distributed along the considered pathways between the two enzymes. An initial analysis of both the LAM enzymes sequences, identified the overall abundance of cysteine residues to be 11 in *Cs*LAM and 9 in the *Bs*LAM whilst the methionine distribution was 9 and 8 in *Cs*LAM and *Bs*LAM respectively. A larger number of sulfur-containing residues in the *Bs*LAM involved either as bottleneck residues or along the tunnels would have been expected to confer additional stability against oxidative stress. In contrast, CAVER identified more methionine residues involved in tunnel formation for the *Cs*LAM relative to its bacillary counter enzyme. This implies that the reason why *Bs*LAM tolerates oxygen and ROS species, does not reside in reactive residues or in constriction of the oxygen flow. The implication, based on the evidence from *Bs*LAM, is that developing an oxygen-tolerant *Cs*LAM mutant, has not naturally evolved

by the simple approach of substituting key amino acid residues to either physically shield or chemically scavenge the buried [Fe$_4$-S$_4$] cluster from the attack of ROS species, potentially limiting protein engineering possibilities on the basis. It also suggests that the tolerance of the bacillary LAM has a different origin, and other aspects of structure may play a role.

4.1.5   Inspection of the tunnel-clusters

The first three clusters of tunnels reported by CAVER, were compared and visually inspected with VMD. A clearer view of the tunnel-clusters was obtained by representing them as a series of interconnected balls of different colours, originating from the FeS cluster of either the oxygen-sensitive or the oxygen-tolerant LAM enzyme (Figure 35). The three clusters of tunnels were identified as the widest cavities from the calculation and they were found being of equal size and position within the structurally aligned enzymes. This superposition suggested they might be involved in substrate or cofactor access, in addition to enabling access to either solvent or radical oxygen species. The tunnels originated in highly fluctuating regions of the structure. The oscillation of the loops enabled the formation of energetically favoured cavities that were detected over the simulation timeframe and directly linked the buried [Fe$_4$-S$_4$] cluster to the protein surface within a short distance. Such tunnels may be wide enough to be involved with the creation of pathways for the access of reactants needed by the LAM enzyme for its catalysis. It is hard to make a distinction regarding which tunnels were involved with the access of which reactant but the tunnels in red and blue (Figure 35, clusters ID 1 and 2 respectively) might play a role in the access of the SAM cofactor and lysine substrate as these two clusters of tunnels seemed to represent the shortest paths to the active site. The opening of the tunnel-cluster 3, represented in green and in more detail in Figure 35, has the external end close to Lys337 (Lys346 in *Bs*LAM), which is the enzyme residue responsible for the coordination of the PLP cofactor when no substrate is present into LAM. A hypothesis might be that tunnel cluster 3 is the favourite pathway for the adsorption of PLP
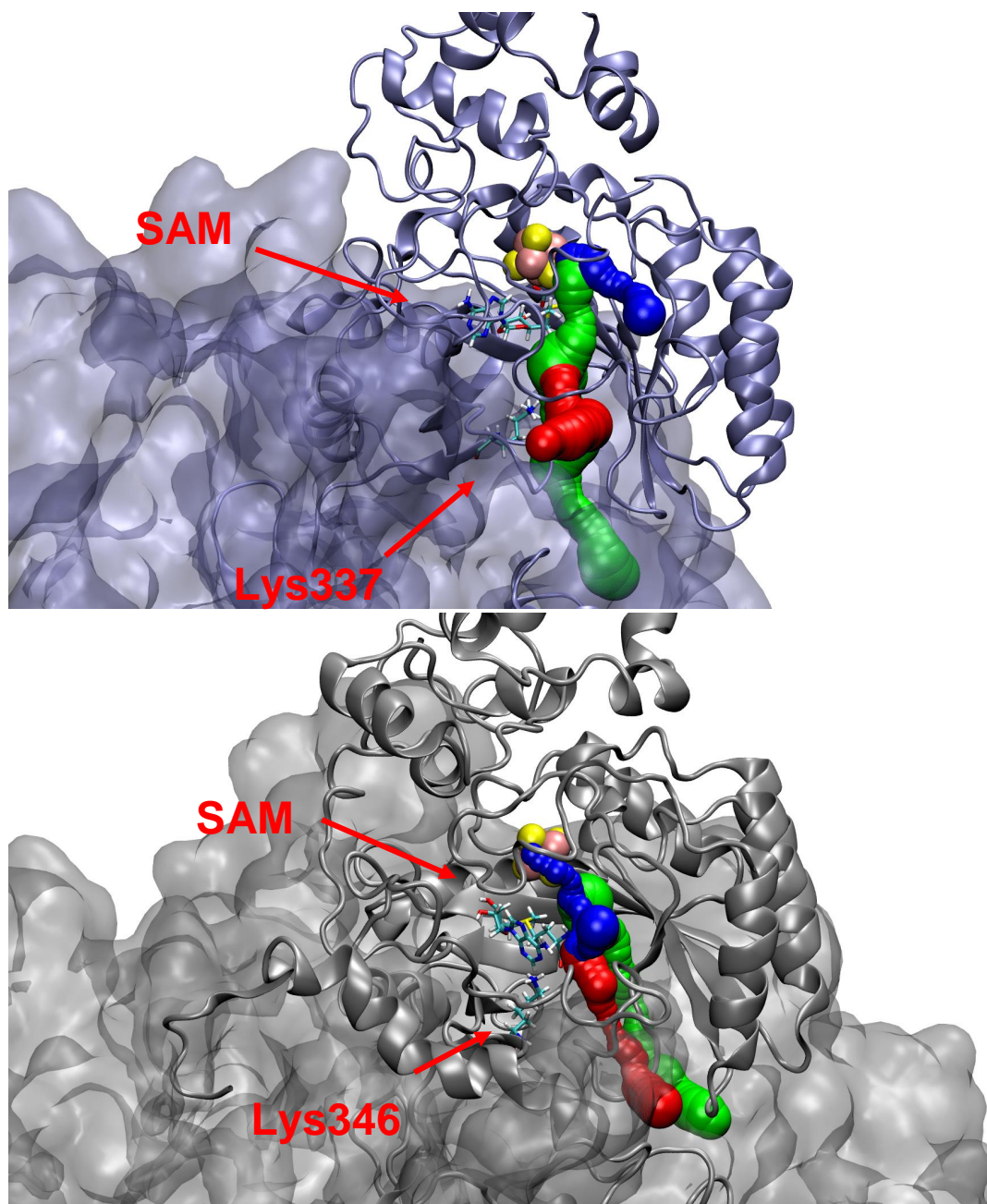
Figure 35. Detail of the lysine active site and the first three tunnel-clusters reported by CAVER. The clusters of tunnels developed into similar areas of the *Cs*LAM (above) and *Bs*LAM (below), the tunnels are depicted in red (Cluster 1), blue (Cluster 2) and green (Cluster 3). Detail of the PGGGK loop and Lys337 (Lys346 in bacillary LAM) that was found to be part of tunnel-cluster 3. Visualisation of the analysed tunnels was set up by representing the selected enzyme monomer of interest in *Ribbon* whilst the rest of the enzyme is represented by the *Quick Surface* representation for better visualisation. The [Fe$_4$-S$_4$] is represented in *VanDerWaals* and the SAM molecule is in *Licorice* to better distinguish each enzyme section. The monomeric structures are shown in silver for *Cs*LAM and in cyan for *Bs*LAM and are also partially coloured in purple. The purple area of the monomer represents those structural loops whose amino acid residues showed the highest fluctuation in the analysis of the MD simulations that was previously discussed (sections 3.4.2 and 3.4.4).

from the environment into the enzyme so that the cofactor could bind to Lys337 and form the PLP-Lys337 imine adduct.[85] Once the lysine substrate enters the catalytic pocket from another tunnel, the adduct undergoes a transimidation reaction to form the PLP-lysine adduct that forces the SAM molecule to trigger the substrate conversion as it was showed in the introduction.[89] These observations imply that the dynamics of reactants and the way they interact with the enzyme to be absorbed to start the catalysis, could be probed through driven docking studies or other appropriate approaches. Although beyond the scope of the current study, this could provide a fruitful avenue to better understand LAM mechanism.

4.1.6 Concluding remarks

In order to obtain insights about the internal structure and the pathways where oxygen could flow to the active site in the two LAM enzymes, a 3D map of the main gas channels was calculated. The analysis of such pathways resulted in the same amount of gas channels identified in the two enzymes and they were also of the same size. From the analysis we could establish the absence of bulkier residues along the channels to hinder the passage of oxygen that could have explained the higher tolerance to oxygen in *Bs*LAM. The analysis of the residues composing the identified pathways, did not report the presence of amino acid residues capable to interact with oxygen molecules upon trapping it. That could have been another reason to elucidate oxygen tolerance in the bacillary LAM enzyme.

The final remarks of this analysis were that LAM enzymes oxygen tolerance seems not to be driven by bulky amino acid residues nor oxygen scavenging residues. In this pair of SAM enzymes the tolerance to the attack of ROS species might be regulated by electrochemical properties that will be discussed in the next section.

4.2 Redox activity of the [Fe$_4$-S$_4$] cluster

4.2.1 The protein environment and its interaction with the [Fe$_4$-S$_4$] cluster

As it was previously discussed in the introduction, FeS cluster-containing proteins can develop tolerance toward oxygen by increasing the reduction-oxidation potential of their metal centre. In ferredoxins[66] and [NiFe]-hydrogenases,[219] either containing a [Fe$_4$-S$_4$] or a [Fe$_3$-S$_4$] cluster, a common trend was observed where higher redox potentials of the metal centre protein could account for better function in presence of air.

The analyses of the MD simulations of the LAM enzymes indicated that *Bs*LAM oxygen-tolerance was not due either to an obstructed air-flow to the active site nor to an obvious presence of oxygen-scavenging residues to prevent the FeS cluster oxidative damage. An increased midpoint potential of the metal cluster mediated by a different protein environment around the cluster, in *Bs*LAM may thus explain the ability of this enzyme to deal with oxygen. Specific residues in the active site are able to modulate the redox potential to higher values in *Bs*LAM and this section attempts to identify those residues and subsequently design *Cs*LAM mutants carrying redox-modulating residues at the corresponding positions.

To be able to detect essential redox-tuning residues in *Bs*LAM, proteins that contain fully cysteine-coordinated FeS clusters and achieve +3, +2 and +1 oxidation states, were drawn from the literature as examples. No information about redox potential tuning is available for the radical SAM family or the [Fe$_4$-S$_4$] cluster in coordination with the *S*-adenosylmethionine. Among the factors that influence the redox modulation of metal clusters in proteins are the solvent accessibility,[220] and the distortion of the cluster driven by the dihedral angles of the cysteine ligands.[221] The two LAM enzymes have no difference in solvent accessibility, as evidenced in the tunnel analysis section 4.1. The workflow employed to look for structural differences in the cuboidal FeS cluster in both LAM enzymes is therefore described. Quantum-mechanics (QM) calculations were performed on [Fe$_4$-S$_4$](SCH$_3$)$_3$ sub-systems, whose

coordinates were retrieved from the MD simulations in order to observe any difference or source of cluster distortion.

4.2.2 Structural investigation of [Fe$_4$-S$_4$](SCH$_3$)$_3$ structures from MD simulations

Initial examination was made of the [Fe$_4$-S$_4$](SCH$_3$)$_3$ systems without added residues to confirm whether any structural difference within the metal cluster could account for the difference in reactivity between the two LAM enzymes. We started the FeS cluster investigation by searching for possible cubane structural distortions and looking at their dihedral angles, as these could have been source of different electrostatic fields. Visual inspection of the [Fe$_4$-S$_4$] RMSD over the timecourse of the MD simulations indicated no particular distortion at the metal cluster in the two LAM enzymes (see Appendix 6.19). The distribution of the cluster-binding cysteine dihedral angles was compared in order to confirm the steadiness of these angles over the timecourse of the simulations (see Appendix 6.21 and 6.22). Once the overall stability was confirmed, the atom coordinates of [Fe$_4$-S$_4$](SCH$_3$)$_3$ systems were obtained from selecting ten MD snapshots from either *Cs*LAM and *Bs*LAM simulations in their unbound-SAM state for ease of calculation. The selected structures were representative of the average structure of the [Fe$_4$-S$_4$] cluster over the MD simulation. The MD structures were extracted at 10, 70, 150, 170, 190, 210, 250, 290, 310, 380 and 400 ns of both LAM MD simulations. To start the structural comparison, the selected structures were overlaid as shown in Figure 36. The RMSD between all the structures was $8.9 \cdot 10^{-2}$ Å for *Cs*LAM and $11.2 \cdot 10^{-2}$ Å for *Bs*LAM.

The measured dihedral angles for each cluster-binding thiolate ligand were considered to look at possible different orientations that might account for the VEA fluctuations over the selected snapshots (Appendix 6.21). Dihedral angles were measured, including the Cβ of the thiolate moiety, the Sγ of the thiolate, and the Fe and the S of the cluster. RMSD was measured

by using the 400 ns FeS cluster + thiolates system as the template structure. The aligned structures of the sensitive LAM looked more similar to each other over the MD simulations compared with those for the tolerant *Bs*LAM, which had a slightly higher RMSD.

From the analysis of both the dihedral angles and structural visualisation after their alignment, it was observed that snapshots 210 and 250 did not align perfectly to the rest of the structures and their dihedral angles moderately diverged from the rest of the structures. The different dihedral angle distributions in the two LAM enzymes looked more ordered in *Cs*LAM than in *Bs*LAM and that might have been a consequence of the different loop motions between the two enzymes. The angle distribution of the analysed cysteine residues was driven by the fluctuation of the cluster-binding loop, which was slightly higher in the *Bs*LAM MD simulation.

The similarity of the FeS cluster orientation in the two enzymes confirmed the structural equality of the prosthetic group within the oxygen-sensitive and tolerant LAM enzymes at the molecular mechanics (MM) level. Nevertheless, the MD need to be analysed in deeper detail at the QM level, to observe how much the structural orientations affect the electrostatics of the metal cluster, and hence the redox potential.

*Cs*LAM   *Bs*LAM
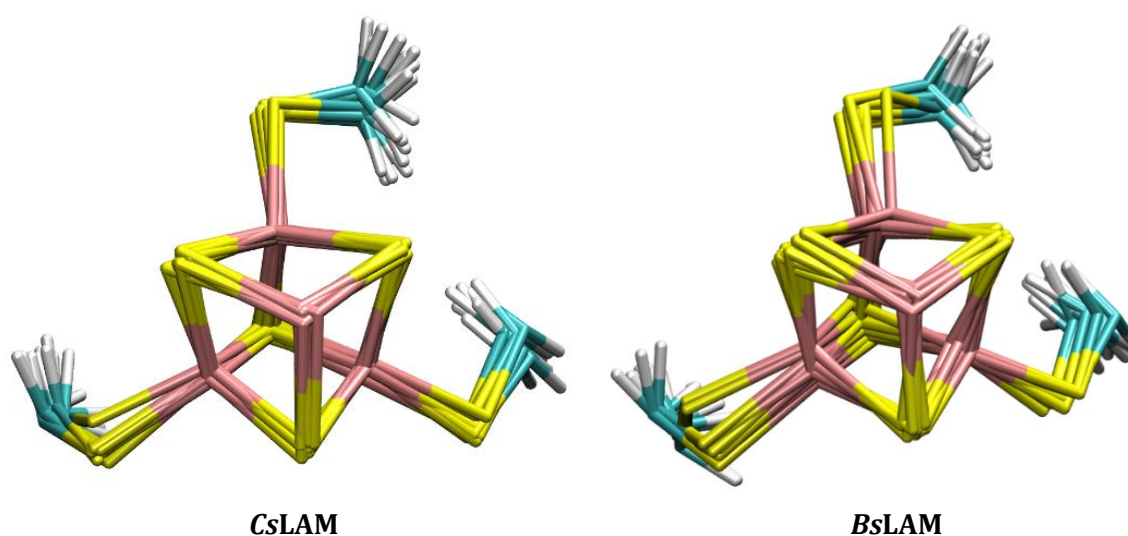
Figure 36. FeS cluster coordinates retrieved from MD simulations and their structural alignment of *Cs*LAM (left) and *Bs*LAM (right).

4.2.3 Quantum Mechanical calculations of [Fe$_4$-S$_4$](SCH$_3$)$_3$ structures

The Quantum Mechanical (QM) approach was employed by using the Gaussian09 software package. The method aimed to structurally optimise the geometries of each snapshot taken into account after energy minimisation, to look at possible conformational differences and to report the VEA of each structure. After minimisation of the retrieved coordinates, a 3-step geometry optimisation at the oxidised state of the metal cluster and *in vacuo*, was performed for each system.

The new aligned geometries after structural optimisation was achieved are shown in Figure 37. The similarity of the atomic systems was estimated through their total RMSD, which provided the value of $5 \cdot 10^{-3}$ Å for the *Cs*LAM and $8 \cdot 10^{-3}$ Å for the *Bs*LAM enzyme. Assessment of the decreased RMSD values compared with those before the geometry optimisation (section 4.2.2) indicated the ten structures for each enzyme looked more similar to each other after been geometrically optimised to a lower energy level.

Despite the low FeS cluster RMSD, the thiolate ligands showed different orientations for their terminal methyl moiety (Figure 37); the dissimilarity was likely due to the optimisation in the

CsLAM   **BsLAM**
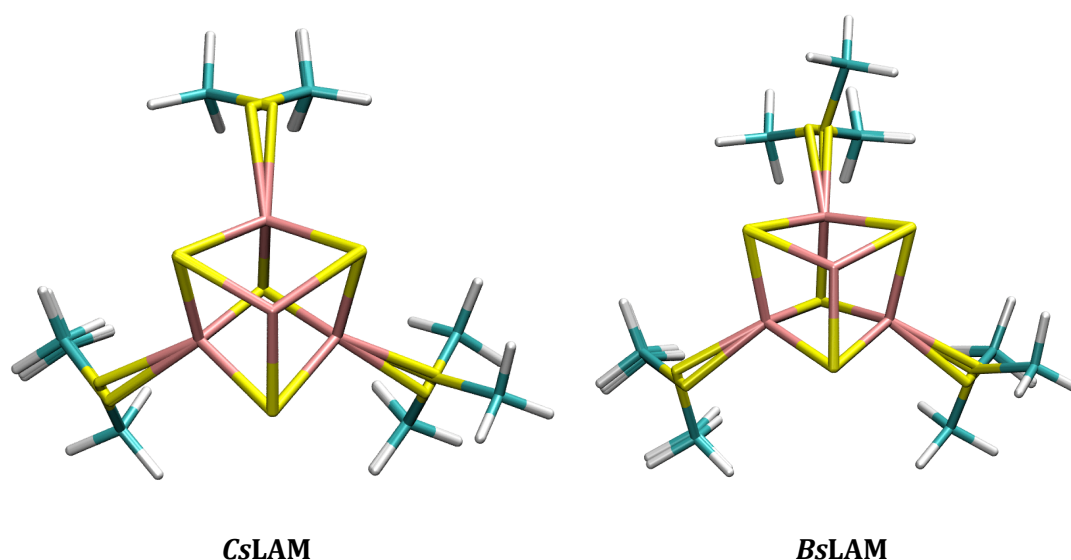
Figure 37. [Fe$_4$-S$_4$](SCH$_3$)$_3$ structures aligned after a 3-step geometry optimisation, the *Cs*LAM systems are on the left whilst *Bs*LAM are on the right. Some cluster conformations could be identified.

gas-phase that emphasized small differences at the quantum level. The dihedral angles of the thiolate ligands (Appendix 6.25) indicate the different atomic orientations after optimisation; a visual representation reporting the different conformations assumed by the thiolate ligands is also shown in the Appendix. Analysis of the dihedral angles and visualisation of the structures, in both LAMs indicated the formation of clusters based on the different thiolate atoms conformations. In *Cs*LAM only two groups of structures with different orientations were present. Group **I** of the table in Appendix 6.25 consists of seven snapshots, the majority of the structures showed the energetically favoured conformation for these molecular systems whilst group **II**, with three snapshots (at 150, 210 and 290 ns), contained explored conformations that accounted for 30% of the conformations analysed, where the only difference is in the orientation of Cys130 (on the right side of the FeS cubane).

In *Bs*LAM four clusters of conformations were grouped with five structures encompassed in cluster **I**. Clusters **II** and **III** contained two conformations, whilst **IV** only contained one structure. The thiolate orientations in cluster **I** and **II** were consistent with the structures highlighted in *Cs*LAM where they differed in the orientation of Cys141 (the equivalent of *Cs*LAM Cys130 on the right side of the cubane). Groups **I** and **II** of *Bs*LAM contained most of the conformations explored by the enzyme that were the same conformations explored by *Cs*LAM. Cluster **I** was reported as the main conformation populated in both the enzymes but the conformation in group **I** of *Cs*LAM was identical to that of group **II** of *Bs*LAM and conversely group **II** of *Cs*LAM related structurally to group **I** of *Bs*LAM. The appearance of clusters **III** and **IV** might be due to the general loop fluctuation that was reported to be higher, relative to that in *Cs*LAM simulations. In particular, cluster **IV** only contained snapshot 380, which seemed to be more of an occasionally assumed conformation. The geometry optimisation highlighted the energetically stable conformations of the cubane cluster and the orientation of the thiolate ligands that coordinate the cluster were of a similar orientation in both the LAM enzymes.

The optimised structures were used to calculate the VEA and the results of these calculations are shown in Figure 38. The trends of the final energy of each system in both their oxidation states were plotted to have a graphical view of the energies (Appendix, 6.26). Plots a) and b) in the Appendix show the energy level of FeS cluster systems either on their oxidised and reduced state of *Cs*LAM and *Bs*LAM respectively.

The overall energy trends were similar between the two enzymes, which could have described a similar electrostatic structure of the prosthetic group, but there is a small fluctuation between the two. The fluctuation throughout every snapshot in the same enzyme, might be due to the slightly different FeS cluster geometry, therefore the dihedral orientation might have affected the overall FeS cluster electrostatics. The slightly higher values in the reduced state of the tolerant LAM might confirm a less stabilised intermediate in the electron transfer process of the SAM cleavage as was reported by Chen *et al*.[49]

Figure 38 displays the mean VEA values of molecular energy levels and their standard deviation from the single-point calculation reported in kJ mol^{-1}. Although the energy means between the two LAM enzymes were different, recording 1753.86 ± 58.47 kJ mol^{-1} for *Cs*LAM and 1797.03 ± 56.94 kJ mol^{-1} for *Bs*LAM, the two energy distributions overlapped (figure 38). To test the statistically relevance of the two data distribution, an independent-samples t-test was conducted. There was not a significant difference between the VEA in the two LAM enzymes (Appendix 6.27) demonstrating that the electrostatics of the systems $[Fe_4\text{-}S_4](SCH_3)_3$ are approximately the same throughout the two enzymes and that different dihedral angles marginally affected the geometrical, and thus the electrostatic structure. There is a general consistency between the structural conformation assumed by the metal cluster and VEA values, indicating that the energy slightly fluctuated over the MD simulation of the LAM enzymes even though the cysteine ligands reorientation did not happen. Such a fluctuation did not manifest as a substantial difference in midpoint potentials of the FeS cluster in the two LAM enzymes. Therefore, this last test suggested that a wider look at the protein environment should be taken, to observe the influence of those key residues around the FeS

cluster on the electrostatics. As the electrostatics of the cuboidal FeS cluster alone were similar, as assessed by the calculations in the gas phase, the different protein environments might affect the redox potential of the metal cluster in the two enzymes, This study helped us identify that the methyl dihedral angles identified in the enzyme by MD, and further refined by QM, are unlikely to have a meaningful effect on the differences in electron affinity of the $[Fe_4-S_4]$ cluster, and thus the cluster reactivity. The oxygen-sensitivity of lysine 2,3-aminomutase radical SAM enzyme does not thus depend on differently oriented cysteine ligands that modify the reactivity of the metal cluster; instead the answer might reside in the protein environment.
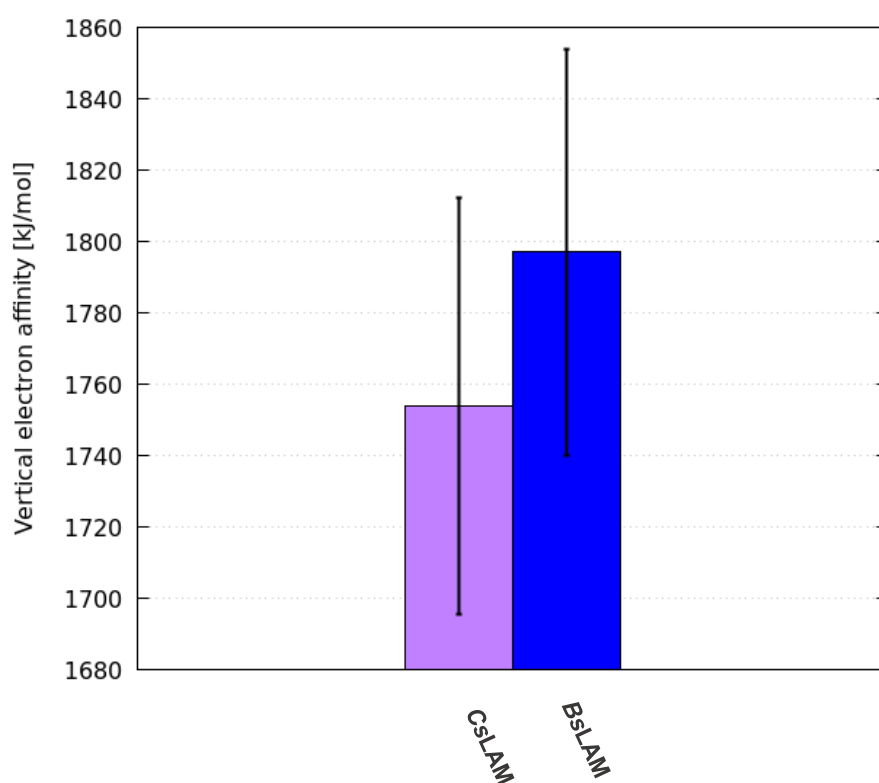
Figure 38. Final energy of the atomic systems $[Fe_4-S_4](SCH_3)_3$ at both oxidation states of *Cs*LAM and *Bs*LAM. VEA mean distribution with standard deviation in both LAM enzymes.

4.2.4 Mapping the redox modulating residues in LAM protein environment

To identify the relevant residues with electrostatic modulating properties, *cpptraj* was used. The *hbond* command could detect the hydrogen bonding residues and specify what atoms they were interacting with (Appendix 6.28). A distance of 4.0 Å was used to identify all the hydrogen-bond amino acid residues interacting with sulfur atoms contained in either the metal cluster (S) or the cysteinyl ligands (S_γ). Other molecular groups, such as either, carbonyl or charged groups, were visually identified with VMD by selecting the specific moiety of interest at the specific distance throughout the simulation. The analysis was run on both the SAM-bound and –unbound simulations of both LAM enzymes to identify only those residues found hydrogen-bonds interacting with the atoms of interest (inorganic S, ligand S_γ and Fe). Those residues that would have interacted only with SAM by establishing hydrogen-bonds were not considered because of the identical nature of the amino acids between the two LAM enzymes. Residues bound to the cysteinyl ligands by the peptide bond with their carbonyl group were added according to the mapping rules. Finally, a model containing the main redox-relevant interacting residues in the four monomers of LAM enzymes over 400ns MD simulation time is depicted in Scheme 2. At the centre of each map, the [Fe_4-S_4] cluster is circled in orange whilst the three cysteinyl ligands next to it are in blue. The residues with peptide bonds and the carbonyl groups neighbouring the cysteinyl ligands (Met124, Tyr128, His131 in *Cs*LAM and Gln133, Tyr137 and Tyr140 in *Bs*LAM) are enclosed in violet hexagons and here the first difference of protein environment could be observed. The residues with neighbouring carbonyls are the methionine and histidine in the sensitive LAM and a glutamine and tyrosine in the bacillary aminomutase. The Gln133 in *Bs*LAM (Met124 in *Cs*LAM) might interact differently with the metal cluster due to its sidechain amide moiety that brings a second dipole whilst the Tyr140 with its aromatic ring and hydroxyl group might be able to tune the redox potential more efficiently than the corresponding His131 in *Cs*LAM. Due to their structural and electronic differences, these residues might play a role in the midpoint potential modulation and, as such, they were chosen as possible positions that

might generate a less oxygen-sensitive *Cs*LAM mutant. There is no difference among the charged residues identified and they are all well known to participate in SAM and PLP orientation. Another difference that was observed among other hydrogen-bond interacting residues was where they are in common to both the enzymes such as His230 and Thr133 in *Cs*LAM and Ser135 in *Bs*LAM (respectively *Bs*LAM His239, Thr142 and *Cs*LAM Ser126). *Bs*LAM Ser147 was found to interact with the S_γ of Cys134 and S in the metal cluster through its hydroxyl group, whilst at the corresponding clostridial position Ala138 is found. Because of the different interaction, this position was also chosen as a mutation candidate.

The same analysis was also performed on both starting structures of LAM enzymes (the crystal structure and the homology model) to visually identify on VMD those residues that
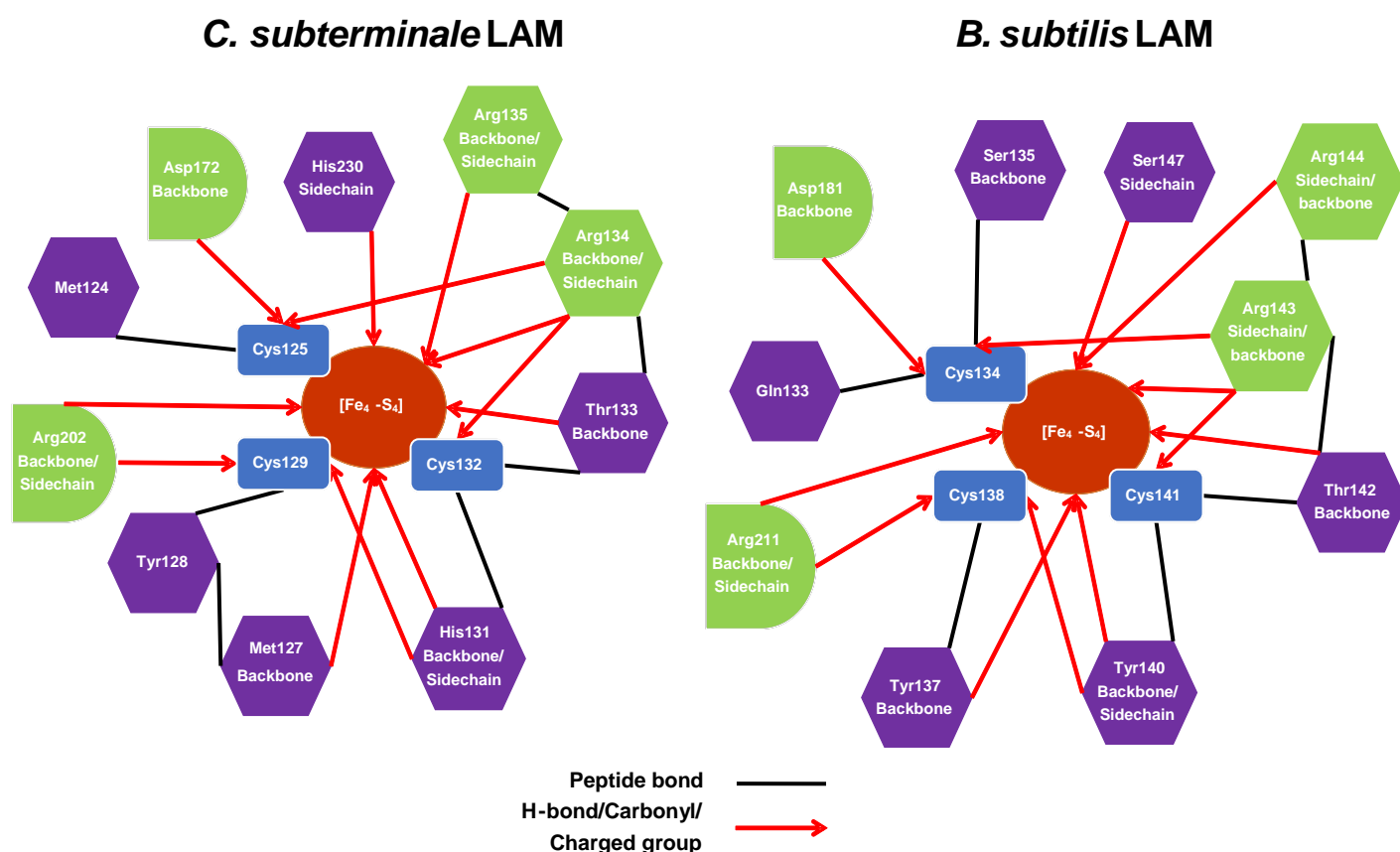
Scheme 2. Maps of the identified redox-modulating residues in *Cs*LAM (left) and *Bs*LAM (right). The [Fe$_4$-S$_4$] cluster is at the centre of the map (orange), the cysteine ligands (blue), the H-bond- and CO-interacting residues (purple) and the charged residues (green). The adjacent residues are linked together by the black peptide bond whilst the electrostatic interactions are represented by the red arrows.

were found within 4.0 Å distance from the metal cluster and cysteinyl ligand S atom (Appendix 6.29). Residues covalently bound to the cysteinyl ligands were easy to identify as well as other residues that were found interacting with the [Fe$_4$–S$_4$] cluster through their hydrogen-bonding atoms (-NH and -OH moieties either from their backbone or sidechain) within the selected distance. The residues Thr133, Asp172 and Arg202 in *Cs*LAM and their corresponding positions in *Bs*LAM, were amongst the relevant residues that could be identified by visual inspection of the environment around the [Fe$_4$–S$_4$] cluster. Other amino acid residues appeared within the interacting distance with the metal cluster by other non-interacting atoms, such as the Cβ of Ser147 in *Bs*LAM that at the same position in *Cs*LAM corresponded to Ala138. The MD simulations later revealed that the hydroxyl group of Ser147 oriented towards the [Fe$_4$–S$_4$] cluster (see Appendix 6.30), therefore it showed it was likely to interact with the metal cofactor and it was identified by the hydrogen-bond screening with *cpptraj*.

4.2.5 Concluding remarks

The QM calculations of the metal clusters provided the electronic energy of different subsystems that were analysed along the MD trajectory for both the sensitive and tolerant LAM enzymes. QM calculations of the simple FeS cluster with SCH$_3$ ligands highlighted a small difference in the electrostatics between the active site in the two enzymes but that was not large enough to account for the difference in oxygen sensitivity of the enzymes. The VEA describing the vertical electron affinity of the metal cluster with thiolates showed minimal difference between the two enzymes and such a difference was confirmed as statistically irrelevant. Thus, the small VEA fluctuation in over the MD simulation could be accounted for by the different orientation of the ligand dihedral angles. The difference in oxygen sensitivity in the lysine 2,3-aminomutase seems to likely reside in the presence of one or more nearby key residues that are able to increase the FeS cluster midpoint potential. By following the rules for hydrogen-bonding network search proposed by Harris and Szilagiy,[176] the relevant

hydrogen bonding residues were able to be grouped into three coordination spheres. There was some divergence in the identified residues of the two enzymes because of the slightly different protein environment. Taking the same coordination sphere for each enzyme, the spheres provide hydrogen-bonding networks because unshared residues interact with the FeS cluster to generate a different electrostatic potential at the active site. The identification of this network for each enzyme allows the identification of the most relevant redox tuning residues between the two enzymes. Most of the amino acid residues of redox-tuning interest were identified by the detailed inspection of the active site together with the application of the set of rules employed to map such amino acids. Other amino acid residues were instead found or confirmed by the hydrogen-bond screening of the MD simulations trajectories of LAM enzymes. This approach was useful for the identification of other redox-modulating amino acids that could act as mutagenesis target. With these differences in hand, selective creation of *Cs*LAM mutants with increased oxygen tolerance should be possible, if this protein region is indeed the controlling factor. Another source of analysis is required to demonstrate the role that key residues might have on the electrostatic environment of the LAM and this approach is discussed in the next section.

4.3   *Cs*LAM mutants creation and the dipole moment approach

4.3.1   Creation of *in-silico Cs*LAM mutants

Following the construction of the active site residue maps of both LAM enzymes to visually compare the two protein environments a few residues belonging to *Bs*LAM were identified as likely redox-modulating amino acids. The list of mutations was engineered by taking into account previous mutagenesis experiments reported in the literature,[222] in order that the specific positions did not host functional residues that may affect catalysis. The new amino acid residues were substituted at the same position in the native *Cs*LAM to generate two single- and two double-substituted *in silico* hetero-electronic mutants. Information relating

their dynamics and electrostatic variability with respect to the native forms of the lysine 2,3-aminomutase is presented in Table 3.

The monitoring of a varied active site electrostatic field of such mutants should provide indirect information about the reactivity towards oxygen of the enzyme mutant, namely whether the impact of this field will be more similar to either the oxygen-sensitive or tolerant LAM. In this manner, it will be possible to give an estimation of which LAM mutant should be more resistant to the attack of radical oxygen species.

The *Cs*LAM mutants included single mutants carrying the substituted M124Q and H131Y, and the doubly substituted mutants M124Q/H131Y and the H131Y/A138S and compared the impact of the different number of substitutions on the enzyme. The different residues modify the electronic environment in the *Cs*LAM active site either by increasing the hydrogen bonding network and bringing in a new moiety that interacts with the electronic structure of the FeS cluster. New N-H—S interactions from the backbone or new –OH—S hydrogen bonds created by the sidechain, are likely to elevate the midpoint potential of the [Fe$_4$-S$_4$] cluster. To create the modifications, the substitution at position 124 of *Cs*LAM removed the methylthiolate moiety of methionine and added the amide moiety of glutamine.

The change increased the hydrogen bonds because of the amide functional group and added a new carbonyl in the nearby of the metal cluster that could affect its electrostatics, although the sidechain of Q124 seemed not to face directly to the metal cluster and the effects on the electrostatics might be low. At position 131, tyrosine brings a phenol group replacing the imidazole sidechain of histidine so that a new and strong hydrogen bond is present in the cuboid cluster surrounds. Similarly, the mutation at position 138 adds a new hydroxyl group onto the alanine to generate a new hydrogen-bond effect. The H131Y and A138S replacements bring a new -OH moiety within the [Fe$_4$-S$_4$] cluster coordination sphere that, accordingly to the work by Denke *et al.*,[81] they might have a strong and positive influence on the overall electrostatics. All these substitutions bring new factors to the *Cs*LAM environment that are consistent with the set of rules developed to identify the redox modulating residues

Table 3. Designed single and double substituted hetero-electronic *Cs*LAM mutants

*Cs*LAM WT	*Bs*LAM
_Cs_LAM M124Q	**_Cs_LAM H131Y**
_Cs_LAM M124Q / H131Y	**_Cs_LAM H131Y / A138S**
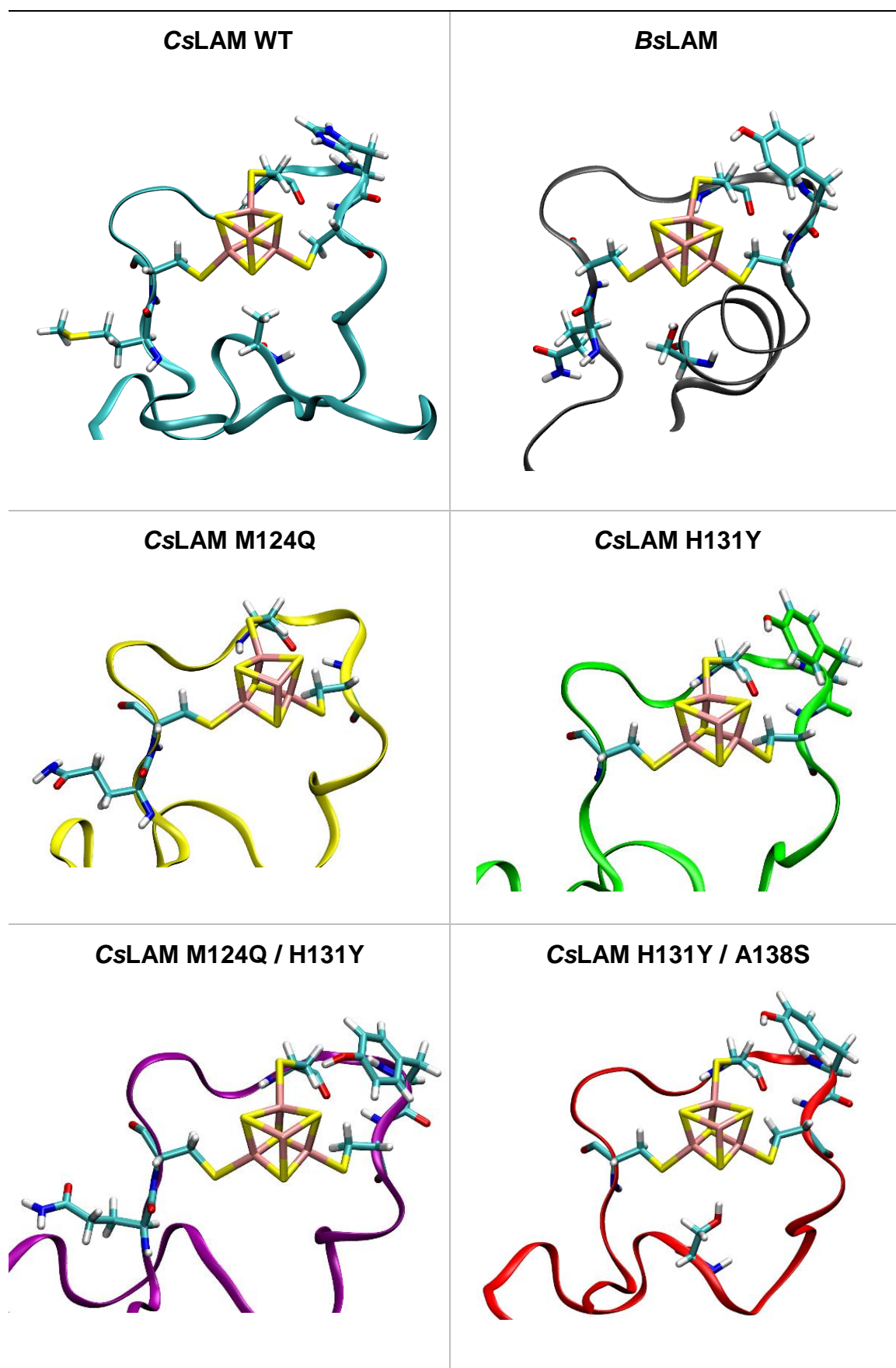
and the engineered mutants should therefore show an increased tolerance to oxygen.

The enzyme mutant model building procedure is easy to achieve in Amber. Starting from the PDB file of the protein of interest, all the positions of the amino acid residues to be mutated were identified over the multiple chains of the *Cs*LAM enzyme. To mutate each residue, the lines contained in the PDB file that describe each atom composing the amino acid were deleted, leaving only those lines that describe common atoms between the native residue and the new one. In addition, the relevant amino acid three-letter codes must be changed according to the new residue names (Appendix 6.31). The lines of the sidechain in the PDB file of the native residue have to be removed up to the first atom that is found to be in common in both the amino acid residues. The overall protonation state including the customised one for the histidine residues, was the same as that chosen for the wild type *Cs*LAM MD simulation. Once all the editing was complete, the mutated PDB file was fed into *xleap*, contained in the Amber software. From the creation of the initial structure to the analysis of the MD trajectories, the same workflow that was previously done for the MD simulation of *Cs*LAM was employed.

4.3.2 Dipole moment orientation of LAM enzymes

The study of the impact of electrostatic variation among the LAM enzyme active sites needed to be achieved by another strategy. Detailed information about the electronic structure of the iron-sulfur cluster and the redox-modulating residues was difficult to obtain via direct QM calculations. At this stage of the project, a new strategy was necessary in order to derive information about the electrostatic field that is experienced by the $[Fe_4\text{-}S_4]$ cluster in the active site of the LAM enzyme. The electrostatics are meant to impact on the reactivity of the metal cluster, thus to make it more or less susceptible to oxidative damage. The different reactivity might also drive the speed of the enzyme catalysis as a consequence of a different redox potential, as it was observed that *Bs*LAM had concomitant low activity.[49] More insights are needed about the variation of the reactivity of the prosthetic group between the wild type

LAM enzymes and the engineered mutants to estimate a change in the electrostatics of each active site. An investigation on the electrostatics could provide insights that correlate to the redox potential of the $[Fe_4\text{-}S_4]^{+2/+1}$ couple in each enzyme, although this is not a direct measure of the midpoint potential itself. This method aimed to give a relative estimation of the redox potential between LAM enzymes. This comparative analysis was performed to look at differences among each active site with the expectation that a varied electrostatic field experienced by the FeS cluster between wild type enzymes and with the designed mutants would result in a crude approximation of the relative redox potential between these LAM enzymes.

The approach chosen to conduct the electrostatic analysis was to observe the variation of the resultant dipole moment generated by the redox-modulating residues identified in the previous section. The variation of the dipole moment provided a preliminary, rough estimation of the different electrostatics, hence a crude approximation of the relative redox potential between the LAM enzymes. The polarizability of the protein was not considered nor the solvent interaction with protein and prosthetic group, as previously reported in the literature.[78] The outcomes from Section 4.1 revealed a very similar solvent accessibility in the active site of the two LAM enzymes, thus in order to make a comparison of the electrostatic environment, it could be disregarded from the analysis as it was considered to be comparable. Furthermore, the aim of this investigation was to estimate differences in the protein environment that largely contributed to the FeS cluster electrostatics and not the measure of the absolute redox potential, which requires solvent accessibility to be taken into account. The DFT+PB calculation strategy would have been more accurate as reported by the lab of Prof Ichiye,[223,224] but at that time of the project we had to rely on the dipole moment variation method to compare the hydrogen-bonding residues that interact with the FeS cluster in the two LAM enzymes.

The dipole moment that originates from the mapped residue selections provided an insight on the overall electrostatic environment in the LAM active site. The difference in dipole

moment between oxygen sensitive and tolerant LAM would result in the electrostatic potential within the catalytic pocket being different and, in turn, would impact the electronic structure of the FeS cluster and therefore the midpoint potential. The presence of electron donating or withdrawing groups in FeS proteins active site modify the overall charge of the cluster to influence the redox potential of the FeS proteins, as seen for Rieske proteins[225] and Ferredoxins.[226] The interaction of each hydrogen-bonding network residue's dipole moment with the cluster influences the electron charge density of the prosthetic group. The total dipole moment for those residues and its variation can account for a varied redox potential. The contribution to the electrostatics arising from the SAM molecule, the PLP cofactor and the substrate are assumed to be of the same magnitude, thus these contributions were not included in our analysis because of the consistency between enzymes.

The interaction between the overall dipole moment of the selected amino acid residues and the $[Fe_4-S_4]$ cluster required a reference point that would describe the metal cluster behaviour with respect to the selected electric dipole moment. As such, the cluster vector was defined as the normal vector from the plane that divides the cuboidal cluster in half and intersects along the space diagonal (Figure 39a), described by the cluster atoms: Fe2 - S2 - S3 - Fe3. The vector perpendicular to the plane, forms an angle with the dipole moment arising from all the redox-modulating residues (Figure 39b). The variation of the angle between the total dipole moment and the cluster vector, is unique for each of the simulated LAM enzymes. The angle between the dipole moment and the $[Fe_4-S_4]$ cluster vector is thus able to provide an approximation of the electrostatic forces and their orientation against the FeS cluster. The intersecting angle was assessed for the duration of the MD simulation of each LAM system. The strength and orientation of the electrostatic field should differ between the two native LAM enzymes, and similar angles that would be found in the LAM mutants may confer similar electrostatic properties that may affect air tolerance and provide an estimation of the effect.
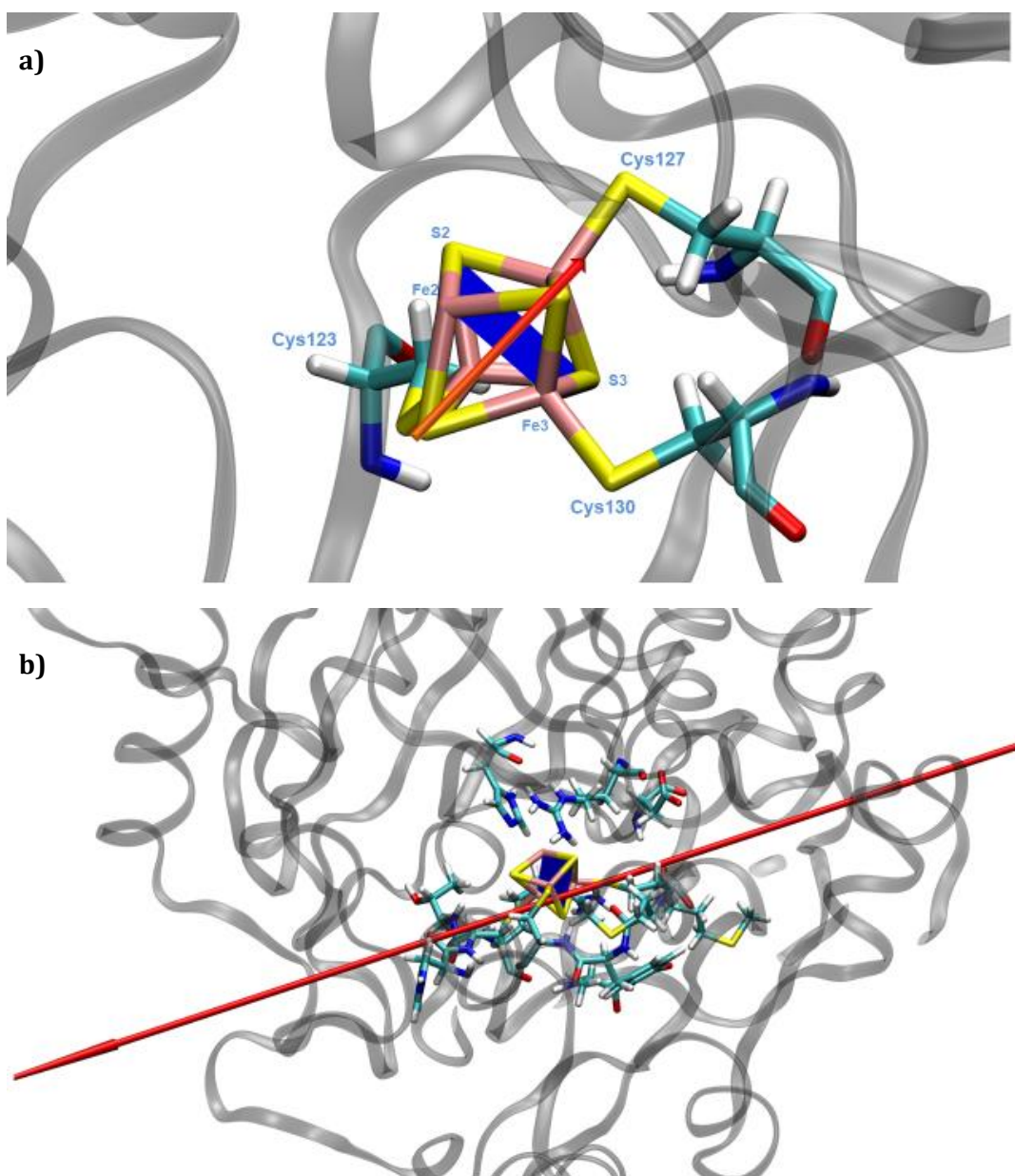
Figure 39. a) Detail of the plane defined within the FeS cluster (in blue) and the square vector that interacts with the dipole moment (the arrow in orange). b) The dipole moment (red arrow) arising from the selected redox-modulating residues shown in licorice. Images were produced using the dipole watcher in VMD.

4.3.3   Cluster vector - dipole moment vector angle analysis

The angle between the vector normal to the cluster plane and the total dipole moment of the redox-modulating residues in the protein environment was analysed with the *cpptraj*

software in Amber. The protein-induced dipole moment and the FeS cluster vector normal to the plane were mapped over the 400ns of MD simulation time scale. The vector perpendicular to the cluster plane, represented by the four atoms as shown in Figure 39a and the centre of mass of the atoms, was used as the reference point and was automatically normalised. The variation in the non-normalised dipole moment vector was thus obtained to assess the influence on the cluster and how it interacts with the previously defined cluster plane, as shown in Figure 39b. By using a custom python script (Appendix 6.33), the angle between the dipole moment and the vector normal to the cluster plane was calculated to monitor its variation over the 400ns of MD simulation for each enzyme variant.

Results shown in Figure 40 highlight the difference in the angle variation between the two vectors in the wild-type LAM enzymes. The *Bs*LAM recorded an angle between 5° to 50° over the four monomers but we could observe the most reliable monomers **B** and **D** to assume values around 20°. On the other hand, the angle recorded for the *Cs*LAM enzyme fluctuated with values around 55° in monomer **A** and from 20° to 40° in monomers **B**, **C** and **D**. Such a fluctuation over the four monomers in *Cs*LAM might be due to the different pH settings of the histidine residues, His131 and His230, that were included in the redox-tuning residues selection of *Cs*LAM (as it was previously discussed in section 3.4.1). The histidine protonation states might have affected either the dipole direction or its magnitude. This assumption was consistent with the lower fluctuation of the dipole moment in the four monomers of the oxygen-tolerant *Bs*LAM where the relevant histidine had the same protonation state throughout the tetrameric structure. Despite the oscillation in the reported angles, two different values were observed as shown in plots a) and d) of Figure 40. The angle reported for monomer **A** in *Bs*LAM (Figure 40a) started from about 25° in the early 150ns of simulation, then decreasing to a mean of 10.7° ± 3.9 whilst the angle in *Cs*LAM steadily moved from 50° to a mean of 59.2° ± 5.25 over the same time interval. A similar trend could be observed in the first 210 ns of MD time in monomer **D** (Figure 40d), where the angles in *Cs*LAM increased to 20° and it assumed an angle of 6.2° ± 4.4, whilst the *Bs*LAM moved from 30° ± 8.6 in the

first part of the simulation then it decreased to 18°. In monomers **B** (Figure 40b) and **C** (Figure 40c) both the angles in the two enzymes had similar, minor fluctuation around the values of 20°. As evaluated previously from section 3.4.4 the monomer **B** of *Bs*LAM was the most reliable of the MD simulations, and so it was in this analysis. Two distinct angles in the *Cs*LAM enzyme were mainly visible in monomers **A** and **D.** The histidine protonation state in monomer **A** that was suggested by H++ and mirrored to form monomer **D**, represented the best structure in terms of structural stability from the analysis in section 3.4.2. Therefore, monomer **A** in *Cs*LAM was chosen to be the most likely *scenario* for the analysis, in contrast to monomers **B**, **C** and **D**.

The angles of 55° and 20° were chosen as the reference values for an oxygen-sensitive and
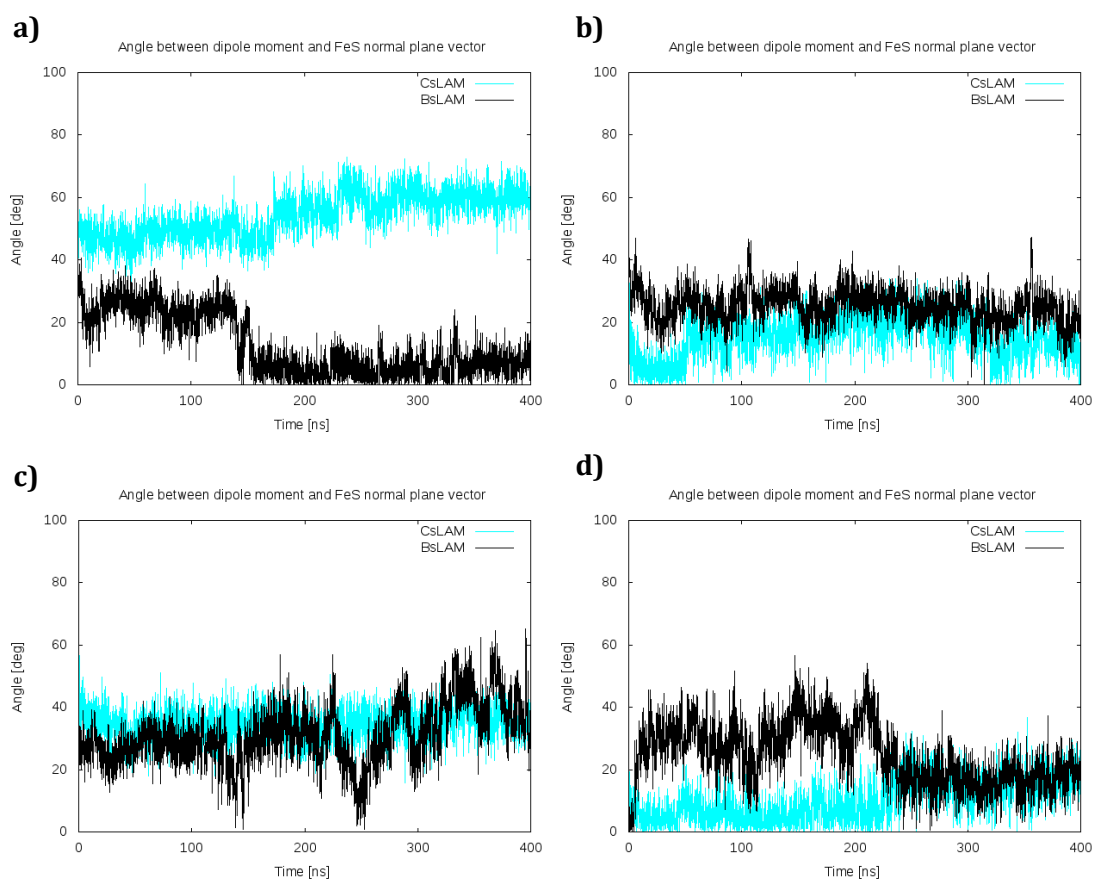
Figure 40. Dipole moment orientation originated from the redox-modulating residues selection and the cluster vector in both wild-type LAM enzymes. From a) to d) the corresponding enzyme monomers are represented; *Cs*LAM is reported in cyan, *Bs*LAM in black.

oxygen-tolerant protein environment, respectively, based on results from monomer **A** and **B** (Figure 40) of each of the LAM enzymes. The difference of 35° between the two angles in the native LAMs marked two different orientations of the dipole moment with respect to the cluster vector that might it be consistent with a different midpoint potential of the $[Fe_4 - S_4]$ cluster in the two enzymes. The angle between the dipole and vector thus, provided an approximated solution to compare different electrostatic environments within the LAM active sites. The usage of MD replicas would have provided more data to average together so to gain more accurate insights but, once again, we were aware about the limited reliability of this last approach.

The same analysis was also run for all the *Cs*LAM engineered mutant enzymes, shown in Table 3. The variations of the dipole moment orientation set by the two vectors for the mutants is illustrated in Figure 41. Throughout the four monomers of the enzymatic structures, it was possible to observe the different values the dipole angle explored during the MD simulations. The mutants, as well as the native *Cs*LAM enzyme, had different options for protonation state of the active site residues His131 and His228, with the suggestion from H++ website implemented, as per monomer **A** (Figure 41). Only the *Cs*LAM M124Q mutant had retained both histidine residues so the same His131-His228 pH was used as had been in the wild-type *Cs*LAM. In monomers **B**, **C** and **D** (Fig. 41) the angle variation recorded for all the enzyme systems spanned over a significant range of values (by to 0° to 60°). That was associated to the histidine protonation state on these monomers that might have provided an overriding influence on the dipole moment, resulting in it assuming a similar orientation in all the *Cs*LAM enzymes. From the MD simulation analysis reported in Appendix 6.34 – 6.37 it was not observed any significant fluctuation among the composing monomers of LAM mutants. That let us believe that different protonation states of His228 alone did not impact the tetramer stability or residue fluctuation but it was likely to have influenced the dipole moment as such to assume the similar orientation in all *Cs*LAM enzymes. We considered in detail the properties of all monomers to determine which is the most representative, with a detailed

discussion below instrumental in identifying the key considerations and comparisons. Informed by the details below, monomer **A** was considered the most representative for the purpose of further analysis.

Due to the difficulty in distinguishing a reliable trend, the analytical focus was put on the **A** monomers of each variant. The dipole moment orientation of the M124Q mutant (Figure 41, in orange) moved from 65° ± 5.5 in monomer **A** to 30° in **C** and 10° in **B** and **D**. This variation across the four monomers, with an angle value in monomer **A** being quite different from the other monomers, was consistent with the angles previously analysed in the wild-type *Cs*LAM. The M124Q substitution only added a carboxylate moiety in the cluster surrounding and that may not be enough to affect the midpoint potential. This could help rationalise the consistency
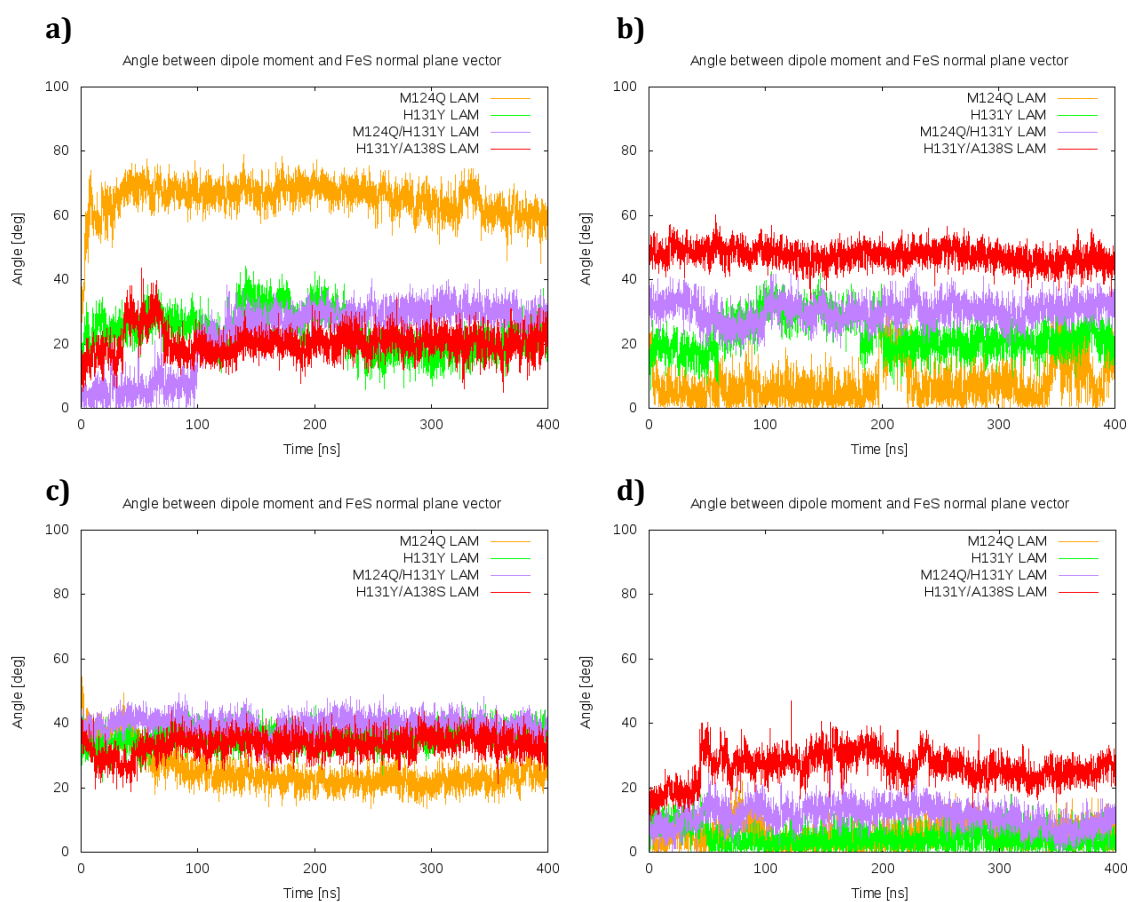
Figure 41. Included angle between the dipole moment and the cluster vector. From a) to d) are represented the angle trends for LAM mutants M124Q (orange), H131Y (green), M124/H131Y (purple), H131Y/A138S (red) over the four monomers of the enzyme structures.

with the value obtained for monomer **A** in *Cs*LAM (55°).

The H131Y mutant angle (figure 41, green) assumed similar values throughout the four monomers with a small fluctuation of 24° and 22° in monomers **A** and **B** respectively, 36° in **C** and 5° in **D**. This mutant had only one histidine residue within the active site that was in a different protonation state between the monomers. As there was only one histidine present, the dipole moment was still governed by the overall dynamics of the residues, and in this case there were not different combinations of the pair of histidine residues. The different electrostatic environment due to having only one histidine might have not been strong enough to affect the angle orientation as much as for the native LAM and the M124Q mutant. That was consistent with the low degree of angle oscillation of H131Y mutant over the four monomers. The presence of another tyrosine close to the metal cluster influences the charge and electrostatics around it, either through interaction with the electron density of the aromatic ring or by adding a hydroxyl moiety to increase the hydrogen-bond network. The angle for the H131Y mutant in figure 41a was close in value of the angle in the same monomer of *Bs*LAM consistent with the influence of the phenol moiety from the tyrosine substitution and its possible effect on the dipole moment and the electrostatic environment on the metal cluster. The similarity in orientation of the electrostatic field because of the common tyrosine in the active site, might indicate an elevated midpoint potential for this mutant.

The doubly substituted M124Q/H131Y *Cs*LAM mutant, reported in purple in figure 41, afforded an average angle of 30° ± 4 for the majority of the simulation in monomers **A** and **B** whilst it increased to 40° in **C** and it decreased to 10° in **D.** The mutant, as well as the single mutant H131Y, only carried one histidine residue, which implied a change in protonation state close to the metal cluster. Perhaps because of the presence of just one histidine residue or because of the combined effect with the mutated residues, this doubly mutated enzyme had angle values similar to the *Bs*LAM but less similar than those angles for the H131Y mutant. For example, for monomer **A,** the angle had a slightly higher values, by ~20° than the oxygen-tolerant LAM and higher by ~10° than the H131Y mutant, yet lower than in *Cs*LAM and

M124Q mutant. The angle seemed to suggest that the electrostatics were similar to the *Bs*LAM but with still some modifications possible to bring it in line. With the dipole angles as a proxy for redox activity, this double mutant was possibly less effective than the H131Y mutant.

The H131Y/A138S double mutant had one differently protonated histidine residue close to the metal cluster. The dipole (red in Figure 41) assumed an angle of 20° in monomer **A**, similar to the *Bs*LAM. Among the other monomers, the angle orientation increased to 50° in monomer **B** and it remained steady at 30° in both monomers **C** and **D**. In this mutant the double hydroxyl moiety carried by inserting a tyrosine and serine residue might positively affect the dipole moment of the active site by increasing the hydrogen-bond network. The orientation was consistent, in trend and values, with the angle reported for *Bs*LAM in the corresponding monomers **A**, **B** and **D** and more consistent than the equivalent values for the H131Y mutant. Considering monomer **A** in Figure 41, we can see two different values were identified for the majority of the simulation in *Cs* and *Bs*LAM and might imply a different redox potential for the metal cluster on the basis of electrostatic influences. Monomer A of the M124Q LAM mutant (Fig. 41, orange) seemed to have the dipole moment orientation quite similar to the native *Cs*LAM, consistent with the presence of only one substitution whose chemical structure did not affect the overall electrostatic much. Either the presence of the amide moiety of glutamine in adding favourable N-H—S bonding was insufficient or its orientation did not have an impact on the dipole to confer oxygen-resistance.

Mutants H131Y and the doubly substituted M124Q/H131Y both had very similar variation and angle values adopted through the overall simulation. Both the mutants only carried one histidine residue close to the cluster that might affect the angle but their average values were closer to the *Bs*LAM than the M124Q mutant. As their common mutation replaced the histidine imidazole ring with the tyrosine aromatic ring with hydroxyl moiety to increase the hydrogen-bond network, a similar behaviour to the *Bs*LAM is expected. Such modifications were consistent with the increased hydrogen-bond network that should have affected the metal cluster electrostatics to increase the redox potential. Experimental site-directed

mutagenesis experiments carried out on this *Cs*LAM mutant with midpoint potential measures would be the next step to establish the oxygen tolerance of this H131Y mutant. The doubly substituted H131Y/A138S LAM also introduced a phenol ring, adding a hydroxyl moiety on top of that alanine residue substituted by serine close to the cluster. As such, two more hydrogen bonds were established that could interact with the S atoms of the prosthetic group (Figure 42). The angle values reported were very similar to the *Bs*LAM fluctuations, and more similar than the other tyrosine-containing mutants. This increase in hydrogen bonding content mirrors the reverse of the experiments done with Rieske proteins where serine and tyrosine were replaced with alanine and phenylalanine respectively to remove hydrogen bonding and result in a lower redox potential and increased oxygen sensitivity.[81] This implies by analogy an increase in the redox potential of the FeS cluster, and thus improvement in oxygen tolerance is a likely outcome from the H131Y/A138S mutant. The tyrosine at position 131 is very likely to positively affect the redox potential and the A138S mutation seems to furtherly increase the effect of raising the LAM reduction potential and possibly, its tolerance to oxygen.

If the dipole angle can be experimentally verified as influencing oxygen tolerance, based on the above mutants, this *in silico* approach can be seen as a good tool for the rapid approximation of relative midpoint potentials, dramatically reducing the time required to screen for oxygen-tolerant mutants. This would be an excellent addition to the biotechnological toolbox to enhance the utilisation of radical and other metallo-enzymes, and provide a mechanism to generate enzymes robust to industrial conditions and provide greener reaction routes in the future.
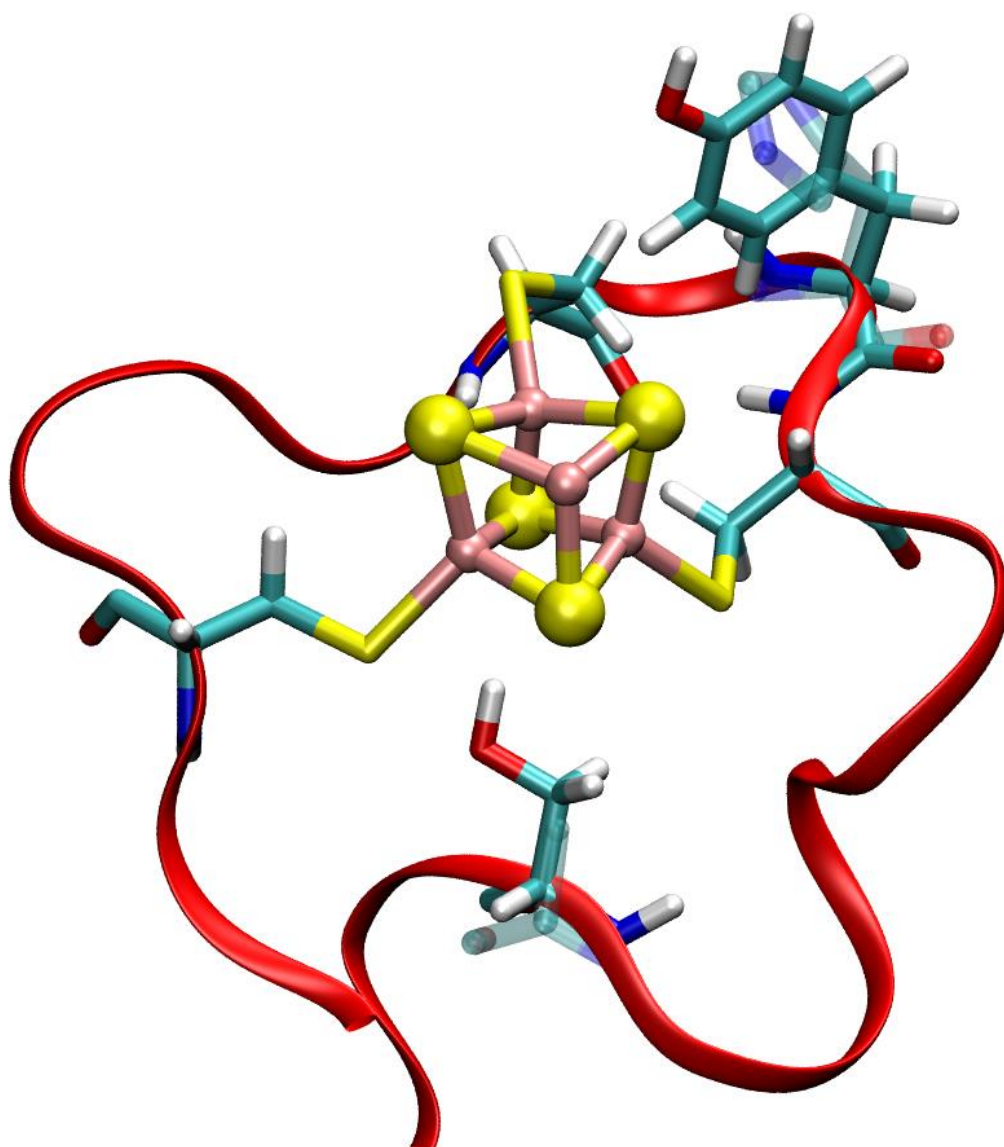
Figure 42. The active site of the *in-silico* LAM mutant H131Y/A138S with native His and Ala residues represented in transparent liquorice for comparison.

5 Conclusion

Radical SAM enzymes are able to catalyse complex biochemical reactions involving organic radical intermediates that can lead to the formation of compounds of pharmacological interest that are complex to achieve with traditional lab techniques. Therefore, enzymes belonging to the radical SAM family are all potential biocatalysts for use in biotechnology, including the focus of this study, lysine 2,3-aminomutase (LAM), which can catalyse the industrially interesting conversion of the common α-lysine into β-lysine. Two variants of LAM that belong to the *S*-adenosylmethionine (SAM) radical enzyme superfamily, specifically the LAM enzymes either from *C. subterminale* or *B. subtilis* bacteria, have been compared with a view to establishing effective principles for inducing oxygen-tolerance. The radical reaction mechanism that is typical of this class of enzymes, is driven by the cuboidal [Fe_4-S_4] cluster that is coordinated by three cysteine residues, whilst the fourth ligand is represented by the SAM molecule. Classically, upon reduction of an electron, the metal cluster homolytically cleavages the S – C bond of SAM creating the dAdo• radical species that attacks the substrate and triggers the complex molecular rearrangement that will provide the final product. This elegant and precise mechanism is affected by the presence of oxygen and other reactive oxygen species (ROS) that that react with the metal cluster in its reduced state by oxidising and decomposing the cofactor, hence preventing the enzymes from working under a regular atmosphere. SAM enzymes that can naturally perform their catalysis in air but at the expenses of product yield, are rare, but provide a starting point for engineering oxygen tolerance for wider use. One such example is LAM from *B. subtilis* (*Bs*LAM). *Bs*LAM has had an activity previously measured to be 1.7% that of the oxygen-sensitive LAM from *C. subterminale* (*Cs*LAM)[49] and represents a fascinating case of study to figure out how the FeS cluster may be stabilised to work in presence of air. If tolerance to oxygen can be artificially transferred to other enzymes by site-directed mutagenesis and new radical biocatalysts that work at regular atmosphere engineered, this will broaden the range of catalytic reactions accessible and allow

advancements in biotechnology substantially beyond the state of the art by reducing the amount of solvent, costs and energy for synthetic chemistry steps, and avoid inefficiently expensive use of inert atmosphere..

The described study of the lysine 2,3-aminomutase employed computational methods, both because of the relatively unstable nature of these radical enzymes and the complexity incurred by use of an inert atmosphere, and because computational approaches often provide an enhanced and more rapid approach to screening new enzymes, once established. The molecular dynamics simulation of LAM enzymes required the use of the 3D structure solved by crystallography of *Cs*LAM, whilst the *Bs*LAM model had to be generated through homology modelling because its structure is unknown. Validation of the homology modelling approach was first established by monitoring the consistency of MD simulations of a known pair of the SAM-radical enzyme THIC, where a homology model had been generated for one of the known crystal structure, and comparisons were made to the other crystal structure. This method showed good reliability of the generated structure, with poor quality parts of the model relating to sections that were created by *ab initio* modelling. These poor quality areas applied equally to the crystal structure and the generated model.

Once this process was established, the *Bs*LAM model was created and validated against the *Cs*LAM crystal structure. After establishing a similar level of reliability for the *Bs*LAM model, 400ns of MD simulations of both LAM enzymes were carried out and no instabilities identified. Fluctuations in both LAM enzymes primarily related to whether or not they contained the SAM molecule in the catalytic pocket, consistent with what presented in the literature about the SuiB[206] and the CteB[210] radical SAM enzymes. The overall outcome was good enough to make further use of the trajectories to look for structural differences between the two enzymes.

The tunnel-searching software CAVER was used to figure out if the air flow throughout the *Bs*LAM was somehow obstructed, impeding oxygen approach to the deep buried FeS cluster, as have been previously seen in the fumarate and nitrate reduction regulatory (FNR)

enzyme.[58] The employment of CAVER also helped with looking at the nature of the residues that composed the tunnels and whether they were able to interact with and scavenge oxygen molecules. No structural differences nor scavenging residues were detected leading to the catalytic pocket, where the FeS cluster could be oxidised and decomposed, between the two LAM enzymes. The redox potential of metal clusters has been previously speculated to be altered by the electrostatic field in the active site, thus increasing the cluster midpoint potential and imparting tolerance to oxygen. Selected key residues were identified in the *Bs*LAM active site and the difference in the electrostatics in the two LAM enzymes calculated, initially by considering solely the $[Fe_4-S_4]$ cluster and the redox-modulating residues using QM/MM methodology. The calculations did not work efficiently and many issues were encountered, therefore structural differences were calculated by performing QM calculations of the $[Fe_4-S_4](SCH_3)_3$ systems taken from different timepoints of the MD simulation. The QM calculations of these reduced models in the absence of protein environment identified slight structural differences among the systems in the two enzymes but without statistically significant difference in their midpoint potential. *Cs*LAM mutants were subsequently created carrying some of the potential redox-tuning residues identified in the *Bs*LAM structure through the mapping approach of *Harris* and *Szilagyi*.[176] An approximation method was developed that identified the product between the dipole moment vector arising from the selected residues of redox-tuning interest and an internal plane in the cluster affording an estimation of the electrostatic field in each LAM active site.

The electrostatics potential, which has been related to the midpoint potential,[78] could be modified by mutations of His131 to Tyr131 and Ala138 to Ser138, consistent with the concept that this could mediate the midpoint potential. This approximated *in-silico* method shows promise as a screening tool and points to key mutations that can be introduced in the *Cs*LAM mutant, after which experiments can validate whether these changes increase its tolerance to oxygen, consistent with similar results reported from the literature.[81]

With a focus on the investigation of the lysine 2,3-aminomutase and its modification, the knowledge and the method employed is envisaged to be transferrable to other radical SAM enzymes, perhaps with the programme to develop a library of new radical SAM biocatalysts. So far, the other SAM enzyme with a proved natural tolerance to oxygen was found in the PqqE from *M. extorquens* which is the enzyme catalysing the first reaction of the PQQ (pyrroloquinoline quinone) biosynthesis pathway.[227] Recently, its crystal structure was solved and published,[228] and there is potential to use this protein as an additional case study to further identify possible structural features that make it tolerant to oxygen, with the advantage that there would be no need to create a homology model of the protein.

In this computational project, the basis has been laid for a new understanding of how to manipulate oxygen tolerance in radical SAM enzymes. By exploring already established factors contributing to iron-sulfur cluster stability: namely (1) steric blocking of the active site tunnels,[56] (2) chemically reactive residues preventing ROS reaching the active site,[57] (3) structural alterations in the geometry of the iron-sulfur cuboidal cluster resulting in altered VEA as a proxy for midpoint potential,[66] and (4) the electrostatic impact arising from the protein environment as measured by the enzyme dipole and acting as a potential proxy for midpoint potential;[176] mechanisms. (1) – (3) have been ruled out for the comparison of *Cs*LAM and *Bs*LAM. In this case, the difference in protein dipole is a likely candidate for the altered oxygen tolerance in line with the increasing evidence for protein electrostatics playing a significant role, in mediating catalysis.[229] Thus, this work, once validated experimentally, helps to contribute to the broader picture of how catalysis is mediated, not just in radical SAM enzymes, but also in the broader milieu of metallo-enzymes. Over the years, more energies were spent on the elucidation of other enzyme mechanism and their improvements such as, the hydrogenase. [NiFe]- and [FeFe]-hydrogenases were the main subject of many studies in terms of oxygen resistance improvements, because of the interest in the H_2 bioproduction as a new sustainable technology.[230]

This methodology shows promise as a screening technology for *in silico*-generated mutant enzymes, allowing a rapid approach to assessing such mutants for their applicability within biotechnological systems and allowing targeted experiments in a low-waste manner. As such, the successful demonstration of this workflow has broader implications in inspiring related computational-based advances for other enzyme-catalysed systems, ensuring a broader accessibility of novel catalysts to improve current chemical process.

5.1   Future perspectives

To enhance the outputs of this work in future, more detailed QM calculations of the FeS cluster surrounded by the redox-tuning residues in the active site could provide more precise calculations informing on the electrostatic field in each LAM enzymes. As was established by previous workers,[231-232] different force fields or atom parameters are also likely to provide improvements and a more detailed view of the enzyme. This could be approached as follows: QM calculations could be run starting from a lower level of theory so to avoid clashes, since deformation of the molecular systems used and SCF convergence issues seemed to be among the main problems in the QM calculations performed. A new software for semiempirical molecular orbital calculation, able to treat large systems could be employed. EMPIRE has been recently released[233] and, because it does not make use of local approximations, it is suitable for electroactive systems like radical SAM enzymes. As it can study large molecular systems it could be used to investigate properties of proteins, such as the electron-transfer in redox-active enzymes. On this basis, other ways to calculate the absolute redox potential could also be developed. In particular, a quick and reliable technique for the absolute redox potential, as reported in Fricker's work,[234] would either negate the need for or validate the proxy methods explored in this work. In addition, another new piece of software that is able to generate the electric field in enzymes, e.g. TITAN,[235] could be employed to better investigate enzyme reactivity and expand the knowledge for radical SAM enzyme reactivity.

The study of the electrostatics forces within the active site and the dipole moment could be improved by the calculations and analysis of the HOMO-LUMO orbitals gap of the metal cluster in LAM (or other radical SAM enzymes) to elucidate the mechanism of charge transfer during catalysis.[236] The study of the dipole moment of the SAM or substrate upon reductive cleavage might clarify the contribution to catalysis and how to speed it up by taking into account the interaction of the protein environment during the enzyme turnover as reported by Fried *et al*.[237] Other studies on the electrostatic field experienced by the prosthetic group and its modulation to increase the catalytic activity could also be run in case the newly engineered *Cs*LAM mutants resulted in a lower activity compared to the wild type. An example of this modulation is given by the application of an external electric field on a (bio)catalyst.[238,239]

One drawback identified in the current set of calculations was that the PLP cofactor was difficult to model and inclusion of this cofactor and the lysine substrate in LAM could help stabilise the enzyme fluctuations and provide a more realistic assessment of the dynamical properties of the enzyme when primed for catalysis. The inclusion of these additional small molecules would help to further establish the roles of the loops and other structural parts of the enzyme involved with the shuttling and control of the cofactors in the enzyme

Lab production and assessment of the already identified mutants provides the ultimate validation of modified sensitivity to oxygen and outcome on activity. This would be the final test to establish the effectiveness of our *in silico* generated LAM enzymes. Once this is completed, the engineering of a broader range of isolectronic LAM mutants, and a streamlined workflow to generate and test these either *in silico* or experimentally, building on the current work. That will help with the study of the multipole reorientation within the active site to better interpret the impact of reorganizational properties of the inner multipole moment and observe how the electrostatics influence catalysis.

A preliminary look at the sequence similarity network (SSN) of LAM search was conducted to find out if other enzyme sequences closely related to the *Bs*LAM family could be found that

may also exhibit oxygen tolerance. Sequences were obtained from the Structure-Function Linkage Database (SFLD, http://sfld.rbvi.ucsf.edu/).[240] This database contains considerable and detailed information about the radical S-adenosylmethionine (SAM) superfamily, which can be found on http://sfld.rbvi.ucsf.edu/django/superfamily/29/. The databases used were structure-function linkage networks specifically for the LAM and LAM-like domain that can be found on http://sfld.rbvi.ucsf.edu/archive/django/family/283/. The networks were downloaded with restrictions of 250K edges at a BLAST bit score cutoff of 60. In metagenomics, the edges indicate the total number of connections among proteins in the database and they are built on the similarity of amino acid sequence; the bit score measures the size of the sequence database required where the current match could be found just by chance. The two nodes representing the *Cs*LAM and the *Bs*LAM were connected, even though at some separation, indicating a distant relatedness between the two enzymes, as was noted in figure 43 (more information at Appendix 6.38). The sequence network represents the distribution of LAM in several bacteria species (nodes). The size of the nodes represents how much data is available or that has been gathered. The interconnecting lines (edges) represents the relatedness between nodes through the similarity of their amino acid sequence. Further interrogation of the available nodes is still needed to figure out whether other strictly related *Bs*LAM enzyme sequences exist that could resist oxidative stress. Other sequences containing the redox-modulating residues or the C-terminus extension may be found that tolerate the presence of oxygen, and would expand the knowledge of the oxygen-tolerant radical SAM enzymes.

In the future our method could be employed to investigate on other radical SAM enzymes that need to be modified to be employed on industrial-scale level. The enhanced comprehension of the oxygen tolerance is pivotal to the development of a number of high-profile applications to use in biotechnology and this project provides a theoretical framework for the screening of protein structural modifications that can identify prospective viable mutants. Air-sensitive enzymes are not often able to undergo some of the unique and efficient chemical

transformations that are crucial to developing cost-effective and low-carbon routes to compounds of pharmacological interest, and there is still a gap that could be appropriately filled by radical enzymes.[241] As such, the validation of the proposed protocol is of considerable importance. With this project we aimed to engineer novel mutant enzymes that could be employed to improve the challenging area of fine chemical production with potential for future investment that could lead to large-scale process. We have produced a computational workflow that has revealed the likely source of oxygen tolerance resides with the impact of the protein structure, via electrostatics, on the reactivity of the active FeS cluster. This provides a powerful and rapid approach once validation of this protocol is achieved, with the potential to influence the development of radical SAM enzymes and other air-sensitive metalloenzymes for industrial development; and thus to grant access to complex new antibiotics, antimicrobial and anticancer compounds through low-carbon approaches with either social and environmental benefits.
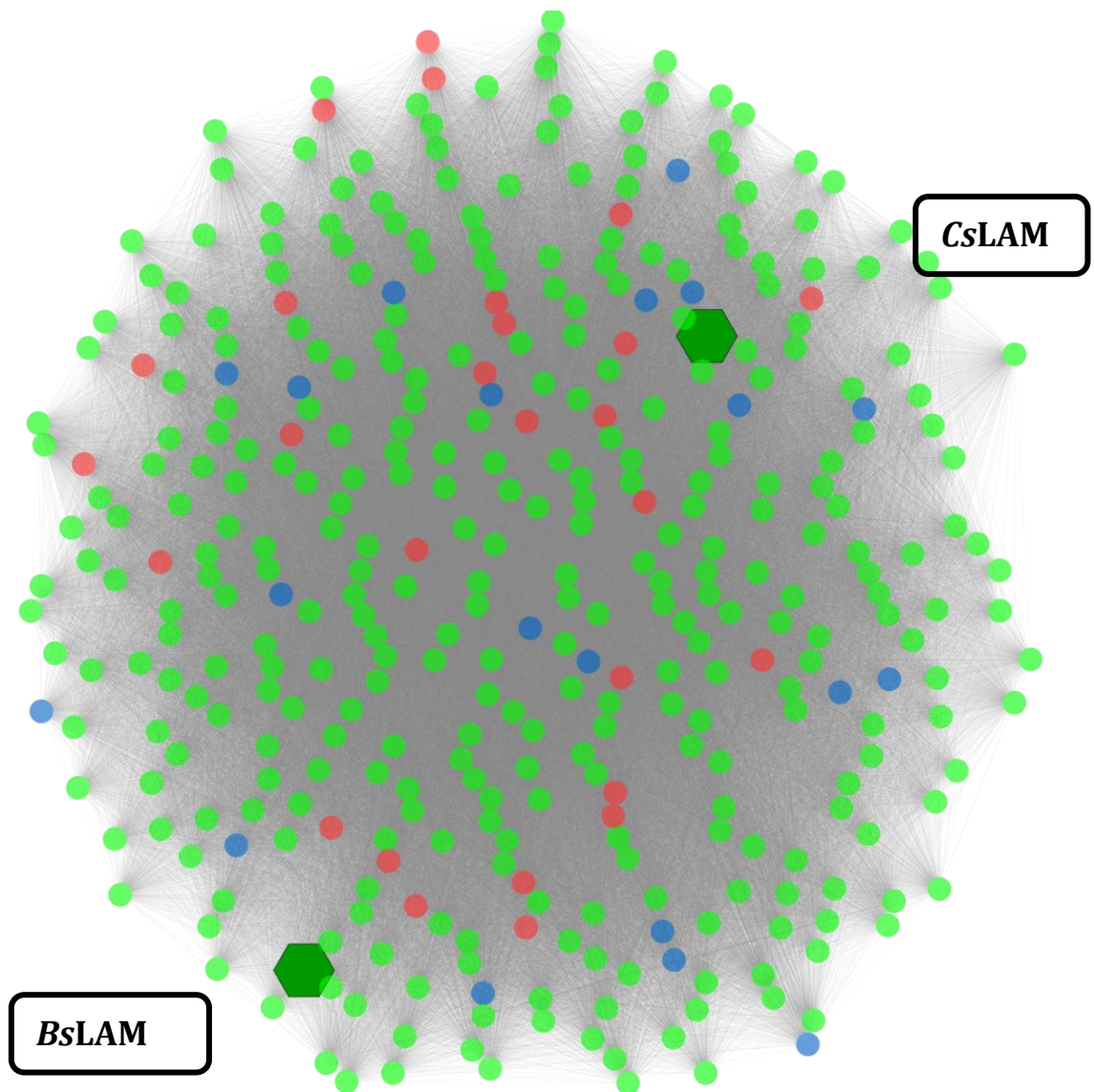
Figure 43. Sequence network of LAM rendered by type of life: red represents *Archaea*, green is for *Bacteria* and blue for *Environmental samples*. The edges, the interconnecting lines, are rendered by size. The two hexagons represent *Cs*LAM and *Bs*LAM nodes.

6 Appendix

6.1 Custom force-field parameters for SAM, [Fe₄-S₄] and cluster-binding cysteine residues.

Force-field modification file for SAM

MASS

SP 32.7798_ 2.900

BOND

CT-SP 345.28 1.78

ANGLE

C -CT-NT   80.000 111.200   same as C -CT-N3, penalty score= 0.1

CT-CT-SP   96.058   96.058

CT-SP-CT 289.219 100.546

H1-CT-SP   33.045 104.439

DIHE

CT-CT-SP-CT 2   0.008   0.000   -3.000

CT-CT-SP-CT 2   0.143   0.000   -2.000

CT-CT-SP-CT 2   1.114   0.000   1.000

H1-CT-SP-CT 2   0.475   0.000   -3.000

H1-CT-SP-CT 2   0.521   0.000   -2.000

H1-CT-SP-CT 2 -0.354   0.000   1.000

SP-CT-CT-CT 2   1.070   0.000   -2.000

SP-CT-CT-CT 2 -1.681   0.000   1.000

SP-CT-CT-H1 2   0.070   0.000   -3.000

SP-CT-CT-H1 2   1.105   0.000   -2.000

SP-CT-CT-H1 2   0.953   0.000   1.000

SP-CT-CT-HC 2   0.070   0.000   -3.000

SP-CT-CT-HC 2   1.105   0.000   -2.000

SP-CT-CT-HC 2 0.953   0.000   1.000

SP-CT-CT-OS 2 -0.021   0.000   -3.000

SP-CT-CT-OS 2 0.550   0.000   -2.000

SP-CT-CT-OS 2 3.566   0.000   1.000

IMPROPER

CT-O2-C -O2   10.5   180.0   2.0   Using general improper torsional angle X-O2- C-O2, penalty score=

3.0)

H5-N*-CK-NB   1.1   180.0   2.0   Using the default value

CA-CB-CB-NB   1.1   180.0   2.0   Using the default value

CA-H -N2-H   1.0   180.0   2.0   Using general improper torsional angle X- X-N2- H, penalty score=

6.0)

H5-NC-CQ-NC   1.1   180.0   2.0   Using the default value

CB-N*-CB-NC   1.1   180.0   2.0   Using the default value

NONBON

 SP   2.3890 0.0053

Force field modification file for [Fe$_4$S$_4$] cluster when no SAM is present in the active site

MASS

FC 55.845

FU 55.845

SC 32.065

BOND

FC-SC 55.2   2.31

FU-SC 55.2   2.31

FC-SH 60.9   2.31

FU-FH 60.9   2.31

ANGLE

SC-FC-SC 50.200   105.60

SC-FC-SH 50.000   113.20

FC-SC-FC 50.800   71.50

FC-SH-CT 50.600   105.60

SC-FU-SC 50.000   105.60 # higher values as in heme parameter set

FU-SC-FC 50.000   71.50 # for unique iron only

FU-SC-FU 50.800   71.50

DIHE

X -SC-FC-X   1   0.000 180.000   3.000

X -SC-FU-X   1   0.000 180.000   3.000

X -SC-SH-X   1   0.000 180.000   3.000

X -SH-FC-X   1   0.000 180.000   3.000

NONBON

FC   1.20   0.050 # iron from heme

FU   1.20   0.050

SC   2.00   0.250

_____

Force field modification file for [Fe$_4$S$_4$] cluster when SAM is present in the catalytic pocket

MASS

FC 55.845

FU 55.845

SC 32.065

BOND

FC-SC 55.2   2.31

FU-SC 55.2   2.31

FC-SH 60.9   2.31

FU-FH 60.9   2.31

ANGLE

SC-FC-SC 8.200   105.60

SC-FC-SH 11.000   113.20

FC-SC-FC 8.800   71.50

FC-SH-CT 15.600   105.60

SC-FU-SC 8.200   105.60

FU-SC-FC 8.800   71.50

DIHE

X -SC-FC-X   1   0.000 180.000   3.000

X -SC-FU-X   1   0.000 180.000   3.000

X -SC-SH-X   1   0.000 180.000   3.000

X -SH-FC-X   1   0.000 180.000   3.000

NONBON

FC   1.20   0.050 # iron from heme

FU   1.20   0.050

SC   2.00   0.250

_____

Force field modification file for [Fe$_4$S$_4$] cluster-coordinating cysteine residues (CYF)

!!index array str

 "CYF"

!entry.CYF.unit.atoms table str name str type int typex int resx int flags int seq int elmnt dbl chg

 "N" "N" 0 1 131073 1 7 -0.415700

 "H" "H" 0 1 131073 2 1 0.271900

 "CA" "CT" 0 1 131073 3 6 0.021300

 "HA" "H1" 0 1 131073 4 1 0.112400

 "CB" "CT" 0 1 131073 5 6 -0.123100

 "HB2" "H1" 0 1 131073 6 1 0.111200

 "HB3" "H1" 0 1 131073 7 1 0.111200

 "SG" "SH" 0 1 131073 8 16 -0.677700

 "C" "C" 0 1 131073 10 6 0.597300

 "O" "O" 0 1 131073 11 8 -0.567900

!entry.CYF.unit.atomspertinfo table str pname str ptype int ptypex int pelmnt dbl pchg

"N" "N" 0 -1 0.0

"H" "H" 0 -1 0.0

"CA" "CT" 0 -1 0.0

"HA" "H1" 0 -1 0.0

"CB" "CT" 0 -1 0.0

"HB2" "H1" 0 -1 0.0

"HB3" "H1" 0 -1 0.0

"SG" "SH" 0 -1 0.0

"C" "C" 0 -1 0.0

"O" "O" 0 -1 0.0

!entry.CYF.unit.boundbox array dbl

-1.000000

0.0

0.0

0.0

0.0

!entry.CYF.unit.childsequence single int

2

!entry.CYF.unit.connect array int

1

9

!entry.CYF.unit.connectivity table int atom1x int atom2x int flags

1 2 1

1 3 1

3 4 1

3 5 1

3 9 1

5 6 1

5 7 1

5 8 1

9 10 1

!entry.CYF.unit.hierarchy table str abovetype int abovex str belowtype int belowx

"U" 0 "R" 1

"R" 1 "A" 1

 "R" 1 "A" 2

 "R" 1 "A" 3

 "R" 1 "A" 4

 "R" 1 "A" 5

 "R" 1 "A" 6

 "R" 1 "A" 7

 "R" 1 "A" 8

 "R" 1 "A" 9

 "R" 1 "A" 10

!entry.CYF.unit.name single str

 "CYS"

!entry.CYF.unit.positions table dbl x dbl y dbl z

 3.325770 1.547909 -1.607204E-06

 3.909407 0.723611 -2.739882E-06

 3.970048 2.845795 -1.311163E-07

 3.671663 3.400129 -0.889820

 3.576965 3.653838 1.232143

 2.496995 3.801075 1.241379

 3.877484 3.115795 2.131197

 4.309573 5.303523 1.366036

 5.485541 2.705207 -4.398755E-06

 6.008824 1.593175 -8.449768E-06

!entry.CYF.unit.residueconnect table int c1x int c2x int c3x int c4x int c5x int c6x

 1 9 0 0 0 0

!entry.CYF.unit.residues table str name int seq int childseq int startatomx str restype int imagingx

 "CYS" 1 12 1 "p" 0

!entry.CYF.unit.residuesPdbSequenceNumber array int

 0

!entry.CYF.unit.solventcap array dbl

 -1.000000

 0.0

 0.0

0.0

0.0

!entry.CYF.unit.velocities table dbl x dbl y dbl z

0.0 0.0 0.0

0.0 0.0 0.0

0.0 0.0 0.0

0.0 0.0 0.0

0.0 0.0 0.0

0.0 0.0 0.0

0.0 0.0 0.0

0.0 0.0 0.0

0.0 0.0 0.0

0.0 0.0 0.0

6.2   Secondary structure of monomeric *Cs*LAM.

Alpha helices and beta-sheets distribution represented in New Ribbon. In the centre of the TIM barrel, the SAM cofactor and the [Fe$_4$-S$_4$] cluster are represented in Licorice.
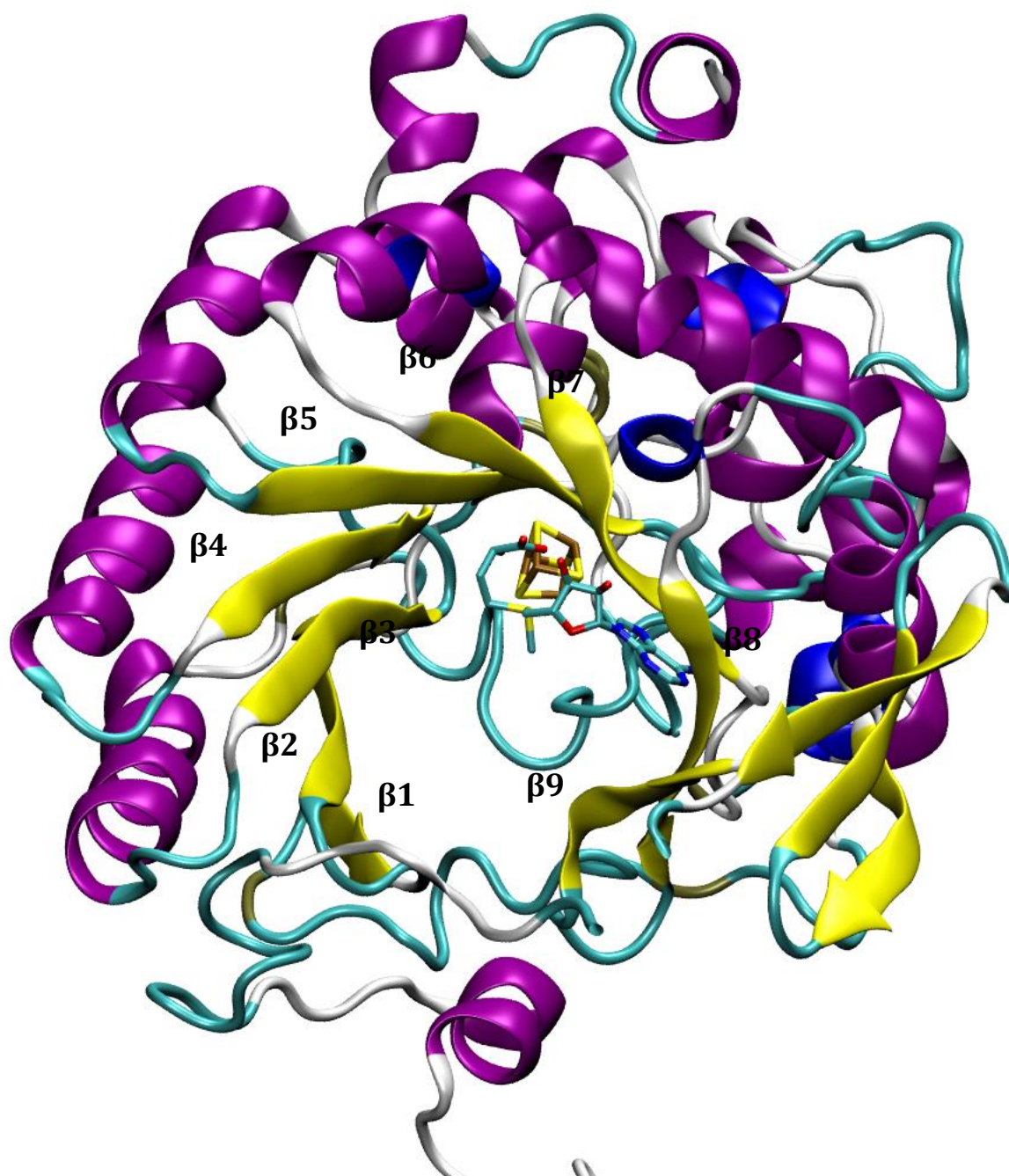
6.3 RMSD and RMSF of ThiC MD simulations with N-terminus removed.

RMSD and RMSF of ThiC crystal structure (a, c) and ThiC homology model (b, d) MD simulations with removed N-terminus (residues 1 to 110) from their monomeric chains.

a)

RMSD of backbone atoms in THIC crystal structure

b)

RMSD of backbone atoms in THIC homology model

c)

RMSF of alpha carbon atoms in ThiC crystal structure

d)

RMSF of alpha carbon atoms in ThiC homology model
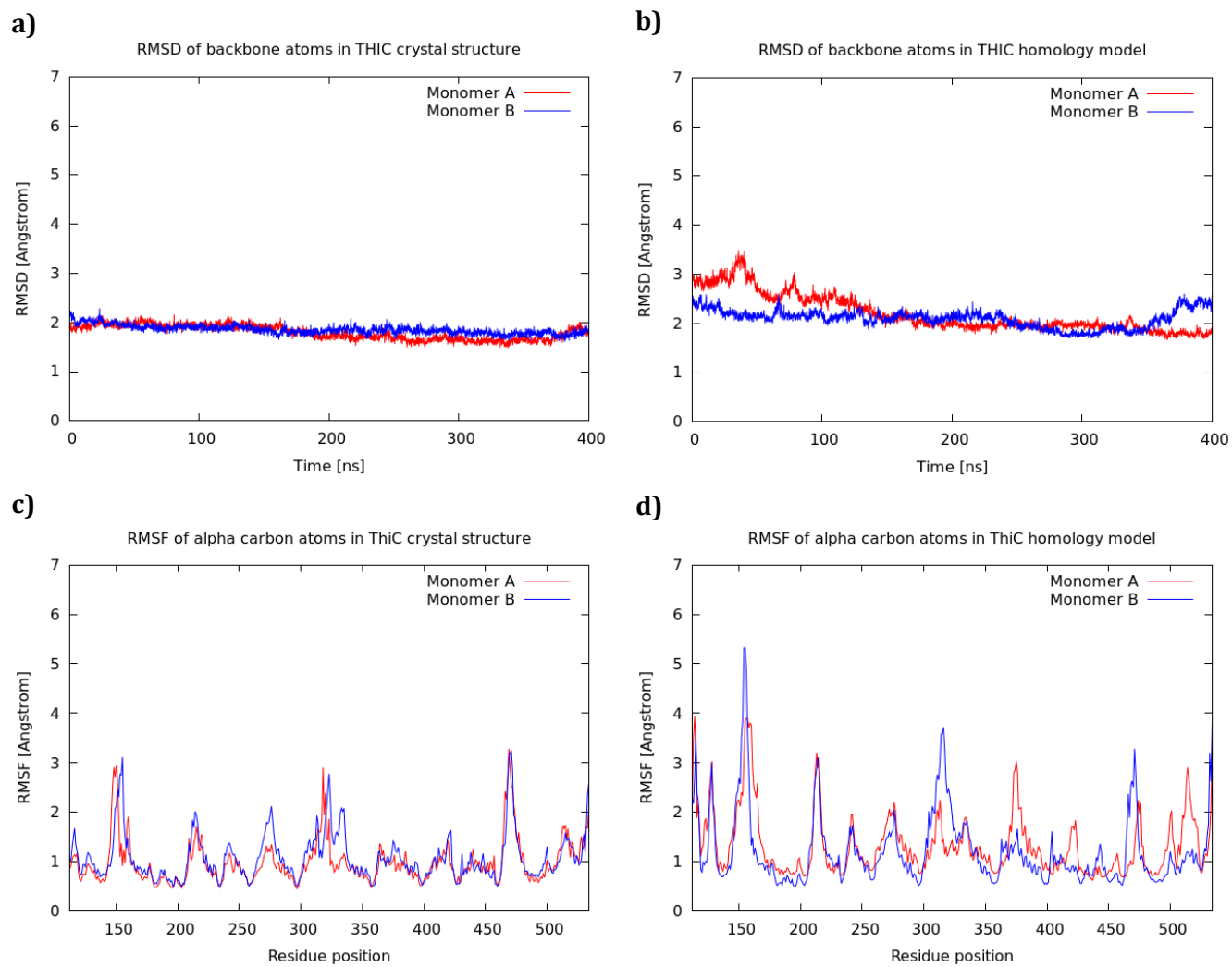
6.4   PDB representation of eigenvectors sampled from combined THIC MD simulations trajectories.
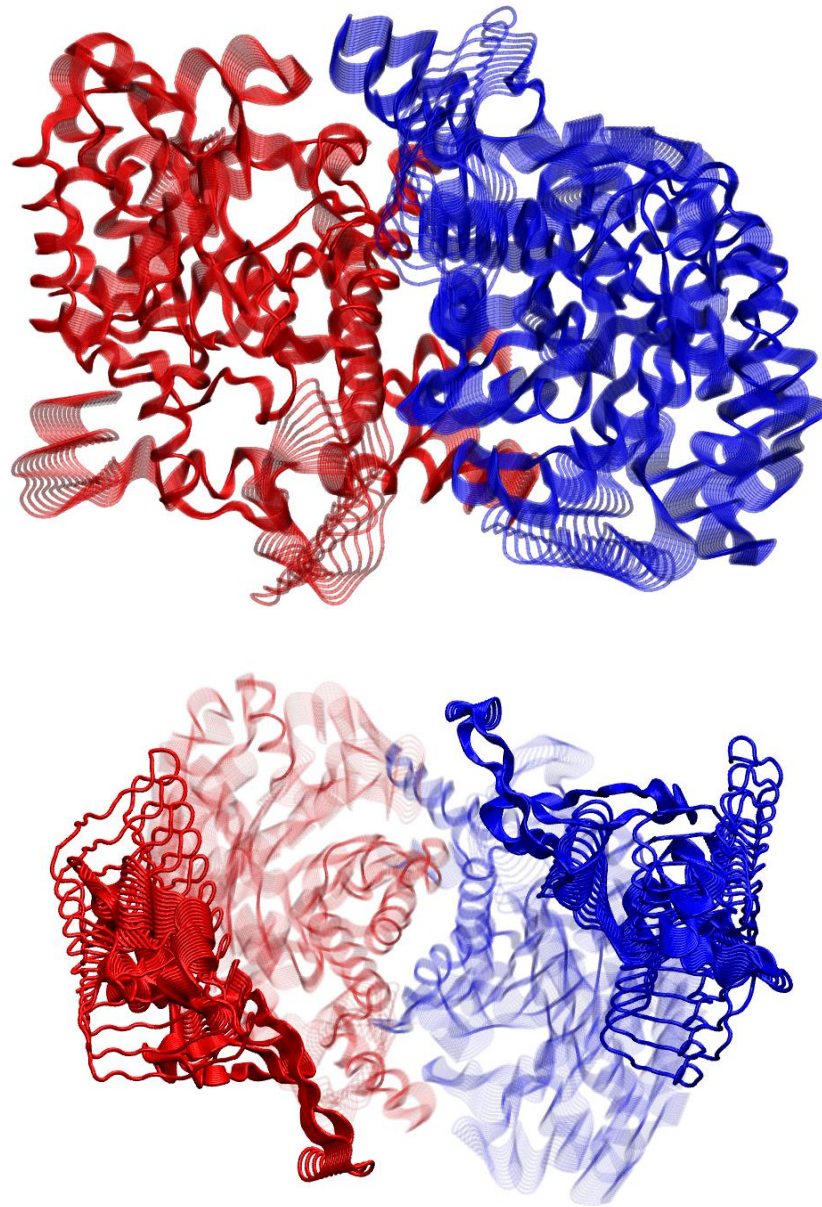
6.5   Structural alignment of the THIC crystal structure and its related homology model.

The image shows the structural alignment of THIC structures after removal of the N-terminus from both enzymes. The colours of the structures are RMSD-based with blue indicating areas of good alignment while red indicates local divergent RMSD.
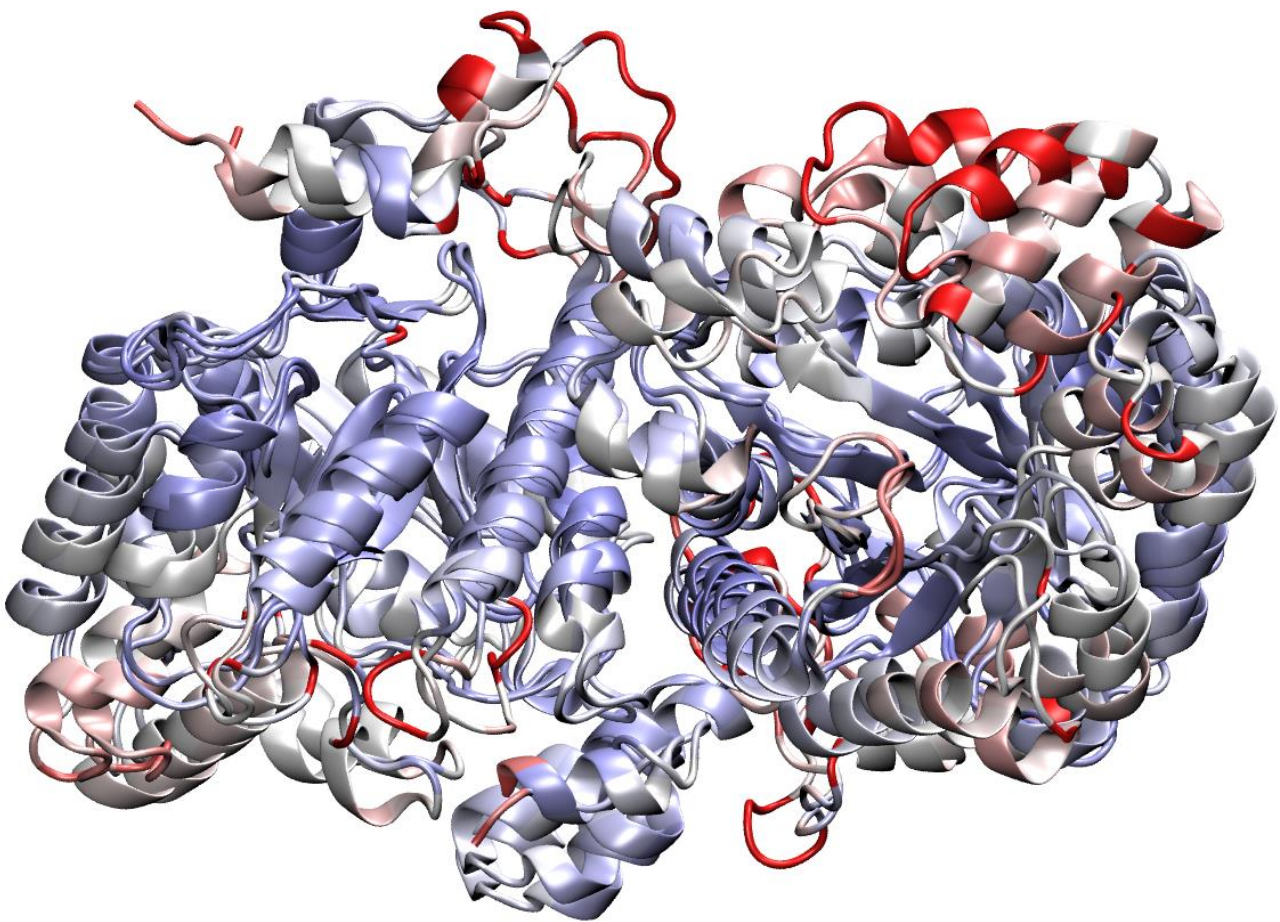
6.6   PROCHECK results of the *Bs*LAM model C-terminus tail.

The last 72 residues were submitted to a separate regular check.

a)

b)
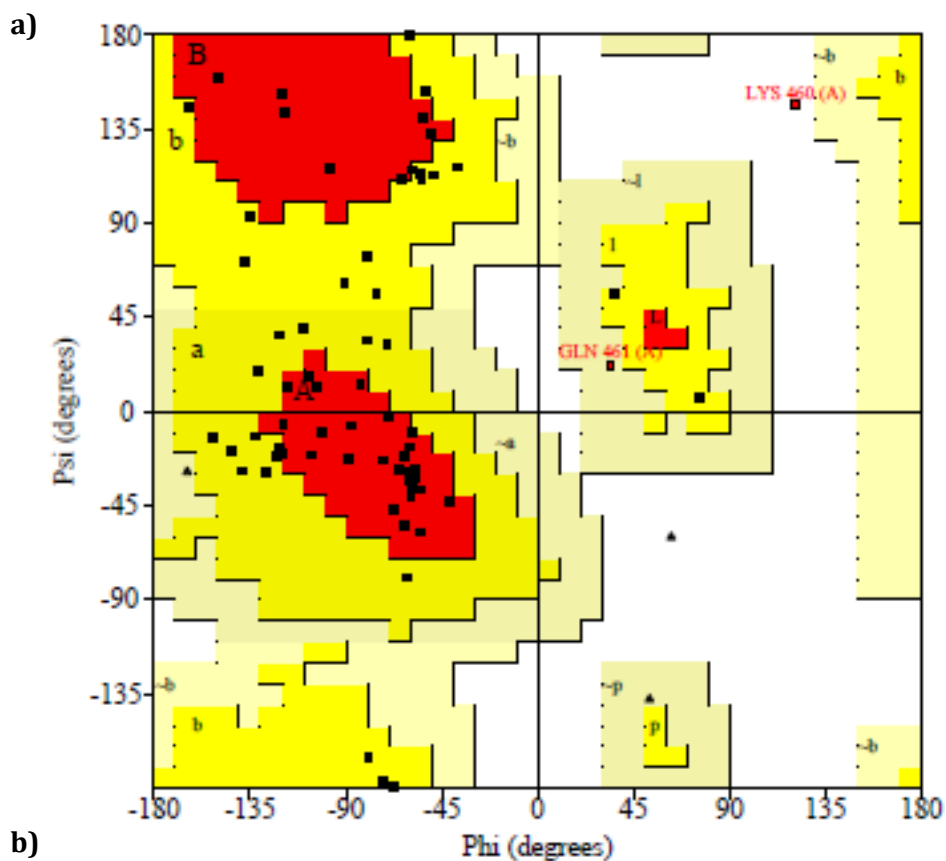
```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
|                                                            |
| /var/www/PROCHECK/Jobs/3129020/3129020.pdb   1.5      72 residues |
|                                                            |
*| Ramachandran plot:   50.0% core   46.9% allow   1.6% gener   1.6% disall |
|                                                            |
*| All Ramachandrans:   13 labelled residues (out of  70)     |
*| Chi1-chi2 plots:      3 labelled residues (out of  51)     |
+| Side-chain params:    2 better    0 inside    3 worse      |
|                                                            |
*| Residue properties: Max.deviation:    9.9          Bad contacts:    1 |
*|                     Bond len/angle:   8.1    Morris et al class:  4  2  2 |
|                                                            |
+| G-factors           Dihedrals: -1.02  Covalent: -0.53   Overall: -0.79 |
|                                                            |
*| Planar groups:   87.5% within limits  12.5% highlighted     1 off graph |
|                                                            |
+------------------------------------------------------------+
    + May be worth investigating further.   * Worth investigating further.
```

193

6.7   Xleap script for the creation of topology and internal coordinates of the enzyme structure.

$xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff12SB

Load ion force field modification file:

> loadamberparams frcmod.ionslrcm_cm_spce

Load non-bonded ion force field modification file:

> loadamberparams frcmod.ionsjc_spce

> loadamberparams /media/RAIDstorage/home/damiano/parameters/SAM.frcmod

> loadamberparams /media/RAIDstorage/home/damiano/parameters/SF4.frcmod

> loadamberparams /media/RAIDstorage/home/damiano/parameters/SF4-stronger.frcmod

> loadoff /media/RAIDstorage/home/damiano/parameters/SF4.lib

> loadoff /media/RAIDstorage/home/damiano/parameters/SAM-0-multi-HF.lib

> loadoff /media/RAIDstorage/home/damiano/parameters/CYF.lib

> SAM = loadpdb 2A5H_SAM.pdb

```
> bond SAM.125.SG SAM.1641.FE1
> bond SAM.129.SG SAM.1641.FE4
> bond SAM.132.SG SAM.1641.FE2
> bond SAM.951.SG SAM.1647.FE2
> bond SAM.948.SG SAM.1647.FE4
> bond SAM.944.SG SAM.1647.FE1
> bond SAM.1360.SG SAM.1650.FE2
> bond SAM.1357.SG SAM.1650.FE4
> bond SAM.1353.SG SAM.1650.FE1
> bond SAM.541.SG SAM.1644.FE2
> bond SAM.538.SG SAM.1644.FE4
> bond SAM.534.SG SAM.1644.FE1
```

check then edit the unit

> saveoff SAM SAM_linear.lib

> savepdb SAM SAM_linear.pdb

> solvateoct SAM SPCBOX 10.0

Check unit for problem, then neutralize:

> addions SAM Cl- 0

After all check unit for problem and save the topology and coordinate files:

> saveamberparm SAM SAM.prmtop SAM.inpcrd

> savepdb SAM SAM_wat.pdb

>quit

6.8   AMBER scripts used for the MD simulations protocol

First step minimisation

```
minimization 1 – Protein restrained
 &cntrl
  imin=1,
  maxcyc=10000,
  ncyc=3000,
  cut=10.0,
  ntb=1,
  ntr=1
 /
Hold protein, Zn, S4F4, SAM fixed
30.0
RES 1 1652
END
END
```

Second step minimisation

```
minimization 2 – hold cofactors and Zn ions restrained
 &cntrl
  imin   = 1,
  maxcyc = 8000,
  ncyc   = 2500,
  ntb    = 1,
  ntr    = 1,
  cut    = 10.0
 /
Hold Fe4S4, Zn, SAM fixed
30.0
RES 1639 1650
END
END
```

Heating step from 0 to 300K

```
Stage 3 heating 0 to 300K with distance restraints on Fe4S4 - CYS
 &cntrl
  imin   = 0,
  irest  = 0,
  ntx    = 1,
  ntb    = 1,
  cut    = 10.0,
  ntr    = 0,
  ntc    = 2,
  ntf    = 2,
  tempi  = 0.0,
```

```
  temp0  = 300.0,
  ntt    = 3,
  ig     = -1,
  gamma_ln = 1.0,
  nstlim = 500000, dt = 0.002,
  ntpr = 1000, ntwx = 1000, ntwr = 1000
  nmropt = 1
 /
&wt
   type = 'DUMPFREQ', istep1 = 50
 /
&wt type = 'END' /
DISANG = SAM_RST1.dist
DUMPAVE = SAM_dist1.dat
```

AMBER script for 1ns of equilibration step

```
1ns equilibration step with decreasing restraints
 &cntrl
  imin   = 0,
  irest  = 1,
  ntx    = 7,
  ntb    = 2,
  ntp    = 1,
  cut    = 10.0,
  ntc    = 2,
  ntf    = 2,
  pres0 = 1.0,
  tempi  = 300.0,
  temp0  = 300.0,
  taup   = 3.0,
  ig     = -1,
  gamma_ln = 1.0,
  nstlim = 10000000, dt = 0.0002,
  ntpr = 10000, ntwx = 100000, ntwr = 100000
  nmropt = 1
  iwrap = 1,
  ioutfm = 1
  /
&wt
  type = 'DUMPFREQ', istep1 = 50
 /
&wt type = 'END' /
DISANG = SAM_RST1.dist
DUMPAVE = SAM_RST1.dat
```

AMBER script for 10ns of trajectory production step

Production step:
10ns MD with distance restraints on Fe4S4 - CYS
```
 &cntrl
  imin = 0, irest = 1, ntx = 7,
  ntb = 2, pres0 = 1.0, ntp = 1,
  taup = 2.0,
  cut = 10.0, ntr = 0,
  ntc = 2,
  ntf = 2,
  tempi = 300.0, temp0 = 300.0,
  ntt = 3, ig = -1, gamma_ln = 1.0,
  nstlim = 5000000, dt = 0.002,
  ioutfm = 1, iwrap = 1,
  ntpr = 1000, ntwx = 1000, ntwr = 1000
  nmropt = 1
 /
 &wt
  type = 'DUMPFREQ', istep1 = 50
 /
 &wt type = 'END' /
 DISANG = SAM_RST1.dist
 DUMPAVE = SAM_RST1.dat
```

6.9 CPPTRAJ script created for the MD simulations analysis

6.9.1 Reduce and merge together all the produced trajectory files.

In just one file it leads to a trajectory files stripped by solvent and counterions.

```
parm fixed.SAM.prmtop
trajin SAM10ns_md.mdcrd 0 last 50
trajin SAM20ns_md.mdcrd 0 last 50
trajin SAM30ns_md.mdcrd 0 last 50
trajin SAM40ns_md.mdcrd 0 last 50
trajin SAM50ns_md.mdcrd 0 last 50
trajin SAM60ns_md.mdcrd 0 last 50
trajin SAM70ns_md.mdcrd 0 last 50
trajin SAM80ns_md.mdcrd 0 last 50
trajin SAM90ns_md.mdcrd 0 last 50
trajin SAM100ns_md.mdcrd 0 last 50
trajin SAM110ns_md.mdcrd 0 last 50
trajin SAM120ns_md.mdcrd 0 last 50
trajin SAM130ns_md.mdcrd 0 last 50
trajin SAM140ns_md.mdcrd 0 last 50
trajin SAM150ns_md.mdcrd 0 last 50
trajin SAM160ns_md.mdcrd 0 last 50
trajin SAM170ns_md.mdcrd 0 last 50
trajin SAM180ns_md.mdcrd 0 last 50
trajin SAM190ns_md.mdcrd 0 last 50
trajin SAM200ns_md.mdcrd 0 last 50
trajin SAM210ns_md.mdcrd 0 last 50
trajin SAM220ns_md.mdcrd 0 last 50
trajin SAM230ns_md.mdcrd 0 last 50
trajin SAM240ns_md.mdcrd 0 last 50
trajin SAM250ns_md.mdcrd 0 last 50
trajin SAM260ns_md.mdcrd 0 last 50
trajin SAM270ns_md.mdcrd 0 last 50
trajin SAM280ns_md.mdcrd 0 last 50
```

trajin SAM290ns_md.mdcrd 0 last 50

trajin SAM300ns_md.mdcrd 0 last 50

trajin SAM310ns_md.mdcrd 0 last 50

trajin SAM320ns_md.mdcrd 0 last 50

trajin SAM330ns_md.mdcrd 0 last 50

trajin SAM340ns_md.mdcrd 0 last 50

trajin SAM350ns_md.mdcrd 0 last 50

trajin SAM360ns_md.mdcrd 0 last 50

trajin SAM370ns_md.mdcrd 0 last 50

trajin SAM380ns_md.mdcrd 0 last 50

trajin SAM390ns_md.mdcrd 0 last 50

trajin SAM400ns_md.mdcrd 0 last 50

unwrap :1-1645

center :1-1645 mass origin

autoimage

center :1-1645 mass origin

image :1-1645,SF4,SAM origin center familiar

autoimage

strip :WAT,Cl-

trajout SAM400nsReimagedStrppdReduced.nc netcdf

6.9.2 Calculate RMSD and RMSF of the retrieved trajectories.

The average PDB structure was created by *cpptraj* in the script and it was used as the reference structure.

```
parm SAM_noWatIons.prmtop
trajin SAM400nsReimagedStrppdReduced.nc
average SAM_avg_str.pdb pdb start 2000 stop 4000 :1-1638
[SAM_avg_str.pdb]
run
reference SAM_avg_str.pdb
rmsd fit :1-1638@CA,C,N
rmsd ref SAM_avg_str.pdb mass :1-409@CA,C,N out
rmsdSAM400_AbackbnMonBI.dat time 2.0 R1b
rmsd ref SAM_avg_str.pdb mass :410-819@CA,C,N out
rmsdSAM400_BbackbnMonBI.dat time 2.0 R2b
rmsd ref SAM_avg_str.pdb mass :820-1228@CA,C,N out
rmsdSAM400_CbackbnMonBI.dat time 2.0 R3b
rmsd ref SAM_avg_str.pdb mass :1229-1638@CA,C,N out
rmsdSAM400_DbackbnMonBI.dat time 2.0 R4b
atomicfluct :1-409@CA byres out rmsfSAM400_AcaB.dat
atomicfluct :410-819@CA byres out rmsfSAM400_BcaB.dat
atomicfluct :820-1228@CA byres out rmsfSAM400_CcaB.dat
atomicfluct :1229-1638@CA byres out rmsfSAM400_DcaB.dat
```

6.10 Distances between Nε atoms of His128 and His230.

The distance between Nε atoms of His128 and His230 is reported in orange in the plots whilst the distance between Nε atoms of His230 and Cζ of Arg200 is depicted in blue for each monomer composing the unbound-SAM structure; in the next page the same data are reported for the SAM-bound MD simulation. Below, a snapshot taken from *Cs*LAM MD simulation outlines in detail the atoms of histidine and arginine residues measured in the active site with SAM and [Fe$_4$S$_4$] cluster reported in transparent.

Distance between histidine residues and argininine in Monomer A

Distance between histidine residues and arginine in Monomer B

Distance between histidine residues and arginine in Monomer C

Distance between histidine residues and arginine in Monomer D
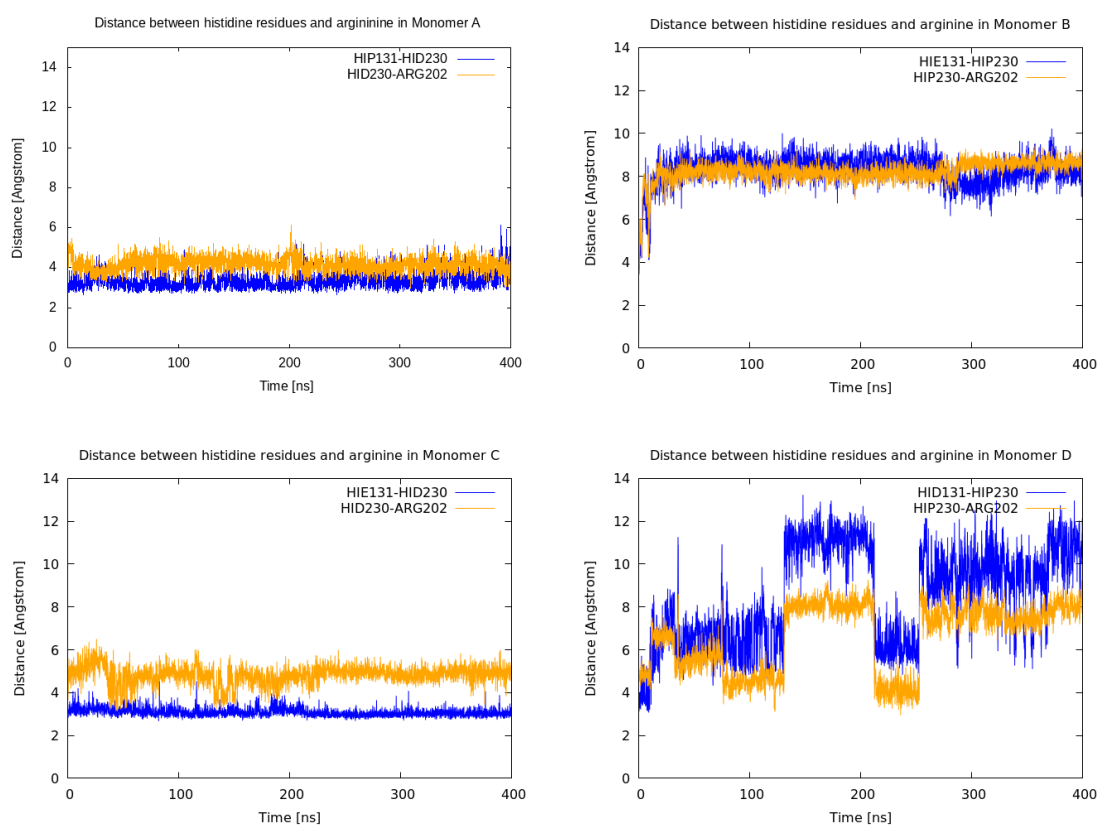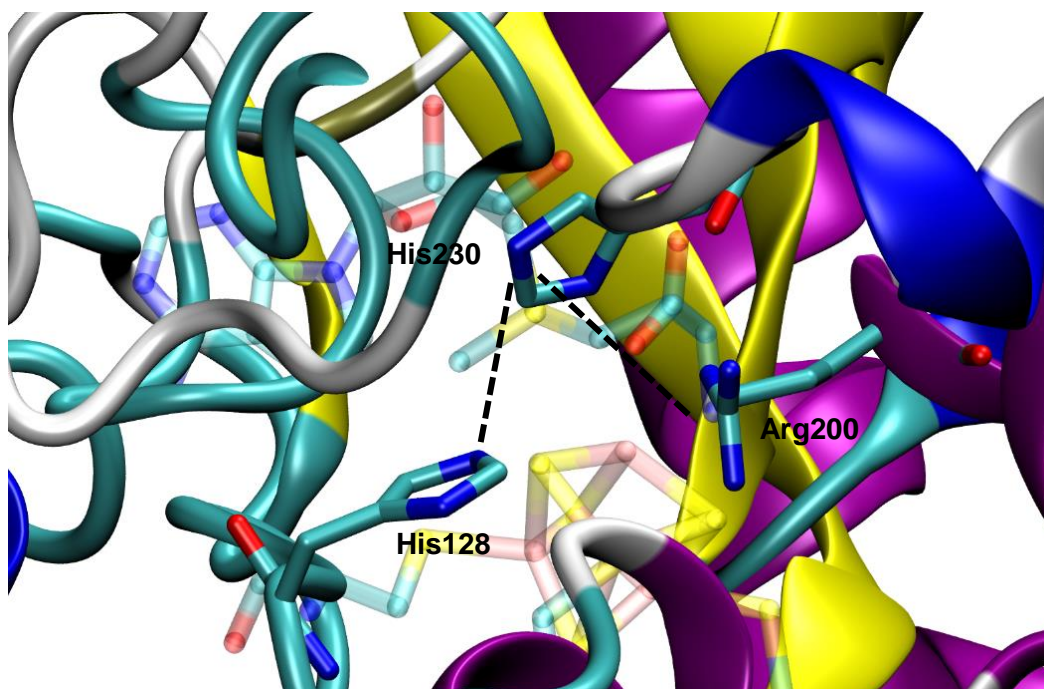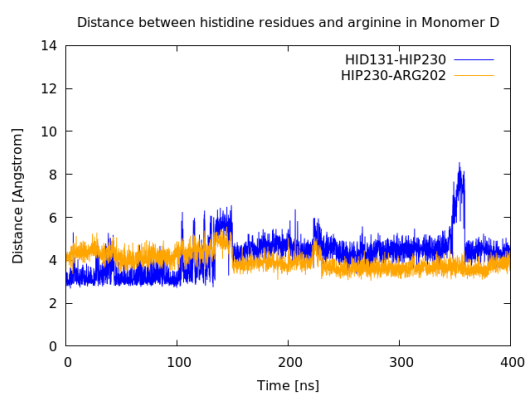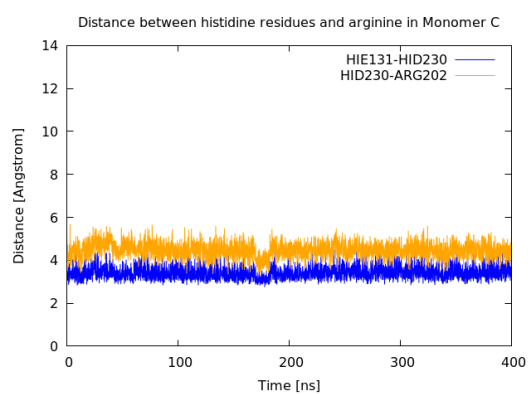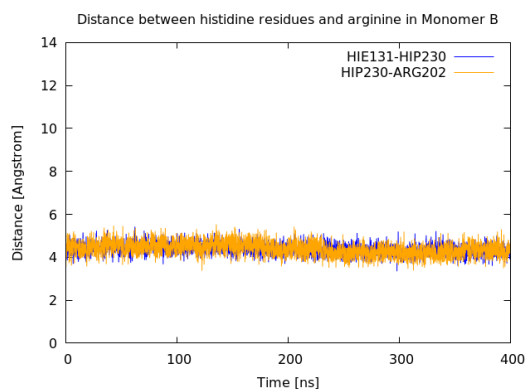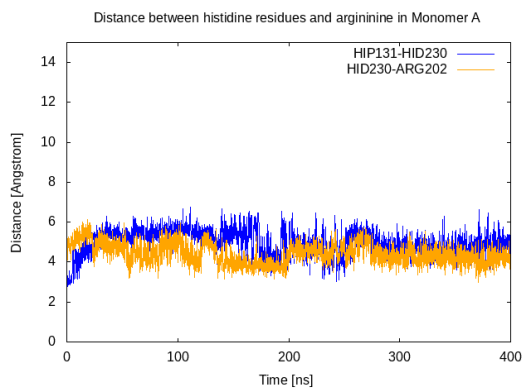
6.11   Principal component analysis of *Cs*LAM MD simulations over time.

PCA of both *Cs*LAM MD simulations sampled at 100 ns (a), 200 ns (b) and 300 ns (c). Colours of each simulation are equal to PCA plot explained in text (section 3.4.2).
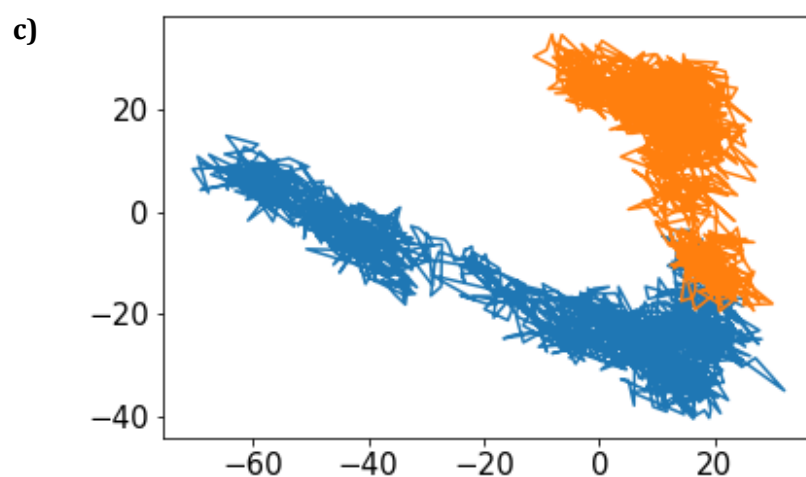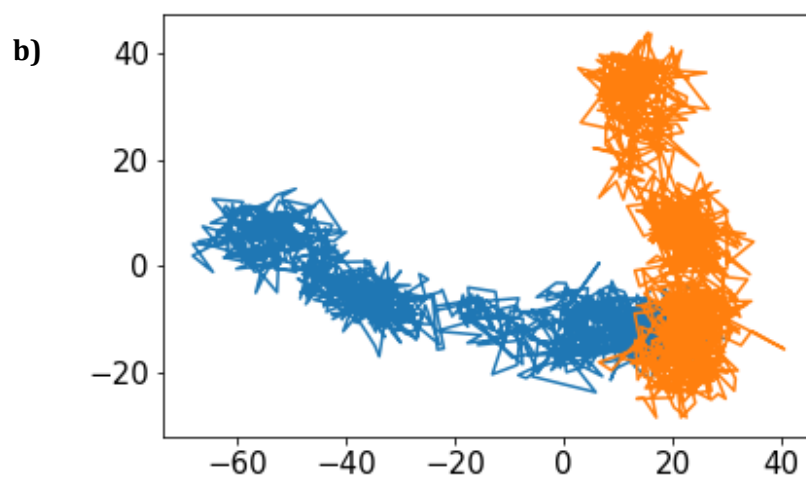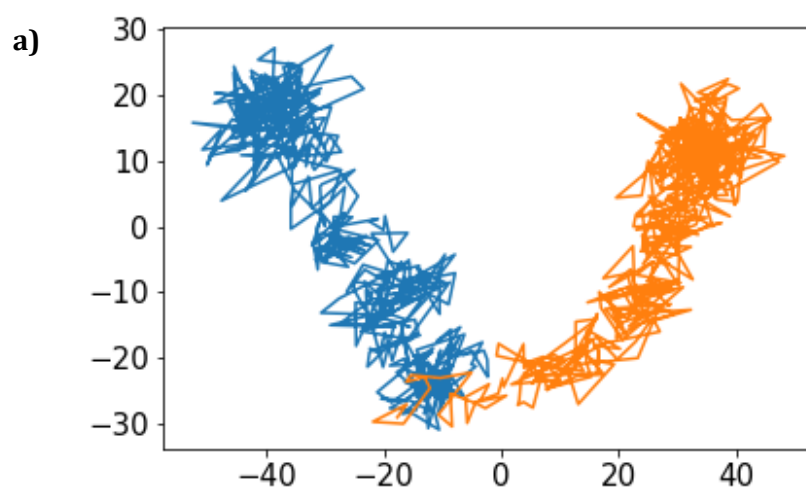
6.12   Porcupine plots of first and second principal components in *Cs*LAM.

PC1 (above) and PC2 (below) of *Cs*LAM with no SAM bound in the active site represented by porcupine plots.
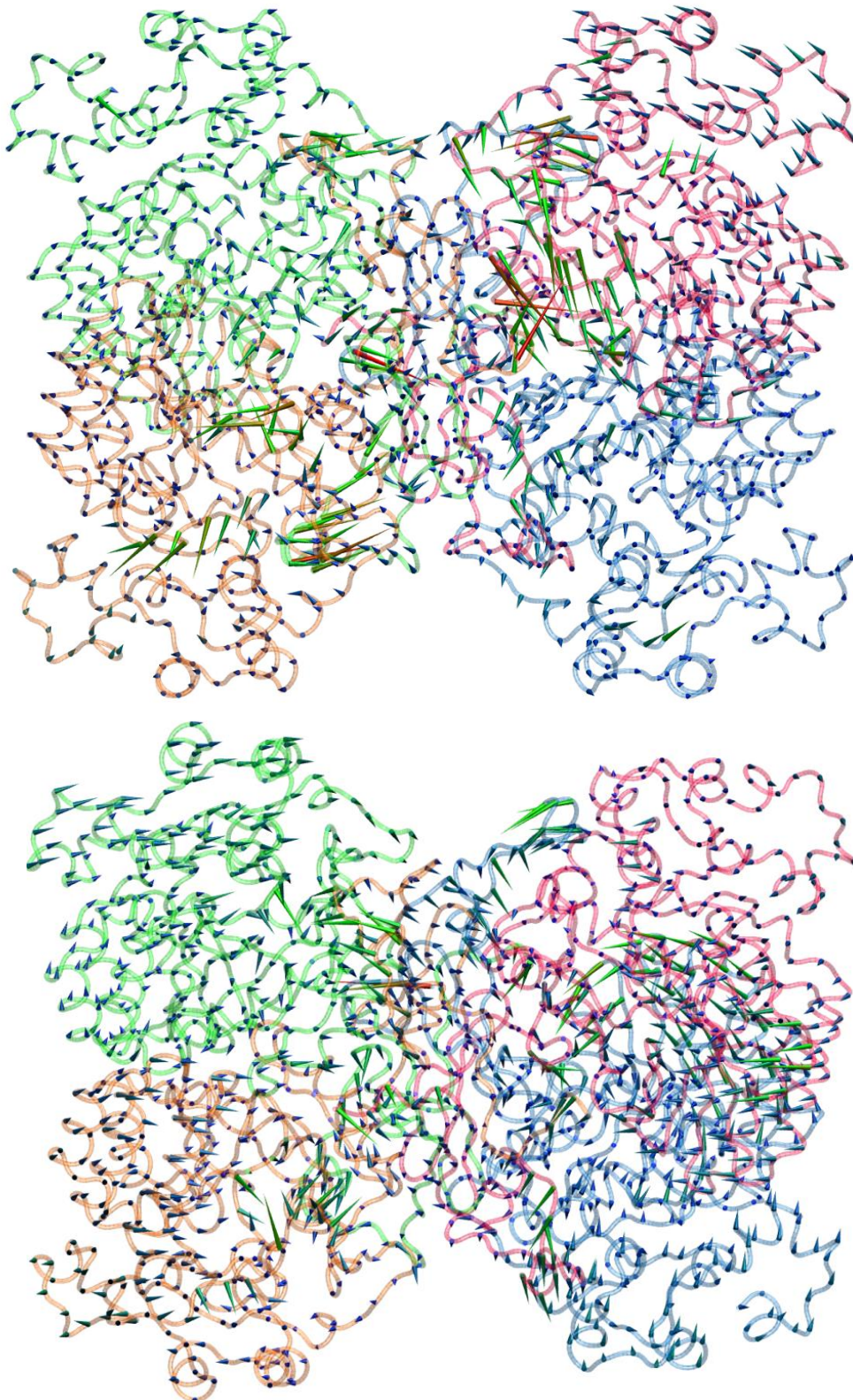
PC1 (above) and PC2 (below) of *Cs*LAM with SAM bound. Each monomer was coloured following the colour representation used in RMSD and RMSF plots.
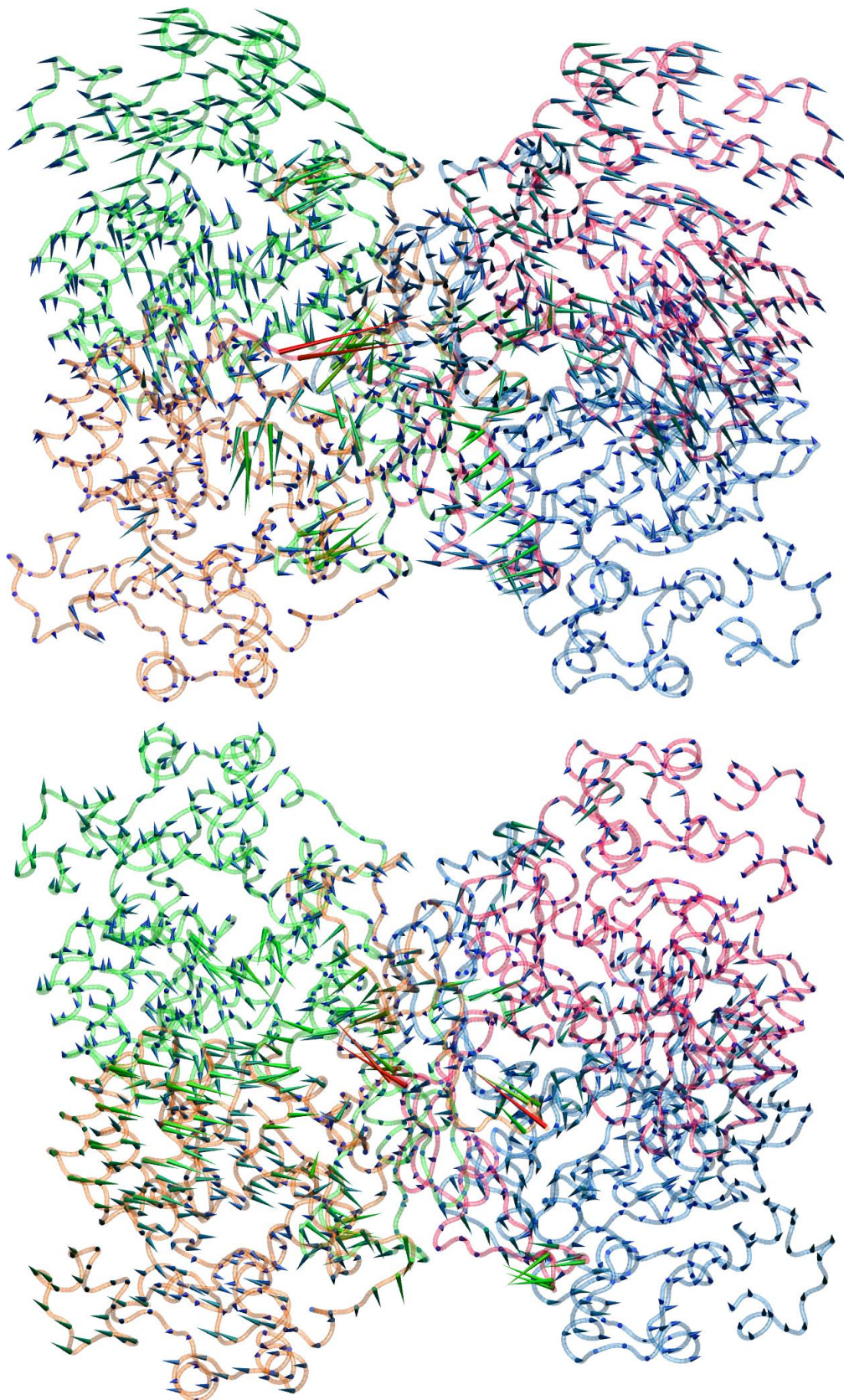
6.13   Other RMSD and RMSF plots of *Bs*LAM MD simulations.

The RMSF of the 0 – 100 ns interval a) and 200 – 400 ns time interval b) in the unbound-SAM

*Bs*LAM MD simulation. RMSD of the unbound-SAM system c) and the bound-SAM enzyme d)

with last 100 amino acid residues removed.
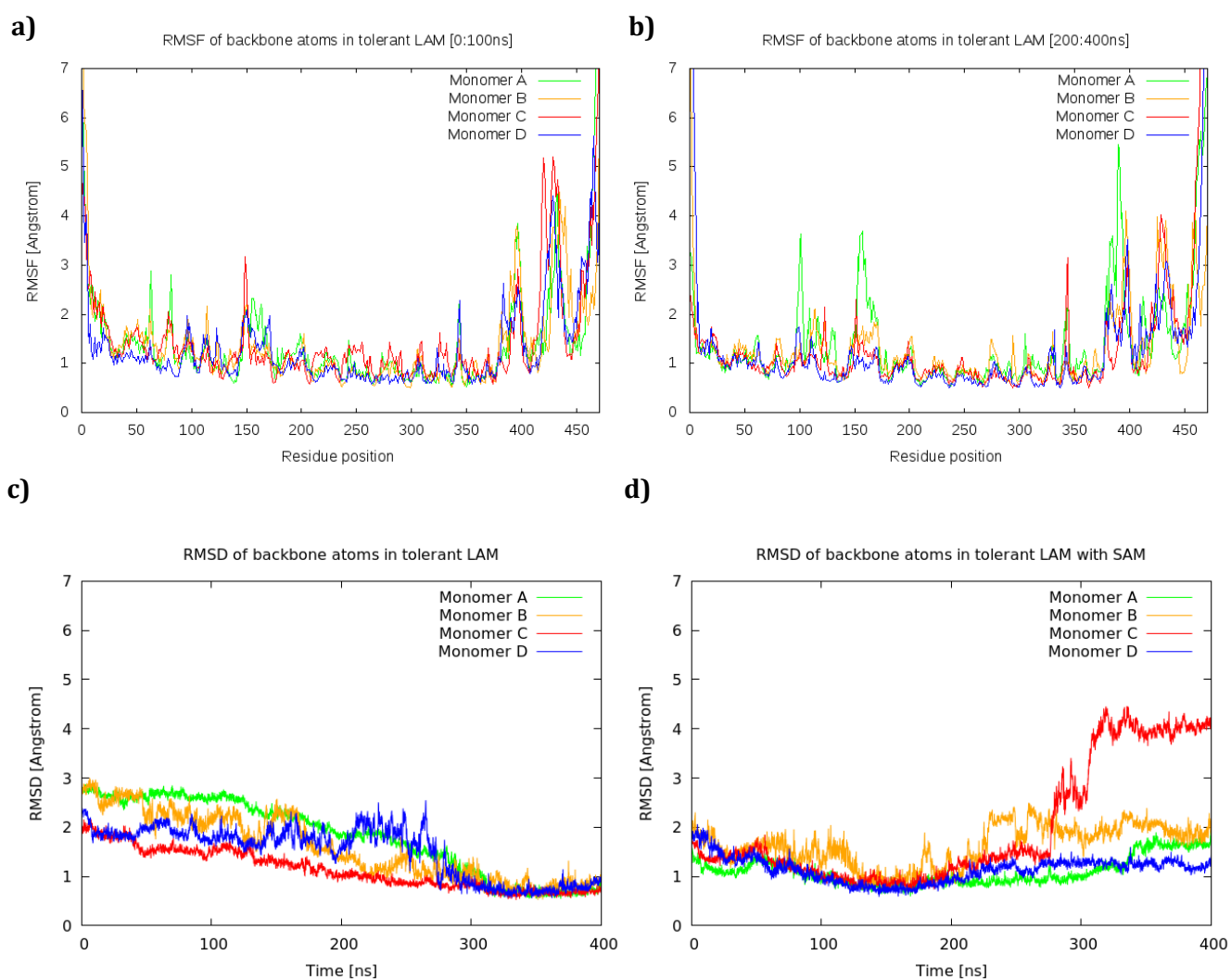
RMSF of both *Bs*LAM enzymes with C-termini (last 100 amino acid) removed, without SAM and with SAM in each pocket (plot e) and f) respectively).

Plots g) and h) respectively reported the RMSD and RMSF of *Bs*LAM with bound-SAM with bond N- and C-termini removed.

e)
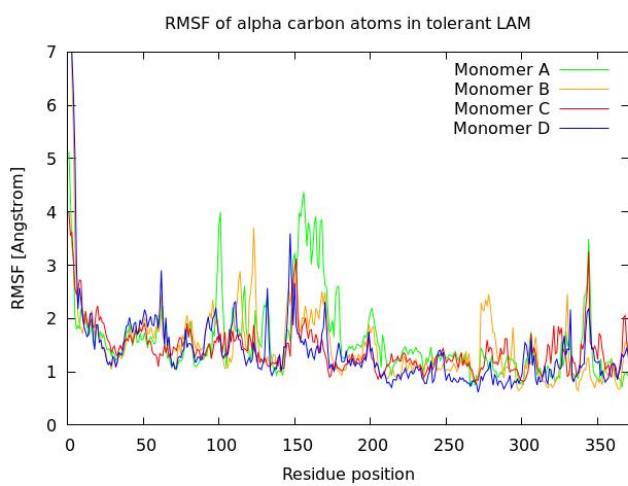
f)
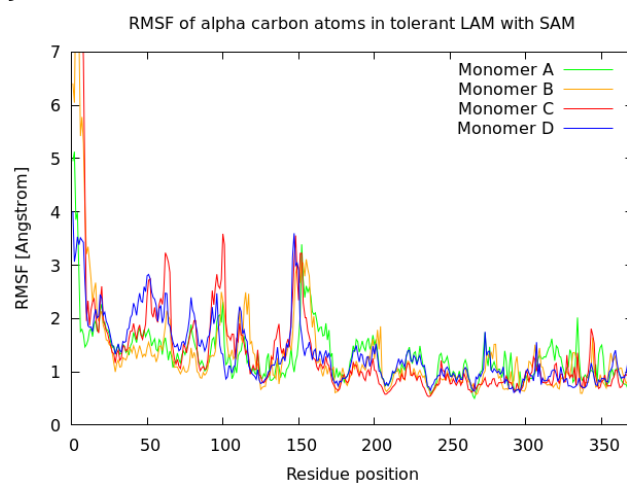
g)
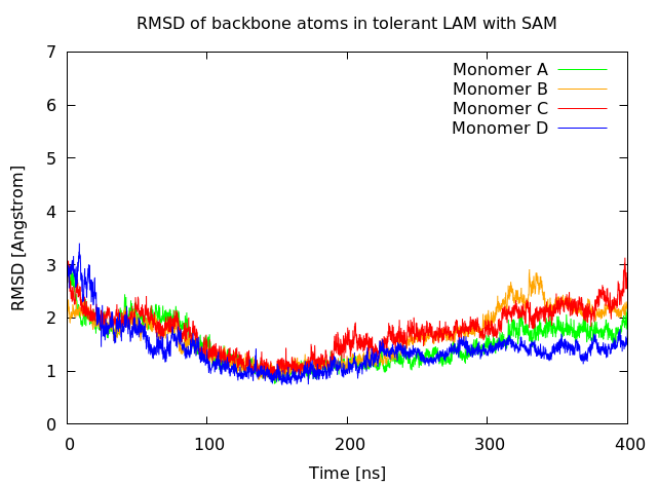
h)
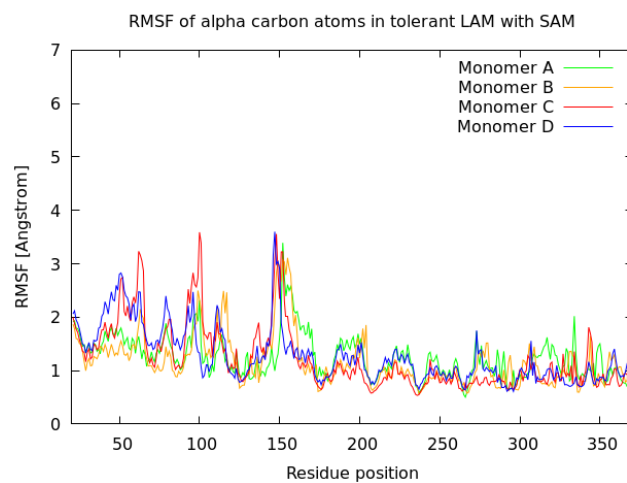
6.14   Principal component analysis of *Bs*LAM MD simulations over time.

PCA of both *Bs*LAM MD simulations sampled at 100 ns (a), 200 ns (b) and 300 ns (c). Colours of each simulation are equal to PCA plot explained in text (section 3.4.4).
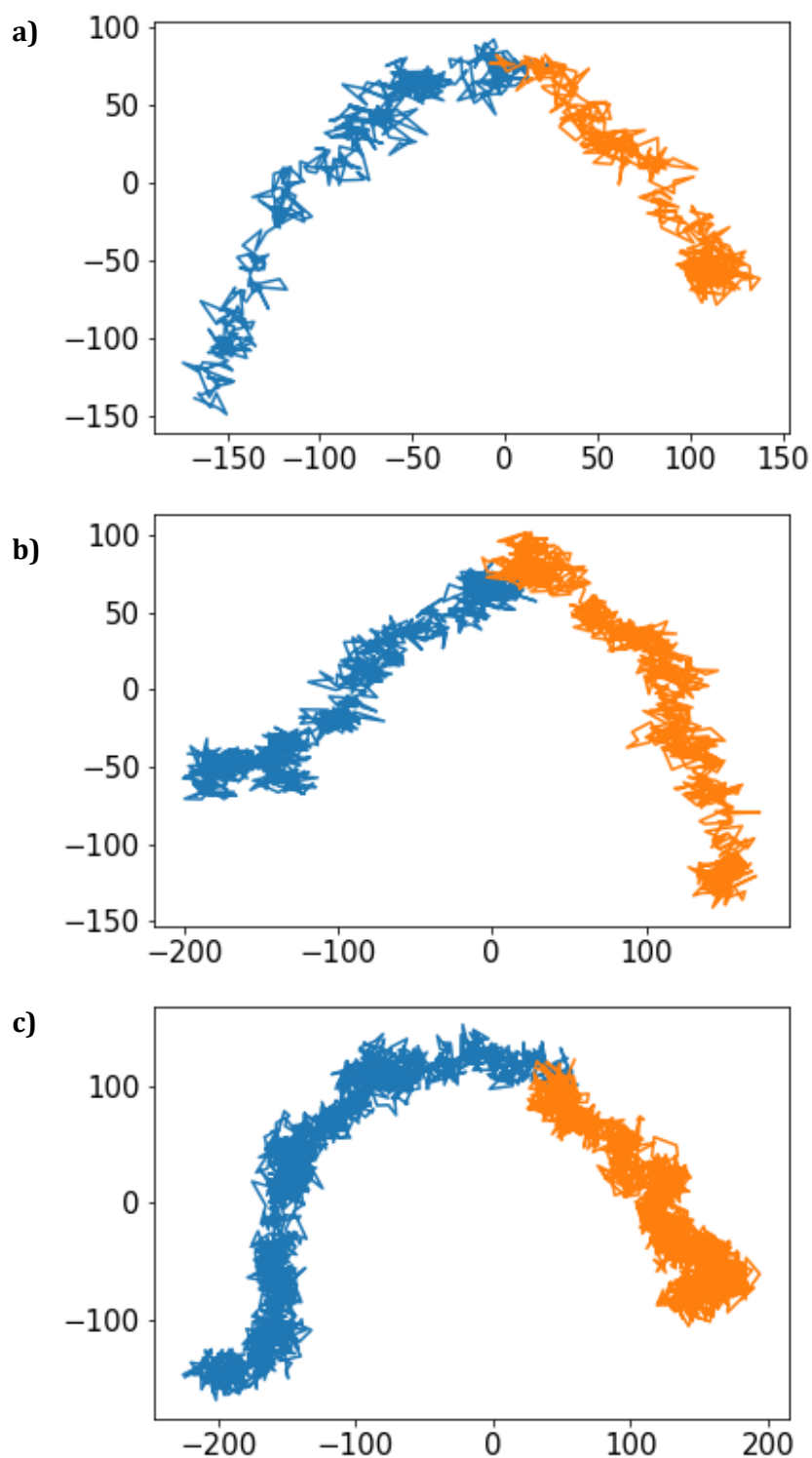
6.15 Porcupine plots of first and second principal components in *Bs*LAM.

PC1 (above) and PC2 (below) of *Bs*LAM with no SAM bound in the active site represented by porcupine plots.
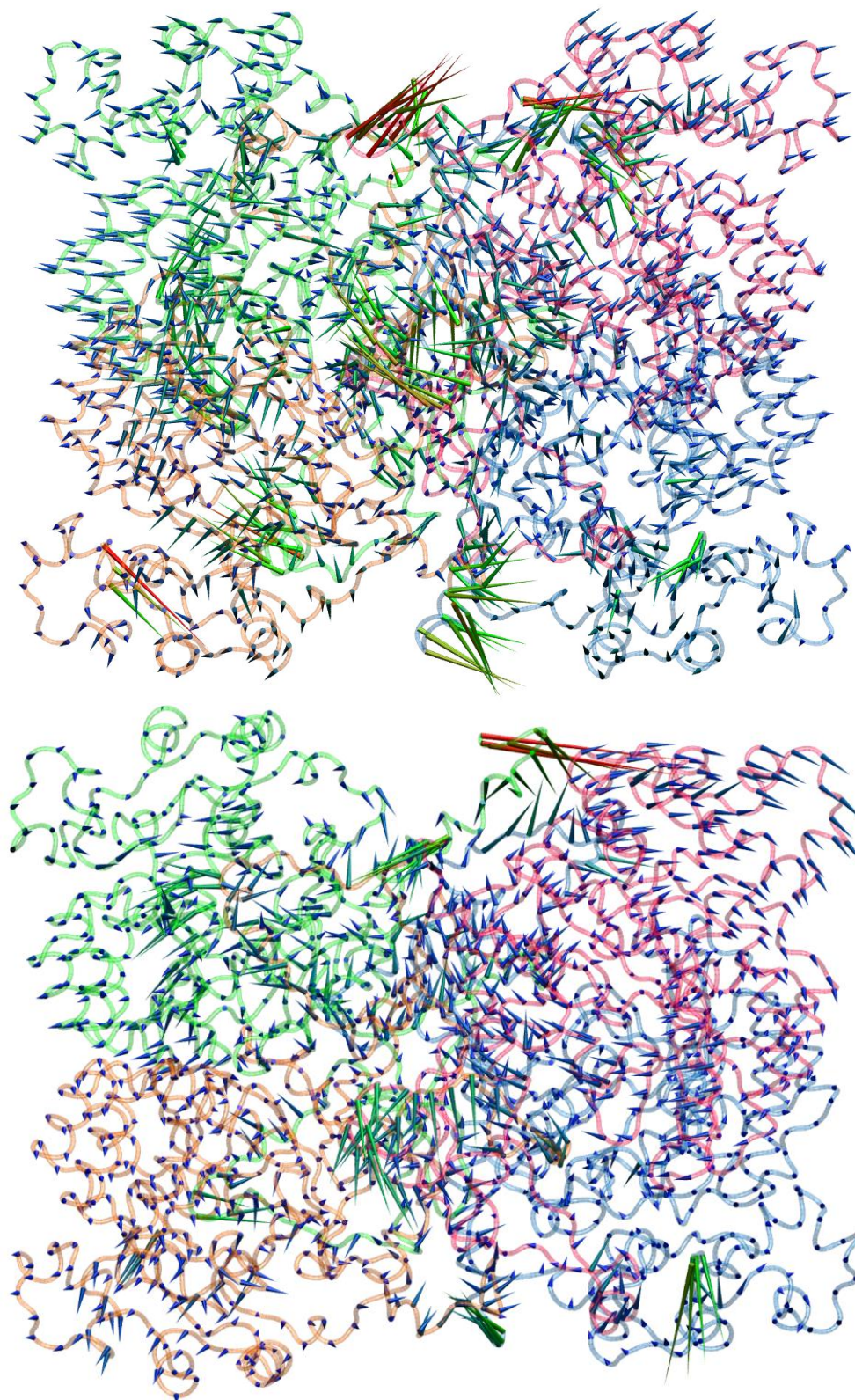
PC1 (above) and PC2 (below) of *Bs*LAM with SAM bound. Each monomer was coloured following the colour representation used in RMSD and RMSF plots.
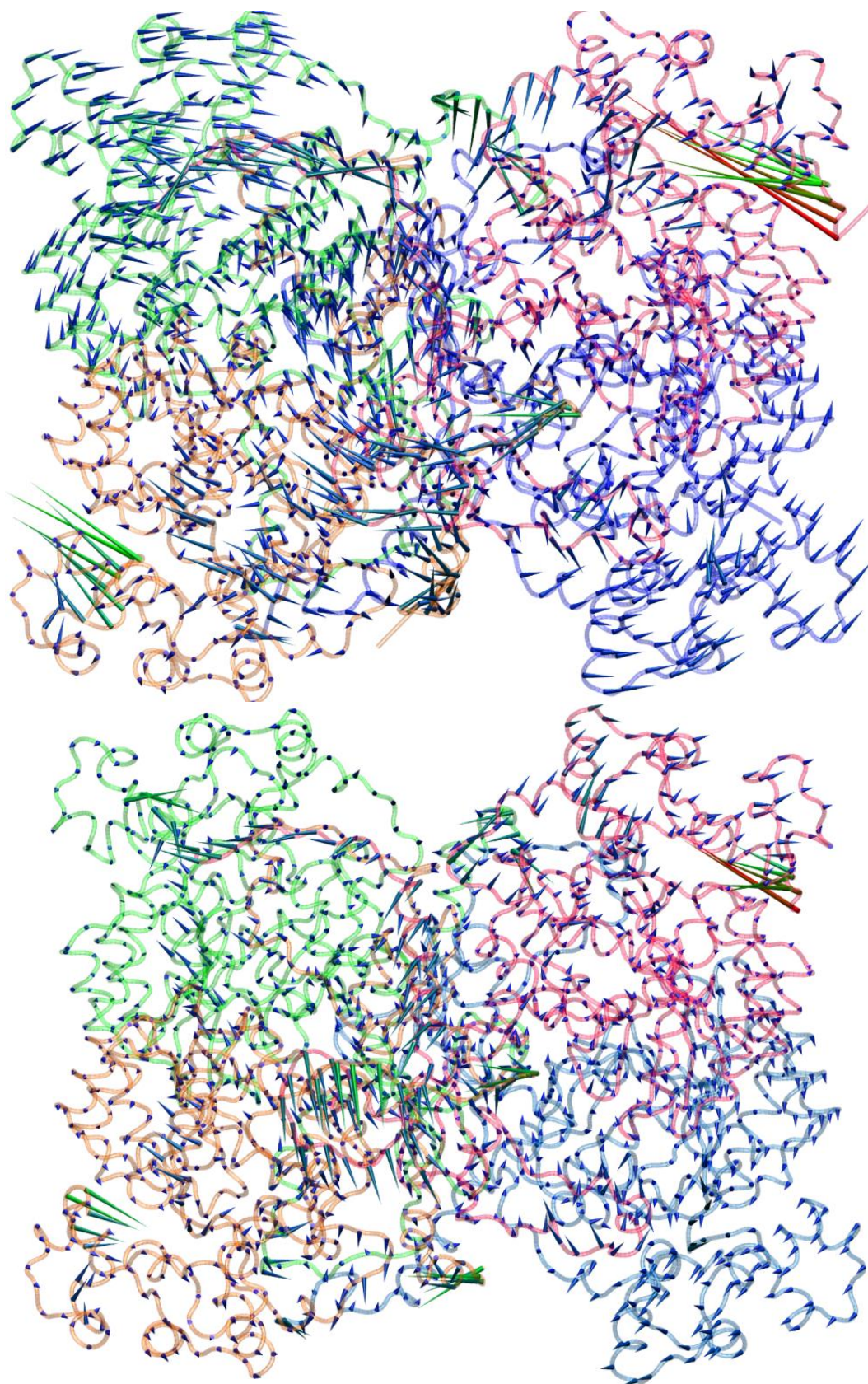
6.16   CAVER input file to start the tunnel search calculation.

The origin of the tunnels is located in the Fe atoms of the cluster.

```
#****************************
# CALCULATION SETUP
#****************************
load_tunnels no
load_cluster_tree no


#****************************
# INPUT DATA
#****************************
time_sparsity 1
first_frame 1
last_frame 100000



#****************************
# TUNNEL CALCULATION
#****************************

starting_point_atom 26134
starting_point_atom 26135
starting_point_atom 26136
starting_point_atom 26137

probe_radius 0.9
shell_radius 3
shell_depth 4


#****************************
# TUNNEL CLUSTERING
#****************************
```

clustering average_link

weighting_coefficient 1

clustering_threshold 4.0

#****************************

GENERATION OF OUTPUTS

#****************************

one_tunnel_in_snapshot cheapest

save_dynamics_visualization yes

generate_summary yes

generate_tunnel_characteristics yes

generate_tunnel_profiles yes

generate_histograms yes

bottleneck_histogram 0.0 2.0 20

throughput_histogram 0 1.0 10

generate_bottleneck_heat_map yes

bottleneck_heat_map_range 1.0 2.0

bottleneck_heat_map_element_size 10 20

generate_profile_heat_map yes

profile_heat_map_resolution 0.5

profile_heat_map_range 1.0 2.0

profile_heat_map_element_size 20 10

compute_tunnel_residues yes

residue_contact_distance 3.0

compute_bottleneck_residues yes

bottleneck_contact_distance 3.0

6.17 Tunnel-cluster characteristics in unbound-SAM LAM enzyme simulations.

	T. cluster ID	N° snapshots	Average bottleneck radius	Maximum bottleneck radius	Priority
CsLAM	1	21	1.252	1.49	37.7%
	2	24	1.228	1.49	35.1%
	3	12	1.170	1.52	18.0%
BsLAM	1	29	1.280	1.52	36.1%
	2	24	1.102	1.38	31.0%
	3	26	1.240	1.52	30.1%

6.18   Amino acid residues identified along the analysed clusters of tunnels.

The amino acids are the same that were identified and discussed in paragraph 4.1.4. The colours adopted describes the hydrophobic residues (in green), polar residues (yellow), negatively charged residues (red) and positively charged residues (blue).

a)

```
sp|Q9XBQ8|KAMA_CLOSU    KEEEEGVAQCVKSLRMAITPYYLSLIDPNDPNDPVRKQAIPTALELNKAA 91
sp|O34676|KAMA_BACSU    EDEEEGVRISTKTIPLNITPYYASLMDPDNPRCPVRMQSVPLSEEMHKTK 100
                         :* ***     .:: : ***** **:* :. . *:* *::*    *:::

sp|Q9XBQ8|KAMA_CLOSU    ADLEDPLHEDTDSPVPGLTHRYPDRVLLLITDMCSMYCRHCTRRRFAGQS 141
sp|O34676|KAMA_BACSU    YDLEDPLHEDEDSPVPGLTHRYPDRVLFLVTNQCSMYCRYCTRRFSGQI 150
                         *   ****** *****************:*:*: ******:******.*.

sp|Q9XBQ8|KAMA_CLOSU    DDSMPMERIDKAIDYIRNTPQVRDVLLSGGDALLVSDETLEYIIAKLREI 191
sp|O34676|KAMA_BACSU    GMGVPKKQLDAAIAYIRETPEIRDCLISGGDGLLINDQILEYILKELRSI 200
                         . .:   .::* ** ** :**::** *:****.**:.*: **   : .** *

sp|Q9XBQ8|KAMA_CLOSU    PHVEIVRIGSRTPVVLPQRITPELVNMLKKYHPVWLNTHFNHPNEITEES 241
sp|O34676|KAMA_BACSU    PHLEVIRIGTRAPVVFPQRITDHLCEILKKYHPVWLNTHFNTSIEMTEES 250
                         **:*::***:*.***:***** .* .:******:*******  . *:* *:

sp|Q9XBQ8|KAMA_CLOSU    TRACQLLADAGVPLGNQSVLLRGVNDCVHVMKELVNKLVKIRVRPYYIYQ 291
sp|O34676|KAMA_BACSU    VEACEKLVNAGVPVGNQAVVLAGINDSVPIMKKLMHDLVKIRVRPYYIYQ 300
                         .**: *.:**:*:***:*:* *:**.: :**.*::.** :*********

sp|Q9XBQ8|KAMA_CLOSU    CDLSLGLEHFRTPVSKGIEIIEGLRGHTSGYCVPTFVVDAPGGGGKTPVM 341
sp|O34676|KAMA_BACSU    CDLSEGIGHFRAPVSKGLEIIEGLRGHTSGYAVPTFVVDAPGGGGKIALQ 350
                         **** *: ***.:**:**:************.************** .:
```

b)

```
sp|Q9XBQ8|KAMA_CLOSU    KEEEEGVAQCVKSLRMAITPYYLSLIDPNDPNDPVRKQAIPTALELNKAA 91
sp|O34676|KAMA_BACSU    EDEEEGVRISTKTIPLNITPYYASLMDPDNPRCPVRMQSVPLSEEMHKTK 100
                         :* ***     .:: : ***** **:* :. . *:* *::*   *::::

sp|Q9XBQ8|KAMA_CLOSU    ADLEDPLHEDTDSPVPGLTHRYPDRVLLLITDMCSMYCRHCTRRRFAGQS 141
sp|O34676|KAMA_BACSU    YDLEDPLHEDEDSPVPGLTHRYPDRVLFLVTNQCSMYCRYCTRRRFSGQI 150
                          *  ****** ***************.*:*: ******:******.*.

sp|Q9XBQ8|KAMA_CLOSU    DDSMPMERIDKAIDYIRNTPQVRDVLLSGGDALLVSDETLEYIIAKLREI 191
sp|O34676|KAMA_BACSU    GMGVPKKQLDAAIAYIRETPEIRDCLISGGDGLLINDQILEYILKELRSI 200
                        . .:   .::* ** ** :**::** *:****.**:.*: **   . .** *

sp|Q9XBQ8|KAMA_CLOSU    PHVEIVRIGSRTPVVLPQRITPELVNMLKKYHPVWLNTHFNHPNEITEES 241
sp|O34676|KAMA_BACSU    PHLEVIRIGTRAPVVFPQRITDHLCEILKKYHPVWLNTHFNTSIEMTEES 250
                        **:*::***:*.***:***** .* .:******:*******  . *:* *:

sp|Q9XBQ8|KAMA_CLOSU    TRACQLLADAGVPLGNQSVLLRGVNDCVHVMKELVNKLVKIRVRPYYIYQ 291
sp|O34676|KAMA_BACSU    VEACEKLVNAGVPVGNQAVVLAGINDSVPIMKKLMHDLVKIRVRPYYIYQ 300
                        .**: *.:**:*:***:*.* *:**.: :**.*::.** :*********

sp|Q9XBQ8|KAMA_CLOSU    CDLSLGLEHFRTPVSKGIEIIEGLRGHTSGYCVPTFVVDAPGGGGKTPVM 341
sp|O34676|KAMA_BACSU    CDLSEGIGHFRAPVSKGLEIIEGLRGHTSGYAVPTFVVDAPGGGGKIALQ 350
                        **** *: ***:**:**:*************.************** .:
```

c)

```
sp|Q9XBQ8|KAMA_CLOSU    KEEEEGVAQCVKSLRMAITPYYLSLIDPNDPNDPVRKQAIPTALELNKAA 91
sp|O34676|KAMA_BACSU    EDEEEGVRISTKTIPLNITPYYASLMDPDNPRCPVRMQSVPLSEEMHKTK 100
                         :* ***     .:: : ***** **:* :. . *:* *::*   *::::

sp|Q9XBQ8|KAMA_CLOSU    ADLEDPLHEDTDSPVPGLTHRYPDRVLLLITDMCSMYCRHCTRRRFAGQS 141
sp|O34676|KAMA_BACSU    YDLEDPLHEDEDSPVPGLTHRYPDRVLFLVTNQCSMYCRYCTRRRFSGQI 150
                          *  ****** ***************.*:*: ******:******.*.

sp|Q9XBQ8|KAMA_CLOSU    DDSMPMERIDKAIDYIRNTPQVRDVLLSGGDALLVSDETLEYIIAKLREI 191
sp|O34676|KAMA_BACSU    GMGVPKKQLDAAIAYIRETPEIRDCLISGGDGLLINDQILEYILKELRSI 200
                        . .:   .::* ** ** :**::** *:****.**:.*: **   . .** *

sp|Q9XBQ8|KAMA_CLOSU    PHVEIVRIGSRTPVVLPQRITPELVNMLKKYHPVWLNTHFNHPNEITEES 241
sp|O34676|KAMA_BACSU    PHLEVIRIGTRAPVVFPQRITDHLCEILKKYHPVWLNTHFNTSIEMTEES 250
                        **:*::***:*.***:***** .* .:******:*******  . *:* *:

sp|Q9XBQ8|KAMA_CLOSU    TRACQLLADAGVPLGNQSVLLRGVNDCVHVMKELVNKLVKIRVRPYYIYQ 291
sp|O34676|KAMA_BACSU    VEACEKLVNAGVPVGNQAVVLAGINDSVPIMKKLMHDLVKIRVRPYYIYQ 300
                        .**: *.:**:*:***:*.* *:**.: :**.*::.** :*********

sp|Q9XBQ8|KAMA_CLOSU    CDLSLGLEHFRTPVSKGIEIIEGLRGHTSGYCVPTFVVDAPGGGGKTPVM 341
sp|O34676|KAMA_BACSU    CDLSEGIGHFRAPVSKGLEIIEGLRGHTSGYAVPTFVVDAPGGGGKIALQ 350
                        **** *: ***:**:**:*************.************** .:
```

6.19 RMSD of the [Fe₄-S₄] cluster in each monomer of LAM simulations.

From monomer A on top, to D on the bottom of *Cs*LAM (left row) and *Bs*LAM (right row) in both SAM-bound (green) and unbound-SAM (red) MD simulation.
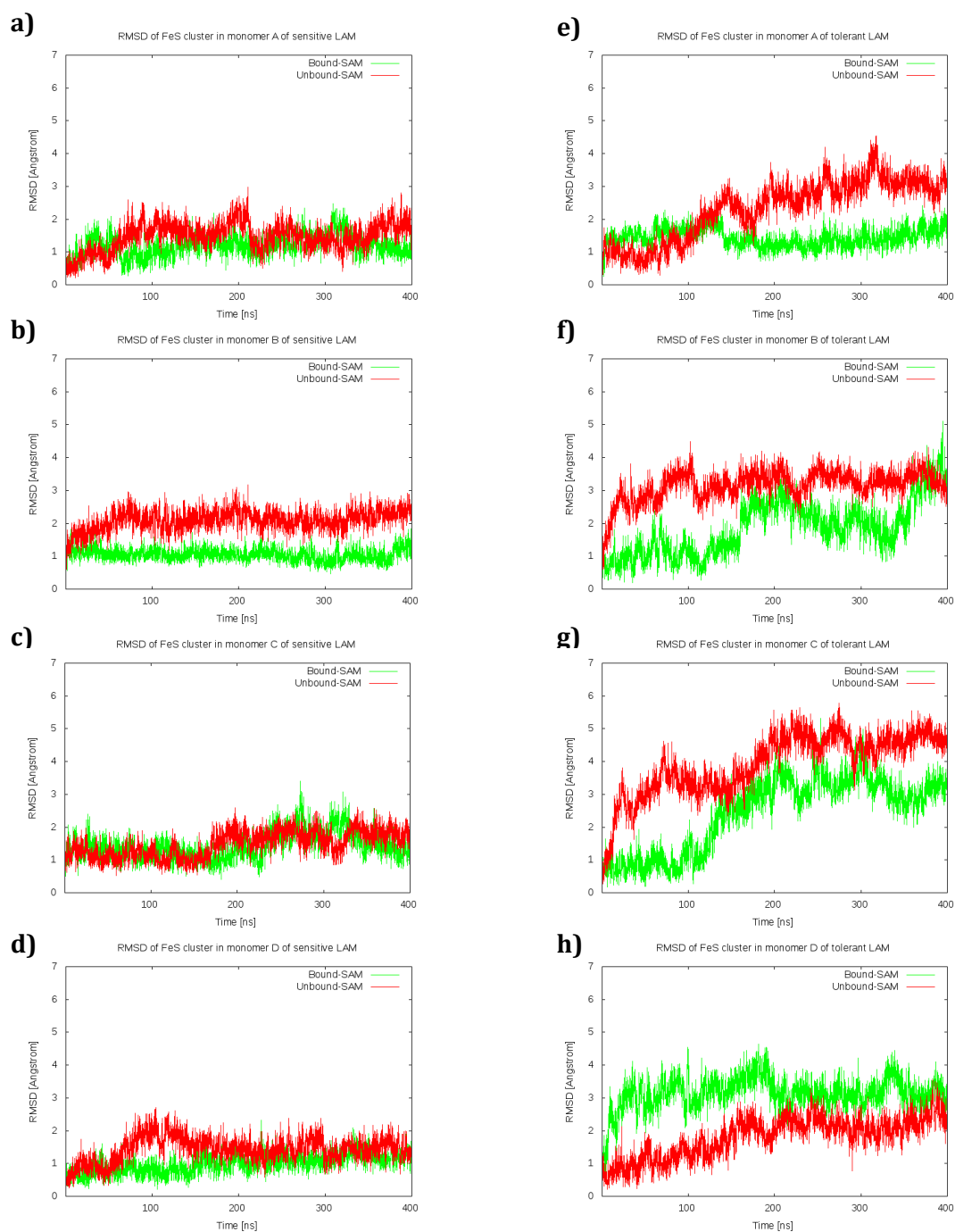
6.20 CPPTRAJ script to calculate dihedral angles of Cys-[Fe$_4$-S$_4$] of each cysteine residues by considering C$_\beta$ – S$_\gamma$ – FeS atoms.

```
parm SAM_noWatIons.prmtop
trajin SAM400nsReimagedStrppdReduced.nc
dihedral :123@CB :123@SG :1639@FE1 :1639@S4 out
CYS123_SAM_dihed.dat range360
dihedral :127@CB :127@SG :1639@FE4 :1639@S3 out
CYS127_SAM_dihed.dat range360
dihedral :130@CB :130@SG :1639@FE2 :1639@S1 out
CYS130_SAM_dihed.dat range360
dihedral :532@CB :532@SG :1642@FE1 :1642@S4 out
CYS532_SAM_dihed.dat range360
dihedral :536@CB :536@SG :1642@FE4 :1642@S3 out
CYS536_SAM_dihed.dat range360
dihedral :539@CB :539@SG :1642@FE2 :1642@S1 out
CYS539_SAM_dihed.dat range360
dihedral :942@CB :942@SG :1645@FE1 :1645@S4 out
CYS942_SAM_dihed.dat range360
dihedral :946@CB :946@SG :1645@FE4 :1645@S3 out
CYS946_SAM_dihed.dat range360
dihedral :949@CB :949@SG :1645@FE2 :1645@S1 out
CYS949_SAM_dihed.dat range360
dihedral :1351@CB :1351@SG :1648@FE1 :1648@S4 out
CYS1351_SAM_dihed.dat range360
dihedral :1355@CB :1355@SG :1648@FE4 :1648@S3 out
CYS1355_SAM_dihed.dat range360
dihedral :1358@CB :1358@SG :1648@FE2 :1648@S1 out
CYS1358_SAM_dihed.dat range360
```

6.21 Dihedral angles of 3-cysteine cluster-binding ligands in unbound-SAM MD simulations.

The *Cs*LAM (a, b, c, d) and *Bs*LAM (e, f, g, h).
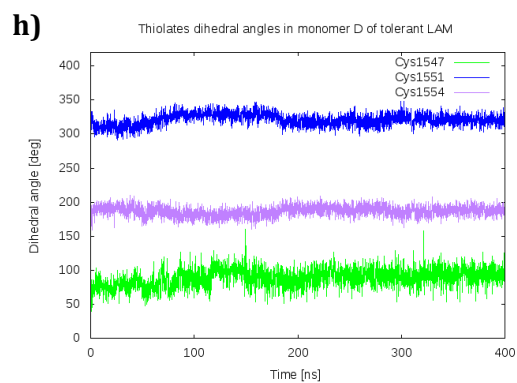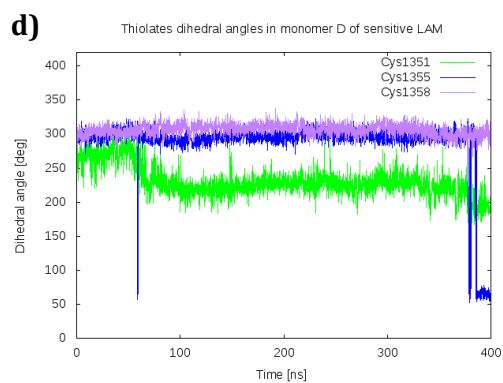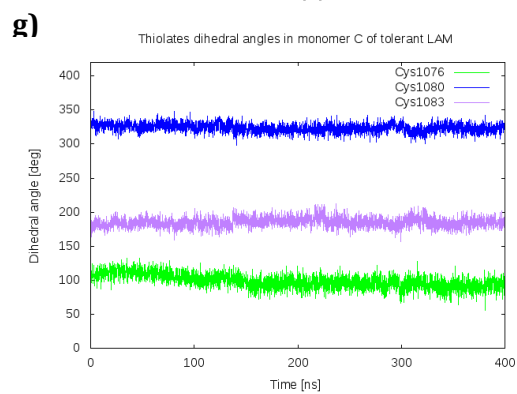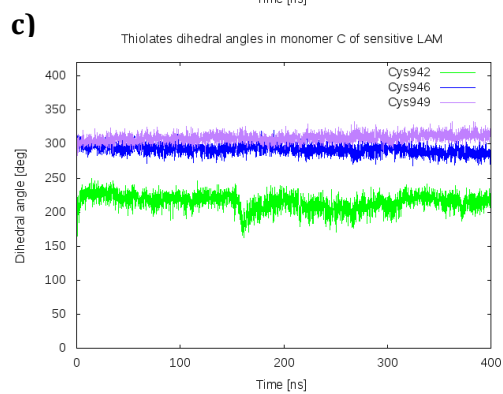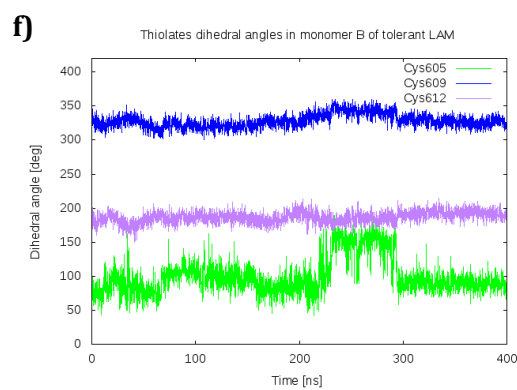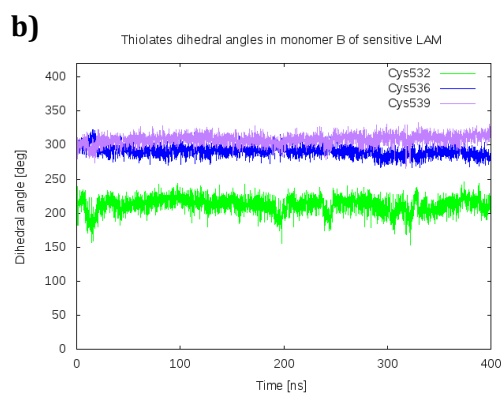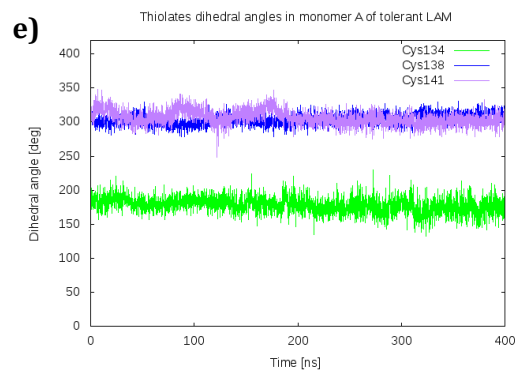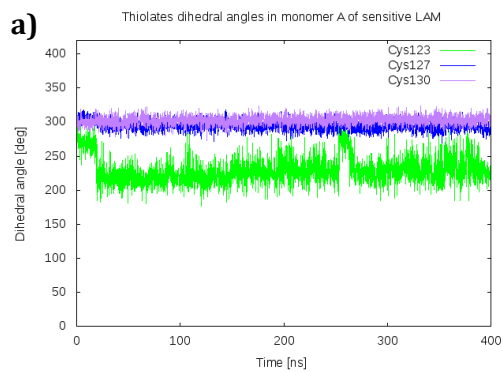
6.22 Dihedral angles of 3-cysteine cluster-binding ligands in the bound-SAM MD simulations.

The *Cs*LAM (a, b, c, d) and *Bs*LAM (e, f, g, h).
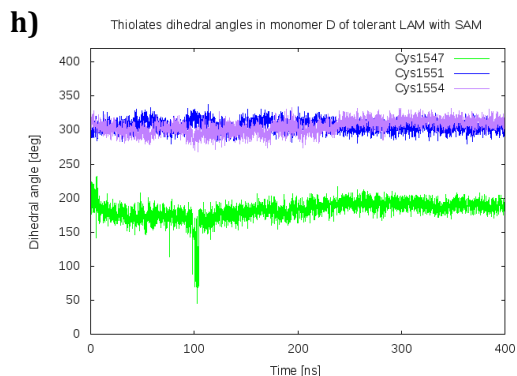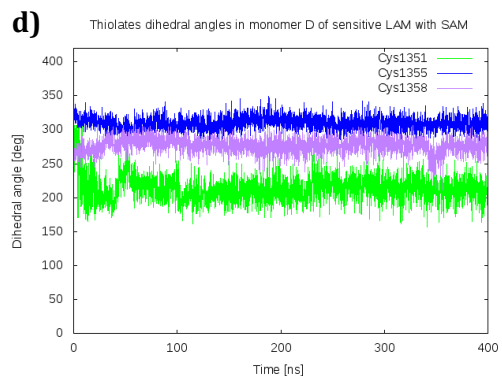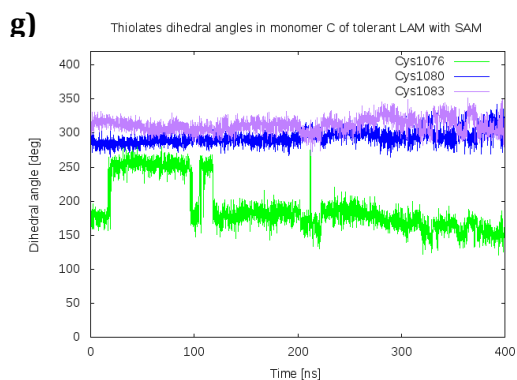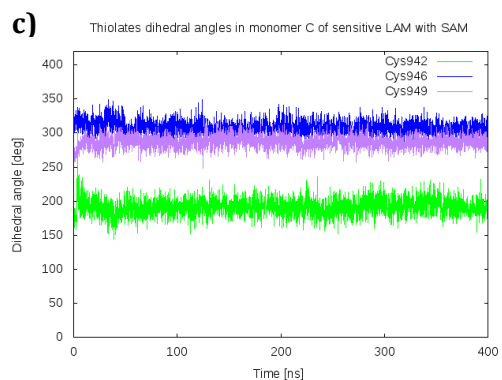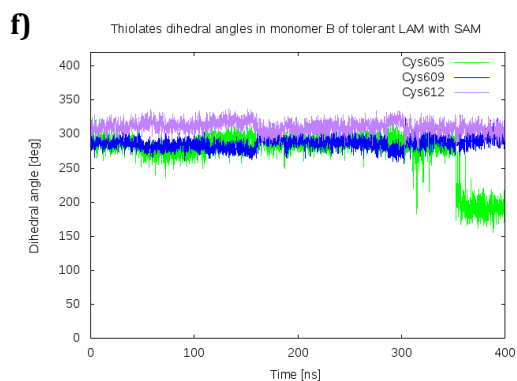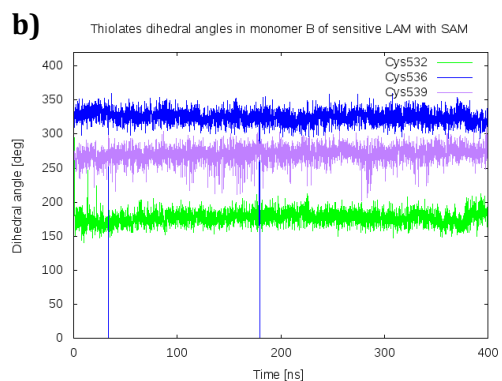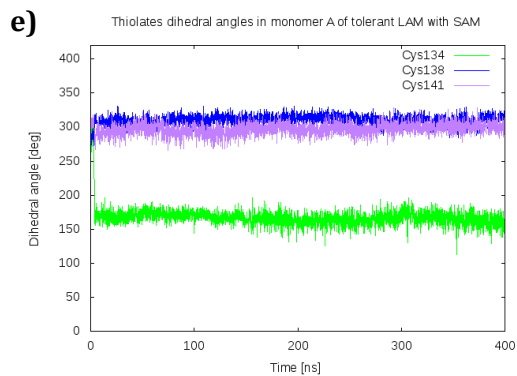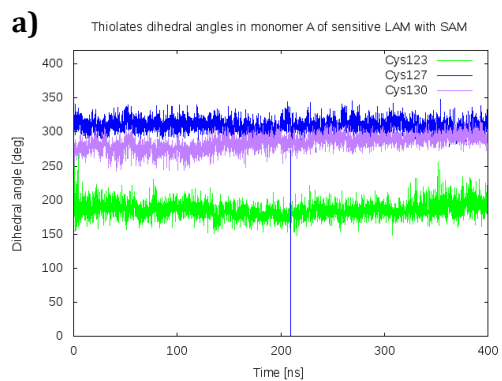
6.23 Gaussian QM optimisation file of [Fe₄-S₄](SCH₃)₃ system.

6.23.1 First step optimisation at the HF level.

%chk=tSF4OxSCH3_OPT_250QM.chk
%NProcShared=10
%mem=80GB
#P maxdisk=64gb OPT=tight HF/6-31G* SCF=(XQC) int=(grid=ultrafine)
noSymm

Gaussian run using SANDER external interface, route from template
gau_job.tpl

-1 1

C	2.4993543464725354	2.4660065040699406	3.2773337895165677
H	2.7043460008414764	2.2082145016640977	2.2382849478055031
H	2.4373390262731438	1.5311496247192151	3.8343766204964278
S	0.9103088389732008	3.3716425302655422	3.4415323035111669
C	-0.0536455660217499	-2.5601121305894479	-0.9354431855797286
H	-0.5583971338302758	-1.7209547100517102	-1.4141306919406702
H	-0.7813238082371259	-3.3716425302655422	-0.9371124830505763
S	0.2022982217444919	-2.2580486727098008	0.8618524700462871
C	-0.9898408543712163	0.9344919074120277	-3.3065027880194222
H	-0.8357104295171141	0.0445233579823601	-2.6963457068015231
H	-0.0396706237751732	1.0158910249986339	-3.8343766204964354
S	-1.4689553281823962	2.3945080862735066	-2.3187975191359769
Fe	-0.3955762949258953	2.1476335466608170	2.0786311041509435
Fe	-0.5108647434856479	-0.2410744612759678	1.0971120602744688
Fe	-2.7043460008414755	1.1821965369202689	1.7216757974934822
Fe	-1.2857294975880147	1.7985384582757253	-0.2699267970256329
S	-2.3496154189022675	-0.0173629453266493	-0.1222631892824282
S	-2.0611623038708231	3.2960257646675286	1.1336930024438843
S	0.8474761789942211	1.3930909411704875	0.3881029387064032
S	-1.2600155260739925	0.2590515851302191	3.1379689589975968
H	3.3642863779403549	2.9995545280402736	3.6714392158409259
H	0.8911191226796608	-2.7333313488028002	-1.4507244538444328
H	-1.8057433155823421	0.6825404845943625	-3.9839418290628075

6.23.2 Second step optimisation at the B3LYP level.

Gaussian optimisation file of [Fe$_4$-S$_4$](SCH$_3$)$_3$ system.

%chk=tSF4OxSCH3_B3LYP_250QM.chk
%NProcShared=10
%mem=80GB
#P maxdisk=64gb OPT=tight B3LYP/6-31G* SCF=(tight) int=(grid=ultrafine)
NoSymm

tSF4OxSCH3_B3LYP_250QM

-1 1

C	2.76713	3.26977	3.01312
H	2.59143	3.67304	2.02491
H	3.14896	2.2618	2.91319
S	1.2505	3.30389	4.03505
C	-0.62017	-3.53649	-0.93572
H	-1.69157	-3.40793	-0.86451
H	-0.41021	-4.57097	-1.17855
S	0.21571	-3.13332	0.63921
C	-0.43733	1.53852	-3.32576
H	-1.0593	0.65463	-3.3586
H	0.5669	1.26459	-3.03191
S	-1.13095	2.78771	-2.18888
Fe	-0.43996	2.19021	2.82444
Fe	-0.56788	-1.01818	1.33165
Fe	-3.20427	0.78677	1.76604
Fe	-1.28019	1.7322	-0.31718
S	-2.39776	-0.13894	-0.28937
S	-2.31416	2.9222	1.21465
S	0.53045	1.23278	0.60347
S	-1.62678	0.0678	3.28709
H	3.52083	3.87392	3.50269
H	-0.25255	-2.91224	-1.73982
H	-0.40292	1.98076	-4.31275

6.23.3 Third step optimisation using the custom hybrid functional at the BP86 level.

%chk=tSF4OxSCH3_BP86_250QM.chk
%NProcShared=10
%mem=80GB
#P maxdisk=64gb OPT=tight BP86/6-311+G(d)
iop(3/76=1000000500,3/77=0720009500,3/78=0810010000) SCF=(verytight)
int=(grid=ultrafine) NoSymm

tSF4OxSCH3_BP86_250QM

-1 1
C	2.659803	3.096525	3.488598
H	2.794108	3.554830	2.506135
H	3.007678	2.061904	3.445968
S	0.889036	3.163902	3.997352
C	-0.443710	-3.352412	-0.666015
H	-1.532672	-3.449982	-0.671848
H	0.004794	-4.350969	-0.684634
S	0.121621	-2.490891	0.862247
C	-1.137054	1.407876	-3.671344
H	-2.213933	1.225972	-3.619729
H	-0.618245	0.446992	-3.618700
S	-0.603163	2.528272	-2.308592
Fe	-0.390361	2.117134	2.633632
Fe	-0.565927	-0.438076	0.836685
Fe	-2.298077	0.933121	1.191472
Fe	-0.848844	1.527055	-0.404819
S	-2.131465	-0.232173	-0.687965
S	-1.701063	2.977363	1.089602
S	0.845977	1.241701	0.965600
S	-1.293212	0.125318	2.889276
H	3.249599	3.646441	4.228062
H	-0.137678	-2.806956	-1.562711
H	-0.901301	1.889573	-4.625813

6.24 Gaussian single-point energy (SP) calculation input files.

Energy calculation of the oxidised state using the custom hybrid functional at the BP86 level.

```
%chk=tSF4OxSCH3_SP_BP86_250QM.chk
%NProcShared=10
%mem=80GB
#P maxdisk=64gb BP86/6-311+G(d)
iop(3/76=1000000500,3/77=0720009500,3/78=0810010000) SCF=(tight)
int=(grid=ultrafine) NoSymm guess=mix

tSF4OxSCH3_BP86_250QM

-1 1
C       2.402291    2.210080    3.476173
H       2.956682    1.719192    2.662355
H       1.998869    1.434442    4.143557
S       1.059392    3.263204    2.800337
C       0.717763   -3.095834    0.098682
H       0.158398   -3.092663   -0.848256
H       1.028070   -4.128803    0.324618
S      -0.314262   -2.472911    1.482421
C      -1.918051    2.022422   -3.733386
H      -2.441070    2.917421   -3.364155
H      -2.578157    1.150758   -3.606640
S      -0.331531    1.772778   -2.844535
Fe     -0.218962    1.942360    1.657303
Fe     -0.839732   -0.433579    0.969763
Fe     -2.551051    1.348087    1.042478
Fe     -0.859607    1.433983   -0.771939
S      -2.344771   -0.086487   -0.511353
S      -1.526291    3.090099    0.389787
S       0.847558    0.685474    0.295855
S      -1.494957    0.612551    2.730329
H       3.096411    2.845657    4.049263
H       1.615960   -2.471601   -0.023350
H      -1.707042    2.155892   -4.806849
```

Energy calculation of the reduced state using the custom hybrid functional at the BP86 level.

%chk=tSF4RedSCH3_SP_BP86_250QM.chk

%NProcShared=10

%mem=80GB

#P maxdisk=64gb BP86/6-311+G(d)

iop(3/76=1000000500,3/77=0720009500,3/78=0810010000) SCF=(tight)

int=(grid=ultrafine) NoSymm guess=mix

tSF4RedSCH3_BP86_250QM

-2 2

C	2.402291	2.210080	3.476173
H	2.956682	1.719192	2.662355
H	1.998869	1.434442	4.143557
S	1.059392	3.263204	2.800337
C	0.717763	-3.095834	0.098682
H	0.158398	-3.092663	-0.848256
H	1.028070	-4.128803	0.324618
S	-0.314262	-2.472911	1.482421
C	-1.918051	2.022422	-3.733386
H	-2.441070	2.917421	-3.364155
H	-2.578157	1.150758	-3.606640
S	-0.331531	1.772778	-2.844535
Fe	-0.218962	1.942360	1.657303
Fe	-0.839732	-0.433579	0.969763
Fe	-2.551051	1.348087	1.042478
Fe	-0.859607	1.433983	-0.771939
S	-2.344771	-0.086487	-0.511353
S	-1.526291	3.090099	0.389787
S	0.847558	0.685474	0.295855
S	-1.494957	0.612551	2.730329
H	3.096411	2.845657	4.049263
H	1.615960	-2.471601	-0.023350
H	-1.707042	2.155892	-4.806849

6.25 Dihedral angles of cysteinyl ligands in [Fe$_4$-S$_4$](Cys)$_3$ systems.
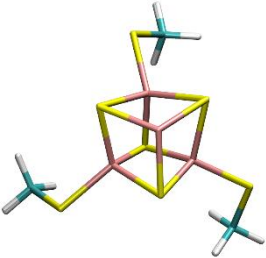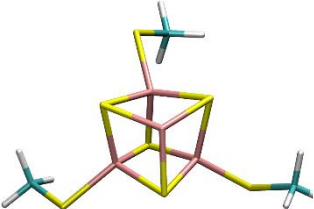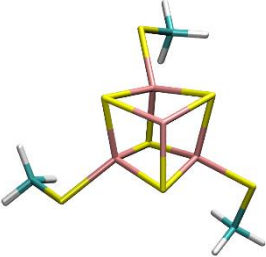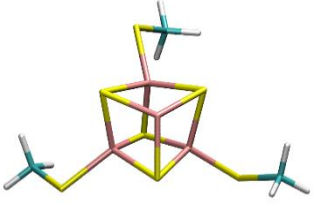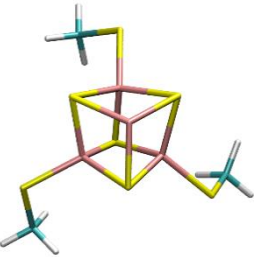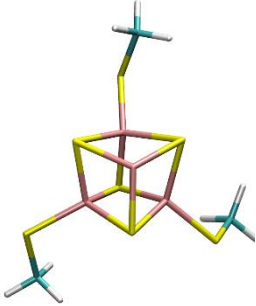
Dihedral angles of cysteinyl ligands in both LAM enzymes before the QM geometry optimisation.

MD snapshots (simulation time)	Pre-optimisation *Cs*LAM geometry			Pre-optimisation *Bs*LAM geometry		
	Cys123 dihedral angle [deg]	Cys127 dihedral angle [deg]	Cys130 dihedral angle [deg]	Cys134 dihedral angle [deg]	Cys138 dihedral angle [deg]	Cys141 dihedral angle [deg]
10	-160,21	-44,71	-160,27	-170.95	-74.53	-56.75
70	-160,37	-44,59	-160,31	-143.74	-61.98	-47.75
150	-165,2	-46,76	51,11	-127.43	-66.99	-59.03
190	-160,5	-44,46	-160,54	-164.36	-68.99	-57.24
210	-165,18	-46,55	50,99	-81.66	-52.14	-66.53
250	-160,22	-44,63	-160,3	-97.08	-70.7	-54.44
290	-78,35	44,7	-78,07	-143.93	-64.93	-44.96
310	-160,39	-44,6	-160,22	-119.48	-65.05	-55.9
350	-160,39	-44,64	-160,2	-146.52	-62.44	-56.47
400	-159,53	-45,03	-161,02	-159.44	-68.45	-53.47

Measured dihedral angles of cysteinyl ligands in $[Fe_4-S_4](Cys)_3$ systems in both LAM enzymes after QM geometry optimisation.

MD snapshots (simulation time)	Optimised *Cs*LAM geometry			Optimised *Bs*LAM geometry		
	Cys123 dihedral angle [deg]	Cys127 dihedral angle [deg]	Cys130 dihedral angle [deg]	Cys134 dihedral angle [deg]	Cys138 dihedral angle [deg]	Cys141 dihedral angle [deg]
10	-167,91	-58,4	-55,76	-165.34	-46.7	51.15
70	-147,73	-58,98	-56,09	-161.97	-44.84	-159.09
150	-145,38	-75,34	-48,03	-165.06	-46.79	51.01
190	-149,23	-74,79	-47,33	-78.35	44.78	-78.32
210	-123,85	-64,54	-62,91	-165.11	-46.8	51.22
250	-146,61	-57,72	-55,01	-78.42	44.66	-78.29
290	-137,81	-78,63	-48,81	-164.96	-46.66	51.11
310	-145,76	-81,19	-48,09	-165.13	-46.76	51.2
350	-147,23	-61,75	-42,79	-73.94	-166.75	-75.94
400	-145,88	-74,61	-43,73	-160.89	-45.48	-159.32

In the table, dihedral angle clusters collected after 3-step geometry optimisation. Structures with equal cysteine orientations were clustered in both LAM enzymes. *Bs*LAM reported the formation of two more conformations than *Cs*LAM. In brackets, the number of structures composing the cluster.

	I	II	III	IV
CsLAM	(x7)	(x3)	-	-
BsLAM	(x5)	(x2)	(x2)	(x1)

6.26 QM-calculated energy levels of [Fe₄-S₄](SCH₃)₃ systems in both LAM enzymes.

The trend in the oxidised state (blue) and the reduced state (red).

a)

Energy level from CsLAM snapshots

b)

Energy level from BsLAM snapshots
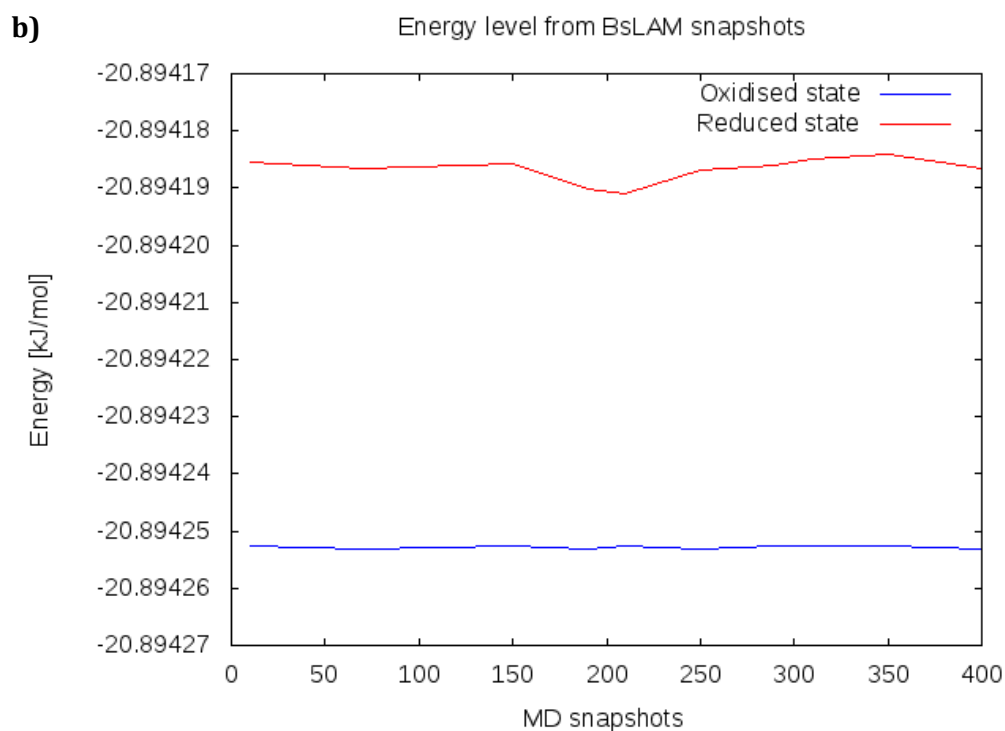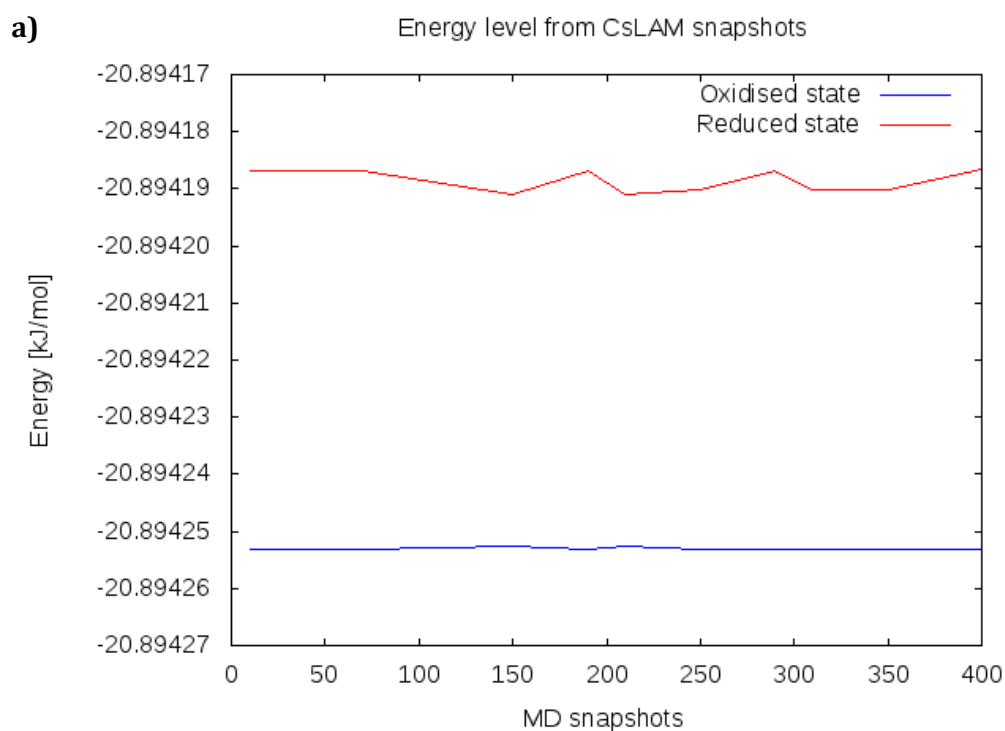
6.27 Calculation of unpaired *t* test.

Input values and outcomes are reported in the following table. By conventional criteria, the reported difference is considered to be not statistically significant. The two-tailed P value equals 0.1352

	*Cs*LAM	*Bs*LAM
Mean	1755.8977	1797.1100
SD	60.8343	56.9361
SEM	19.2375	18.0047
N	10	10

Intermediate values used in calculations:

t	1.5641
Df	18
standard error of difference	26.349

6.28   CPPTRAJ script to calculate H-bonds.

The H-bond network was calculated by considering as acceptor mask the cysteine residues and FeS cluster sulfur atoms.

```
parm SAM_noWatIons.prmtop
trajin SAM400nsReimagedStrppdReduced.nc
hbond dist 4.0 acceptormask :1642,532,536,539 out hbondBacc.dat avgout
hbondAVG_Bacc4.dat
hbond dist 4.0 acceptormask :1639,123,127,130 out hbondAacc.dat avgout
hbondAVG_Aacc4.dat
hbond dist 4.0 acceptormask :1645,942,946,949 out hbondCacc.dat avgout
hbondAVG_Cacc4.dat
hbond dist 4.0 acceptormask :1648,1351,1355,1358 out hbondDacc.dat
avgout hbondAVG_Dacc4.dat
```

6.29 Visual inspection of H-bonding residues on static structures.

Identification of hydrogen-bonding residues around the [Fe$_4$–S$_4$] cluster in *Cs*LAM (left) and *Bs*LAM (right) active site. The inspection was run on the crystal (2a5h.pdb) and on the homology model structures using the *within* command on VMD. The H-bond distance was set at 4.5 Å instead of 4 Å to have a wider view of possible H-bond interacting residues. Below the table, the corresponding residues are enlisted foe each LAM enzyme.
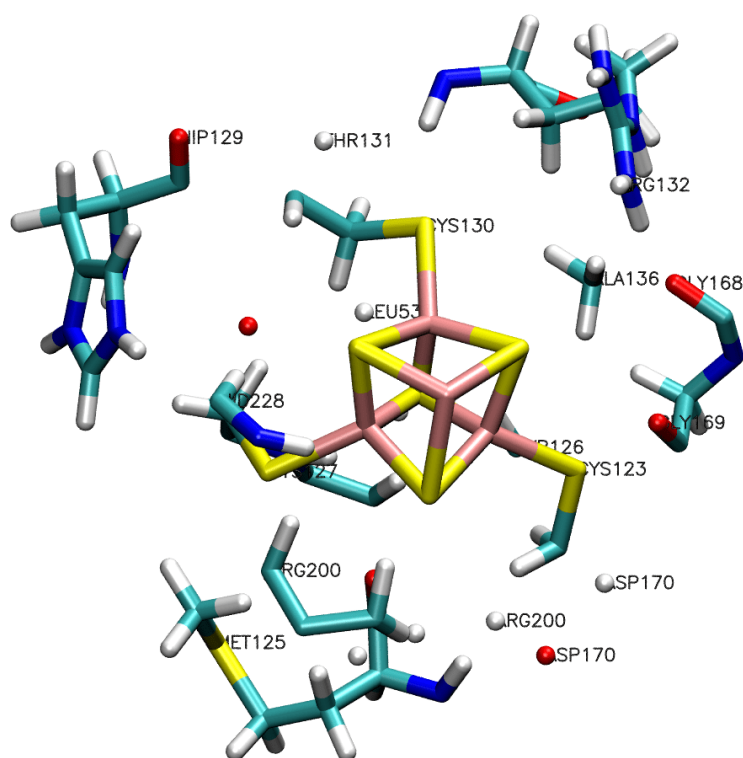
Cs*LAM: CYS125, CYS129, CYS132,
LEU55, MET127, TYR128, HIP131, THR133, ARG134,
ALA138, GLY170, GLY171, ASP172, ARG202, HID230
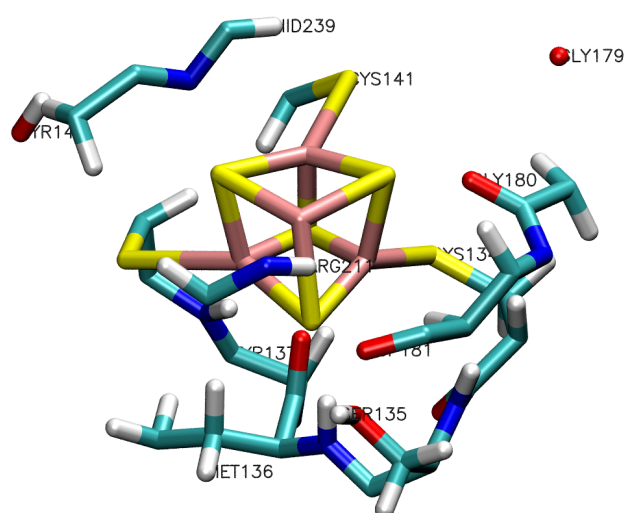
Bs*LAM: CYS134, CYS138, CYS141
SER135, MET136, TYR137, TYR140, GLY179,
GLY180, ASP181, ARG211, HID239

6.30 Detail of Serine136 interacting with S atoms in the active site.

During the MD simulations of both *Bs*LAM and *Cs*LAM H131Y/A136S mutant, it was possible to notice the hydroxyl group of Ser136 (*Bs*LAM Ser148) orienting towards the cuboidal cluster and cysteine ligands. By the shortest measured bonds, we could observe the likely interactions with Sγ of two cysteine ligands and the inorganic S atoms of the [Fe$_4$–S$_4$] cluster. Such an interaction may positively modulate the redox potential of the active site conferring a higher tolerance to the attack of ROS species.
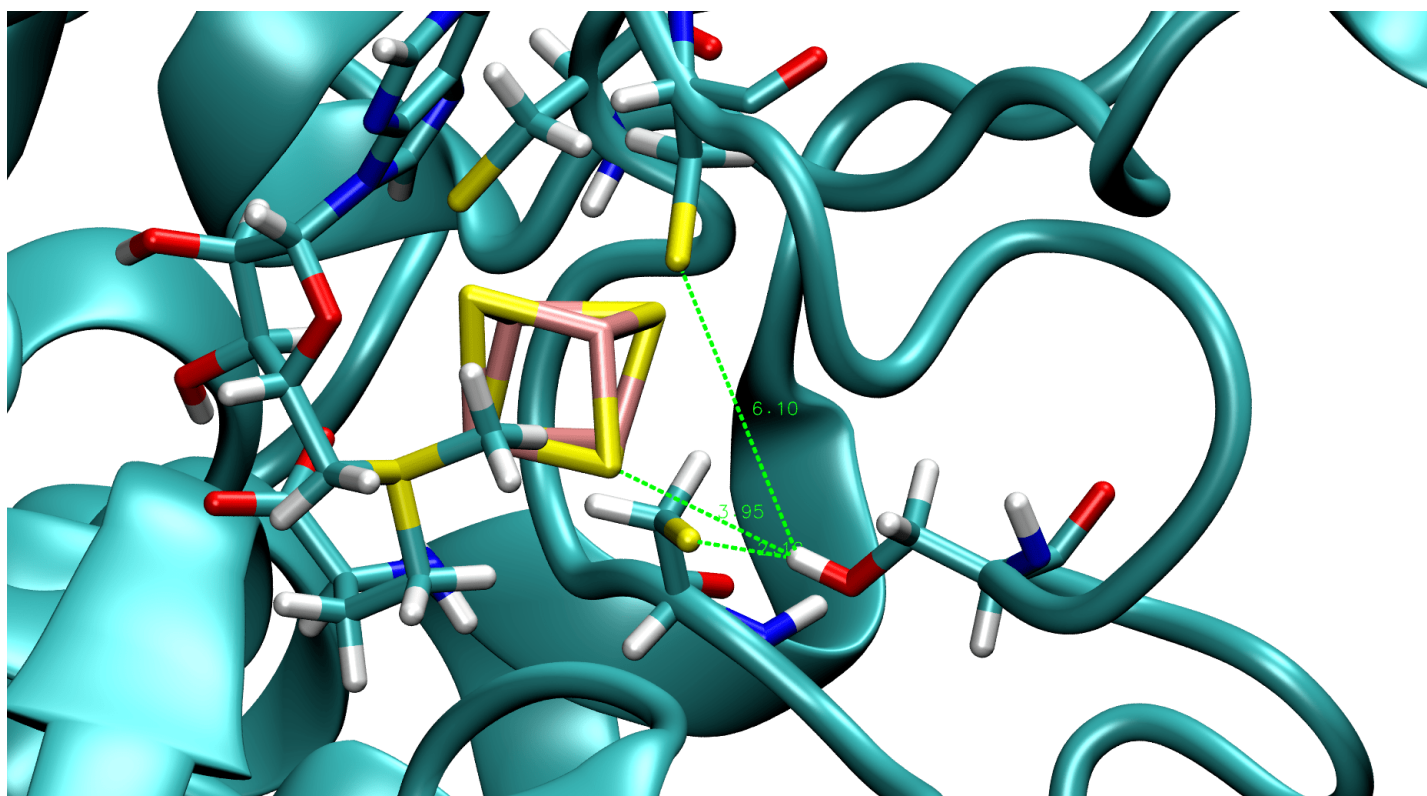
6.31 Manual PDB editing to insert raw residue mutations.

The final atoms and their correct distribution into space without generating clashes with neighbouring atoms will be automatically created by *xleap*. In the example below is reported the mutation of the starting Met124 into Gln124.

```
ATOM    995  N   ASP A 123
ATOM    996  CA  ASP A 123
ATOM    997  C   ASP A 123
ATOM    998  O   ASP A 123
ATOM    999  CB  ASP A 123
ATOM   1000  CG  ASP A 123
ATOM   1001  OD1 ASP A 123
ATOM   1002  OD2 ASP A 123
ATOM   1003  N   MET A 124
ATOM   1004  CA  MET A 124
ATOM   1005  C   MET A 124
ATOM   1006  O   MET A 124
ATOM   1007  CB  MET A 124
ATOM   1008  CG  MET A 124
ATOM   1009  SD  MET A 124
ATOM   1010  CE  MET A 124
ATOM   1011  N   CYF A 125
ATOM   1012  CA  CYF A 125
ATOM   1013  C   CYF A 125
ATOM   1014  O   CYF A 125
ATOM   1015  CB  CYF A 125
ATOM   1016  SG  CYF A 125
ATOM   1017  N   SER A 126
ATOM   1018  CA  SER A 126
```

$\Rightarrow$
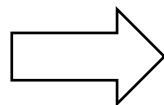
```
ATOM    995  N   ASP A 123
ATOM    996  CA  ASP A 123
ATOM    997  C   ASP A 123
ATOM    998  O   ASP A 123
ATOM    999  CB  ASP A 123
ATOM   1000  CG  ASP A 123
ATOM   1001  OD1 ASP A 123
ATOM   1002  OD2 ASP A 123
ATOM   1003  N       A 124
ATOM   1004  CA      A 124
ATOM   1005  C       A 124
ATOM   1006  O       A 124
ATOM   1007  CB      A 124
ATOM   1008  CG      A 124
ATOM   1011  N   CYF A 125
ATOM   1012  CA  CYF A 125
ATOM   1013  C   CYF A 125
ATOM   1014  O   CYF A 125
ATOM   1015  CB  CYF A 125
ATOM   1016  SG  CYF A 125
ATOM   1017  N   SER A 126
ATOM   1018  CA  SER A 126
```

$\Rightarrow$
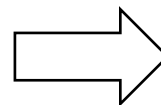
```
ATOM    995  N   ASP A 123
ATOM    996  CA  ASP A 123
ATOM    997  C   ASP A 123
ATOM    998  O   ASP A 123
ATOM    999  CB  ASP A 123
ATOM   1000  CG  ASP A 123
ATOM   1001  OD1 ASP A 123
ATOM   1002  OD2 ASP A 123
ATOM   1003  N   GLN A 124
ATOM   1004  CA  GLN A 124
ATOM   1005  C   GLN A 124
ATOM   1006  O   GLN A 124
ATOM   1007  CB  GLN A 124
ATOM   1008  CG  GLN A 124
ATOM   1011  N   CYF A 125
ATOM   1012  CA  CYF A 125
ATOM   1013  C   CYF A 125
ATOM   1014  O   CYF A 125
ATOM   1015  CB  CYF A 125
ATOM   1016  SG  CYF A 125
ATOM   1017  N   SER A 126
ATOM   1018  CA  SER A 126
```

6.32 CPPTRAJ script to retrieve the dipole moment and the vector.

The dipole moment of the residues selection can be retrieved by the following script as well as the orthogonal vector to the plane defined into the cluster.

```
parm SAM_noWatIons.prmtop
trajin SAM400nsReimagedStrppdReduced.nc
reference SAM_linear.pdb :1-1638@CA,C,N
vector v2 dipole out SAM_DResselA.dat /
:123,127,130,122,170,126,125,129,131,132,200,136,228 magnitude
vector v3 dipole out SAM_DResselB.dat /
:532,536,539,531,579,535,534,538,540,541,609,545,637 magnitude
vector v4 dipole out SAM_DResselC.dat /
:942,946,949,941,989,945,944,948,950,951,1019,955,1047 magnitude
vector v5 dipole out SAM_DResselD.dat /
:1351,1355,1358,1350,1398,1354,1353,1357,1359,1360,1428,1364,1456
magnitude
vector vI out SAM_FeSnormalA.dat :1639@S2,S3,FE2,FE3 magnitude corrplane
vector vII out SAM_FeSnormalB.dat :1642@S2,S3,FE2,FE3 magnitude corrplane
vector vIII out SAM_FeSnormalC.dat :1645@S2,S3,FE2,FE3 magnitude corrplane
vector vIV out SAM_FeSnormalD.dat :1648@S2,S3,FE2,FE3 magnitude corrplane
```

6.33 Custom python script for operations between vectors.

Retrieve the angle between the dipole moment and plane defined by normal vector of the plane included in the [Fe$_4$S$_4$] cluster.

```
##
##########################################
#                                        #
#  angle between vector (a) and  plane   #
#  defined by normal vector of plane (b) #
#                                        #
##########################################
#
#
sub startplaneangle
{
print "\n calculating the angle between vector a and a plane defined
by normal vector b\n";
print "name of vector A?\n";
$A = <STDIN>;
chomp $A;

print "name of vector B?\n";
$B = <STDIN>;
chomp $B;
#
print "and name of file to print angle\n";
$C = <STDIN>;
chomp $C;
#
# reading in vector A:
#
&readinA;
#
# reading in vector B:
#
&readinB;
#
# Do the vector files have the same size ?
#
if ($#framenr != $#Bframenr)
{
   print "CAUTION!!!!!!\n\n";
   print "vectors do not have the same dimension!!!\n\n" and die;
};
# vector norms:
for ($i=0; $i<=$#framenr; $i++)
{
   $normA[$i] = sqrt($vx[$i]*$vx[$i] + $vy[$i]*$vy[$i] +
$vz[$i]*$vz[$i]);
   $normA[$i] = sprintf('%.4f',$normA[$i]);
   push(@normA, $normA[$i]);
};
```

```
#

for ($i=0; $i<=$#framenr; $i++)
{
    $normB[$i] = sqrt($Bvx[$i]*$Bvx[$i] + $Bvy[$i]*$Bvy[$i] +
$Bvz[$i]*$Bvz[$i]);
    $normB[$i] = sprintf('%.4f',$normB[$i]);
    push(@normB, $normB[$i]);
};
#
for ($i=0; $i<=$#framenr; $i++)


{
  $vecprod = abs($vx[$i]*$Bvx[$i] + $vy[$i]*$Bvy[$i]+
$vz[$i]*$Bvz[$i]);
  $sina    = $vecprod / ($normA[$i]*$normB[$i]);
  $alpha[$i] = asin($sina)*180/pi;
  $alpha[$i] = sprintf('%.4f', $alpha[$i]);
  #print "$vecprod, $normA[$i], $normB[$i], $sina, $alpha[$i]\n";
  push(@alpha,$alpha[$i]);
}
# and print vector C  into file
#
open (INFO, ">>$C");
print INFO "# FORMAT: angle (deg) between vector a and plane defined
by normal vector b\n";
print INFO "# FORMAT: A: $A ; B: $B\n";
#
for ($i=0; $i<=$#framenr; $i++)
{
    print INFO "$framenr[$i] $alpha[$i]\n";
};
close (INFO);
};
#
```

6.34 MD simulations of *C. subterminale* LAM M124Q mutant.
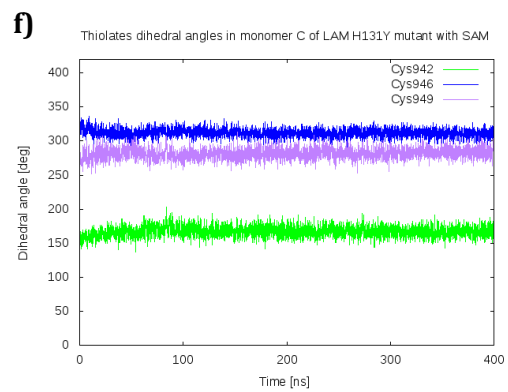
RMSD, RMSF and PCA of unbound-SAM (a, c) and bound-SAM (b, d)
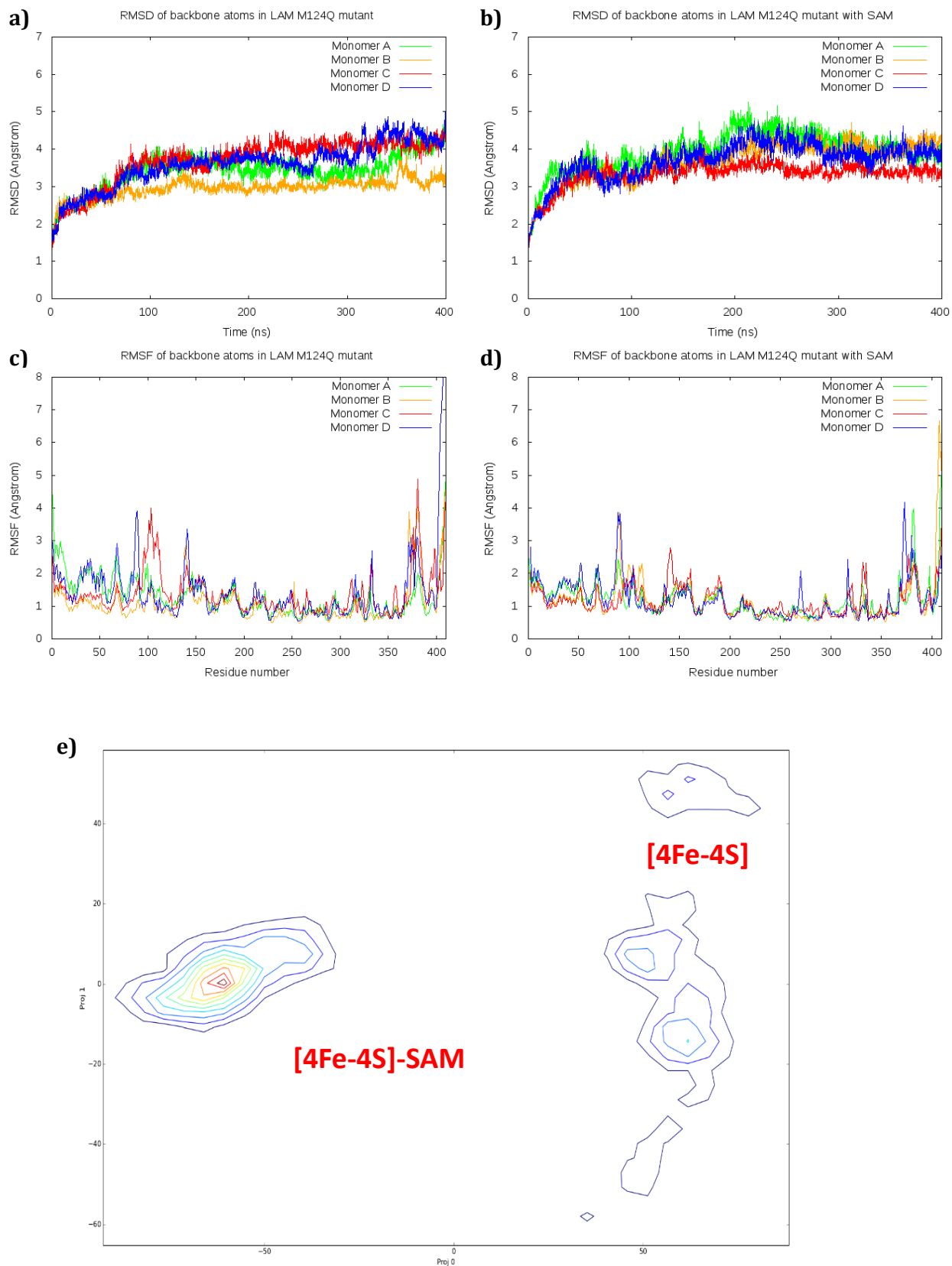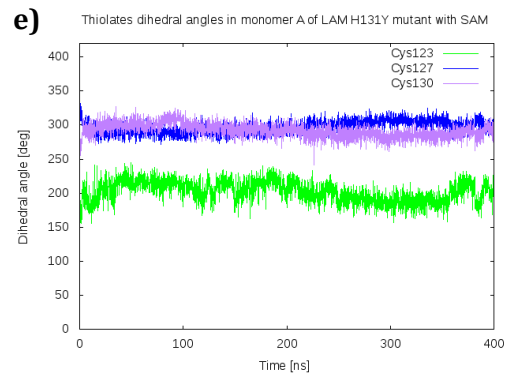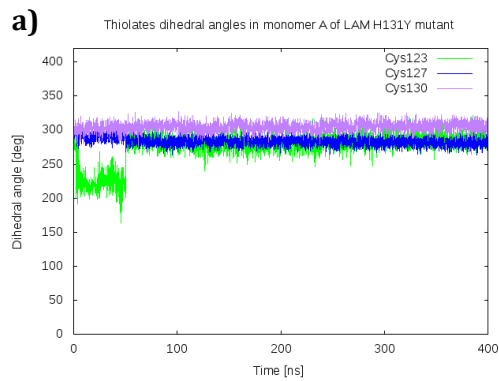
Retrieved SAM distances from the [Fe$_4$-S$_4$] cluster throughout the four monomers
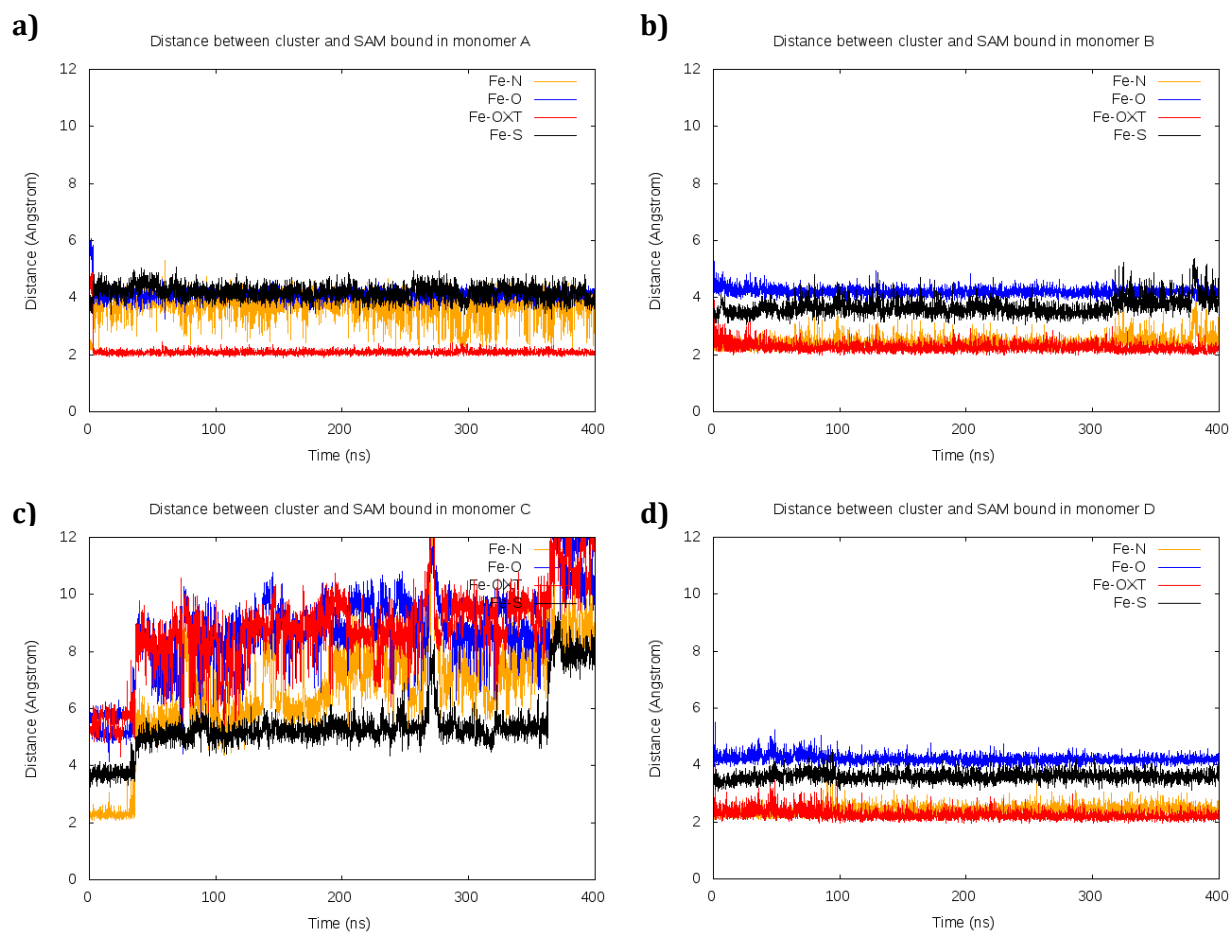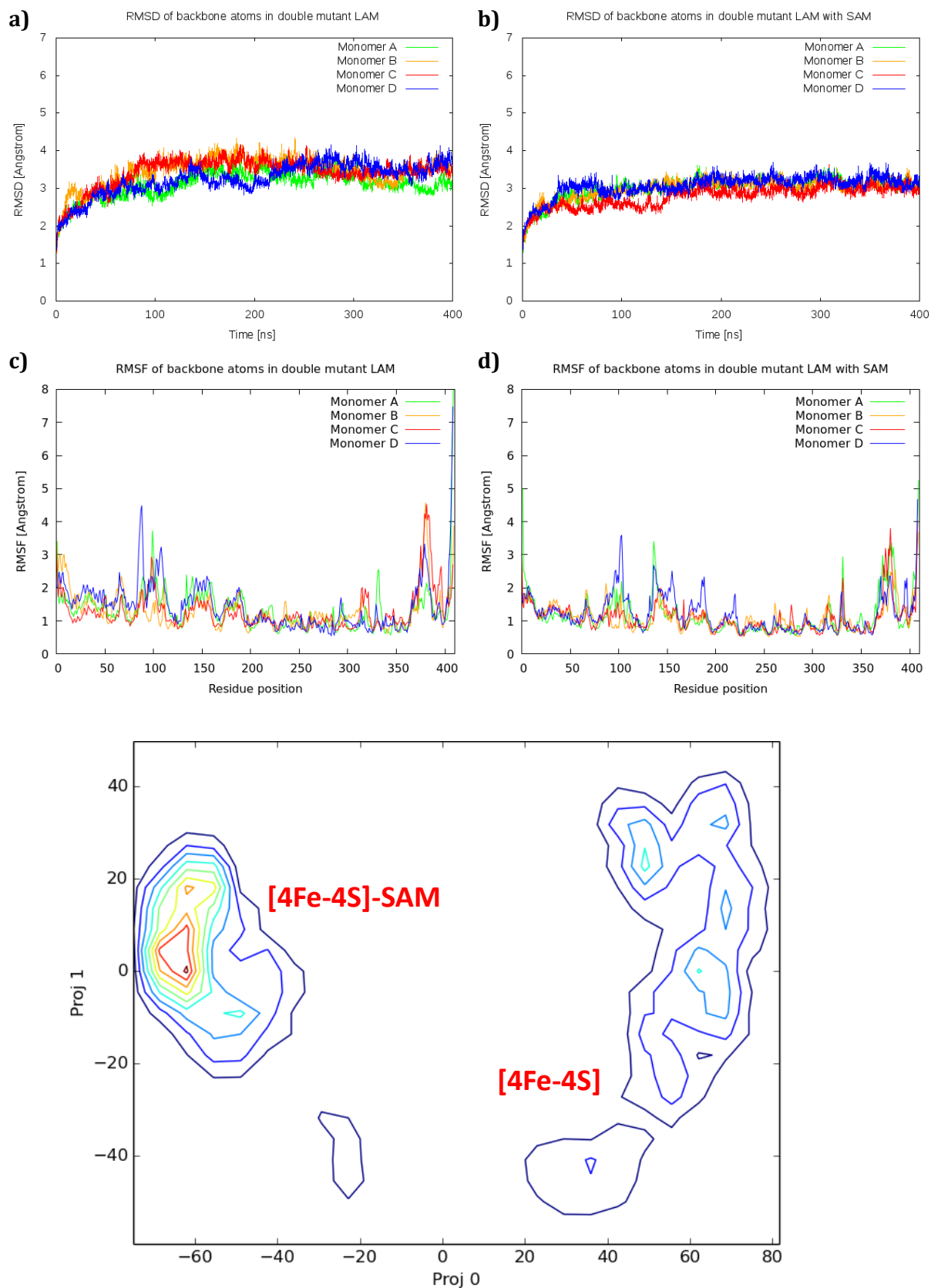

238

Dihedral angles of the three [Fe$_4$-S$_4$]-binding cysteine residues in M124Q mutant throughout the four monomers.

Unbound-SAM (a, b, c, d) and bound-SAM (e, f, g, h).
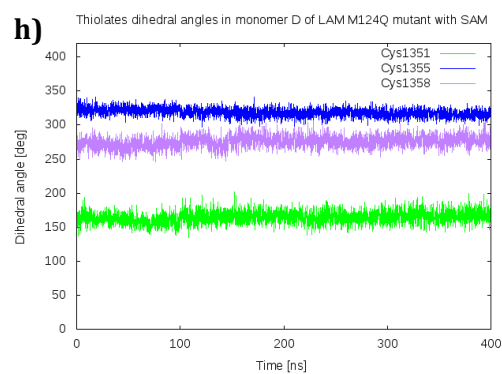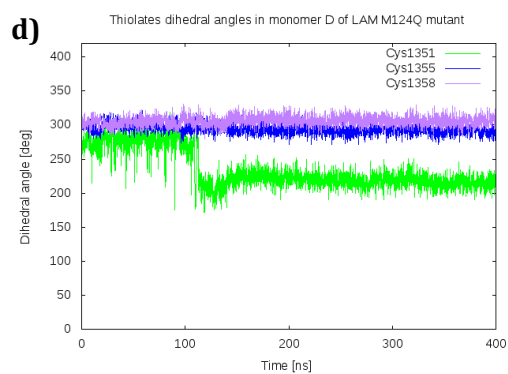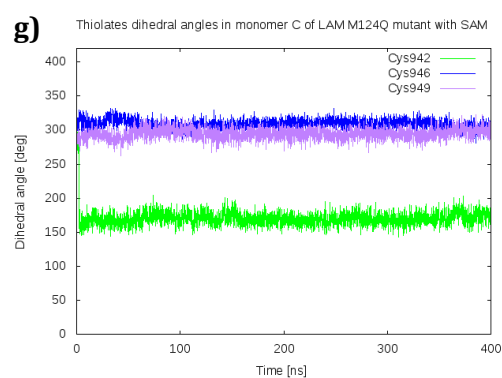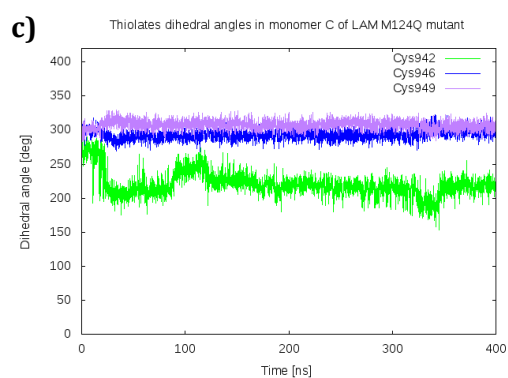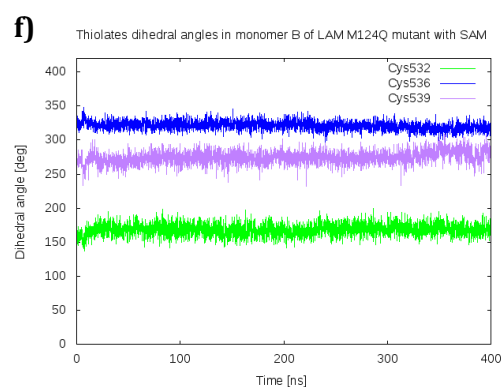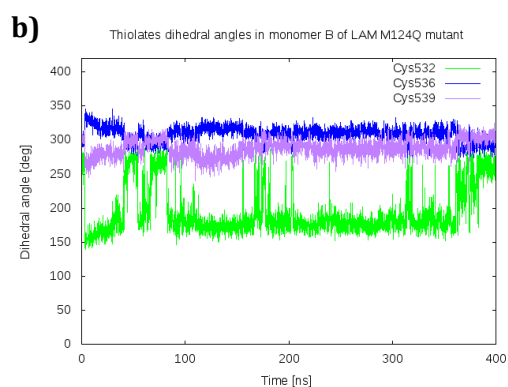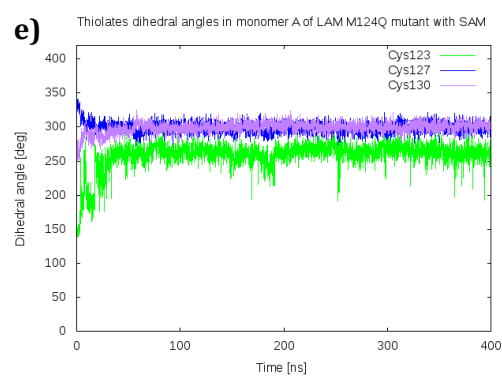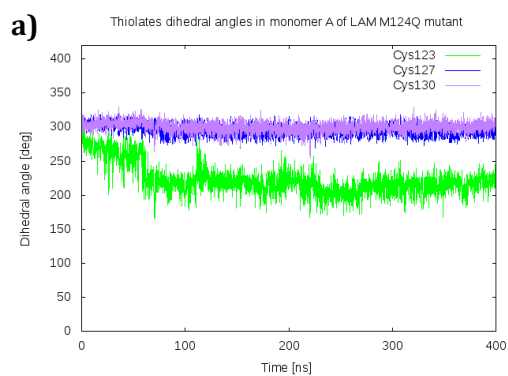
6.35   MD simulations of *C. subterminale* LAM H131Y mutant

RMSD, RMSF and PCA of unbound-SAM (a, c) and bound-SAM (b, d)
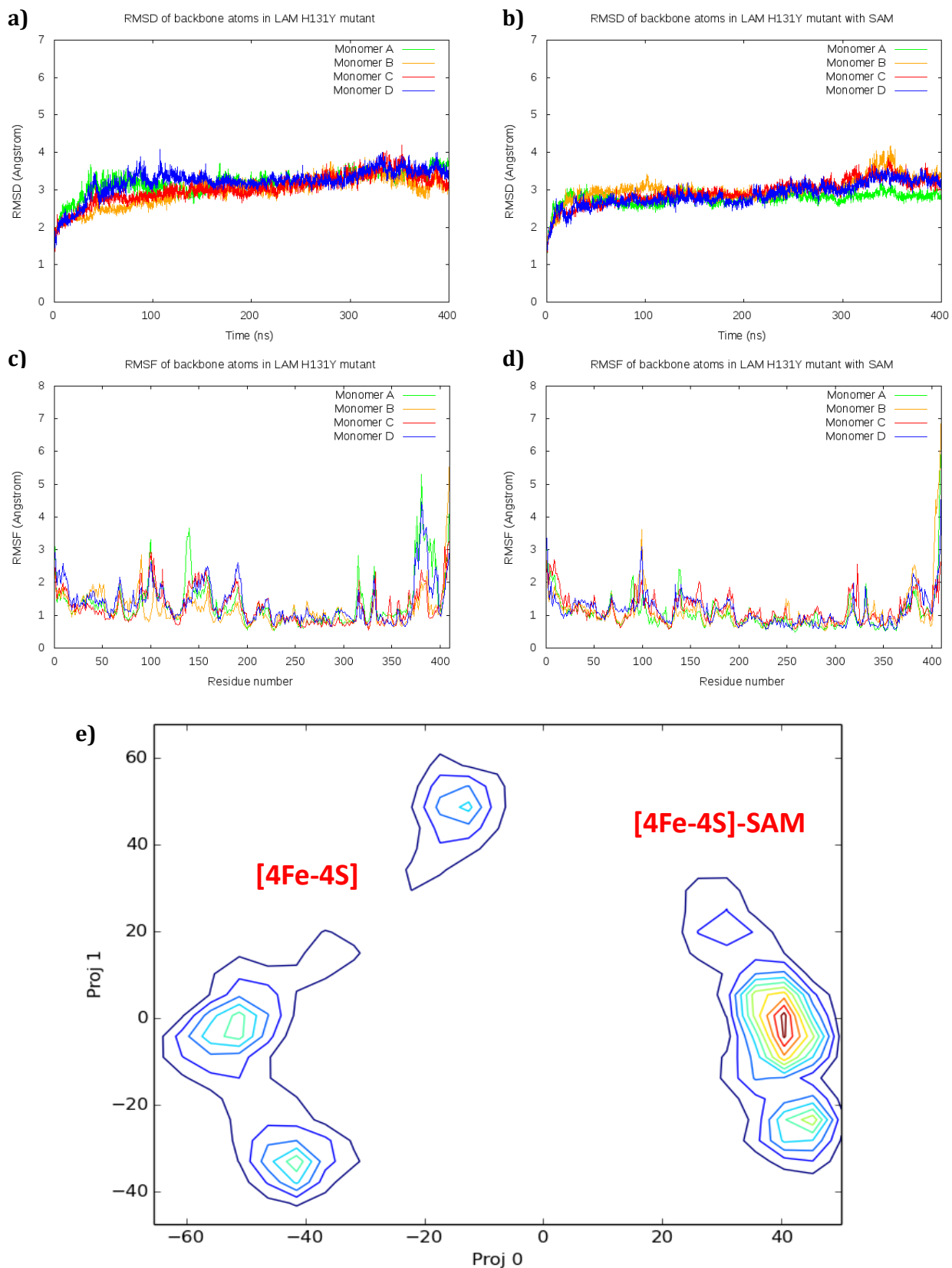
Retrieved SAM distances from [Fe$_4$-S$_4$] cluster throughout the four monomers

a)

b)

c)

d)
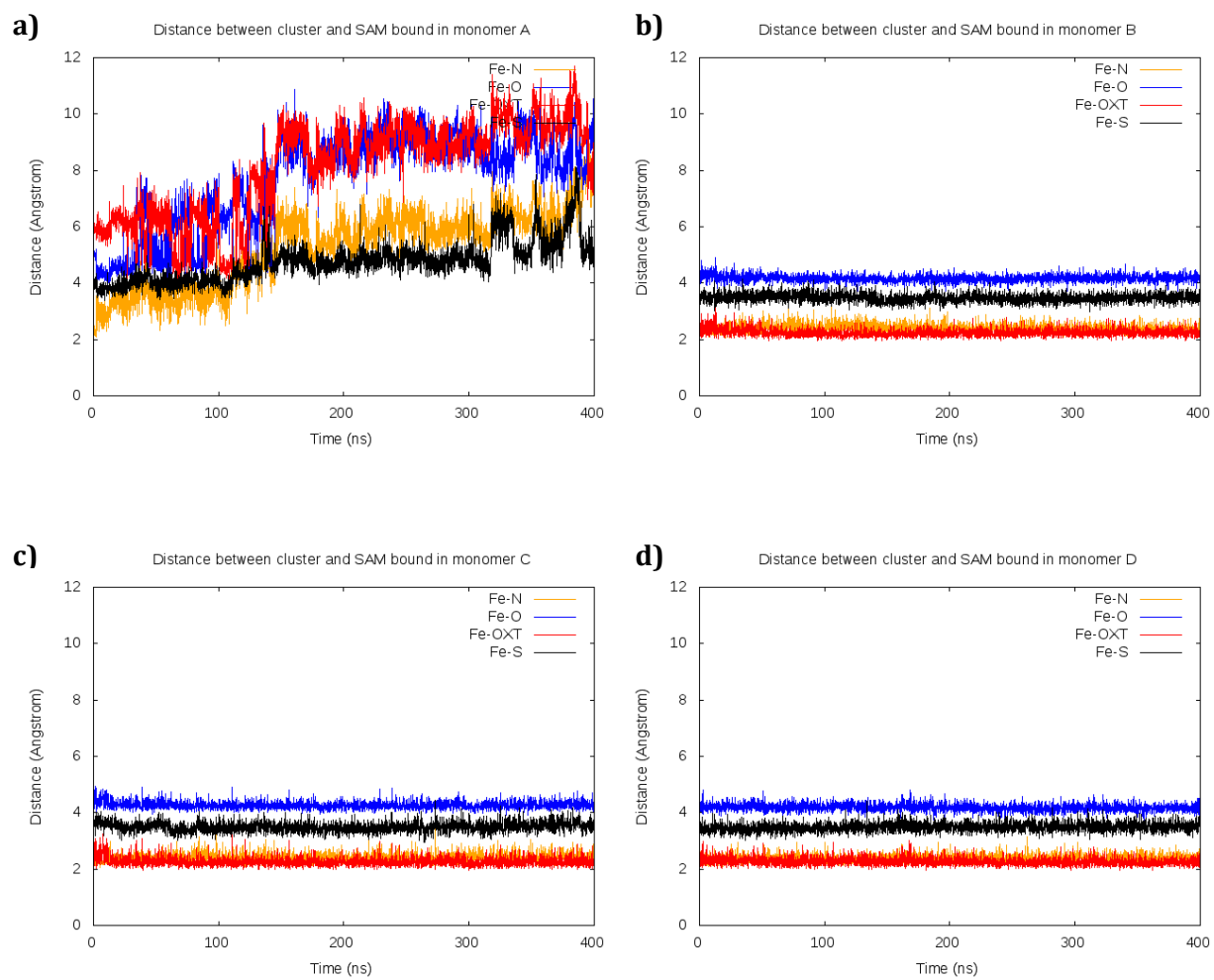
Dihedral angles of the three [Fe₄-S₄]-binding cysteine residues in M124Q mutant throughout the four monomers.

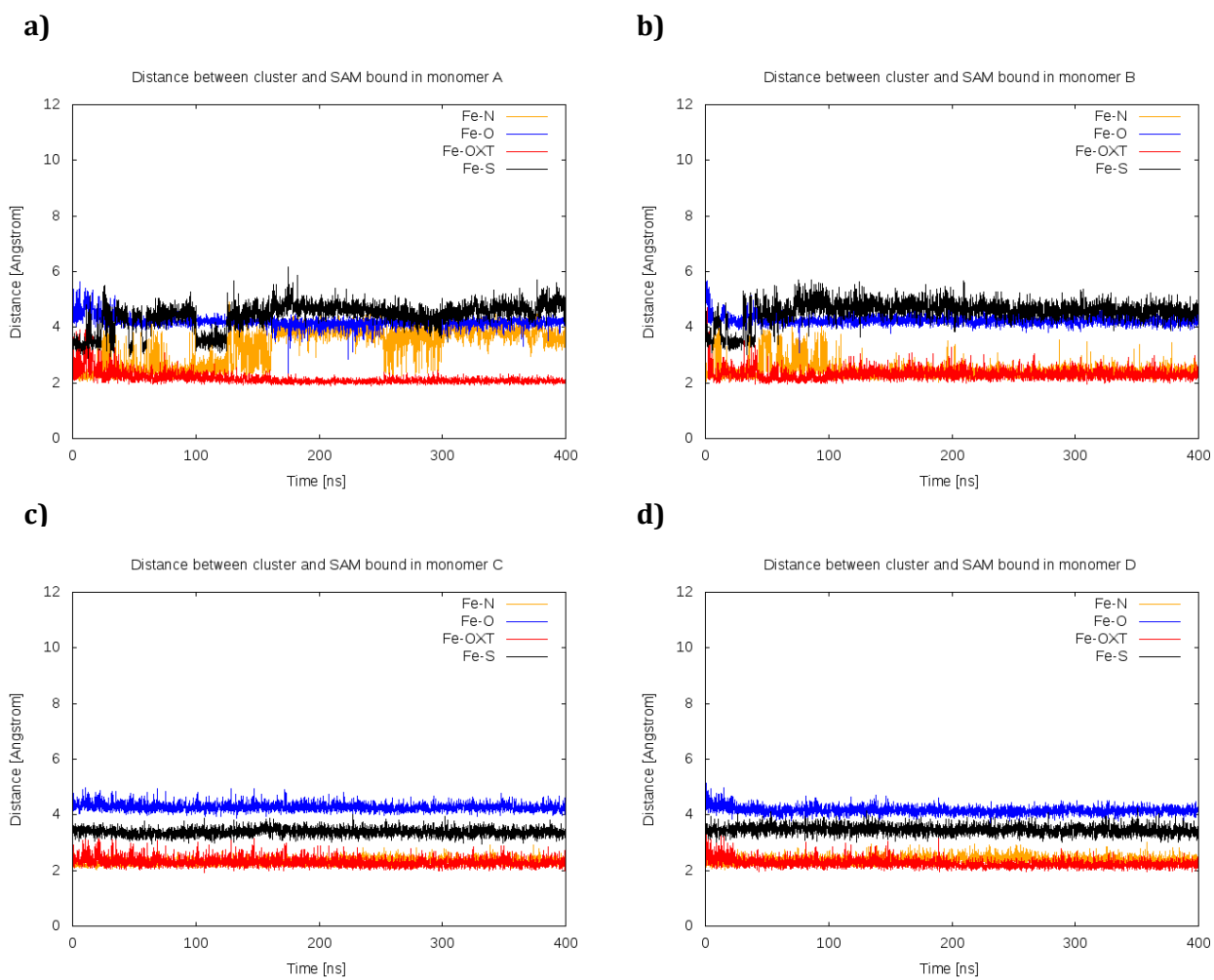Unbound-SAM (a, b, c, d) and bound-SAM (e, f, g, h).
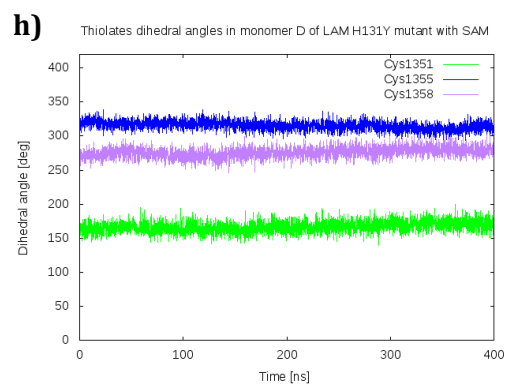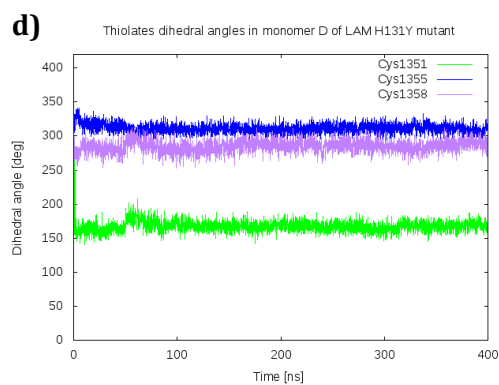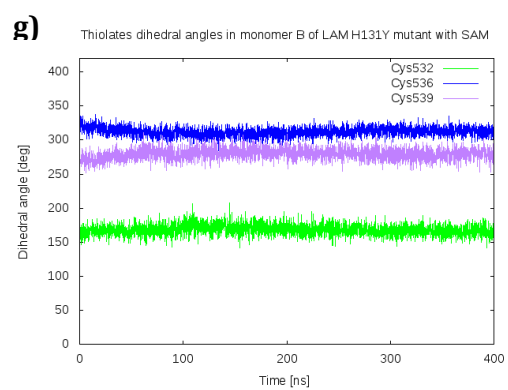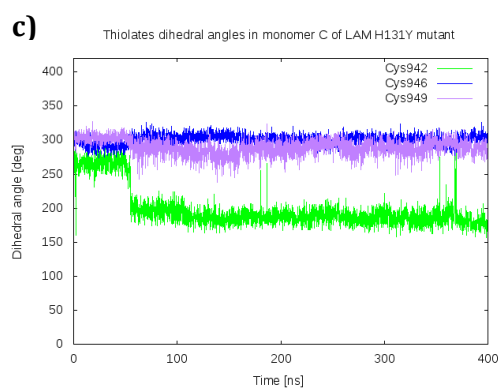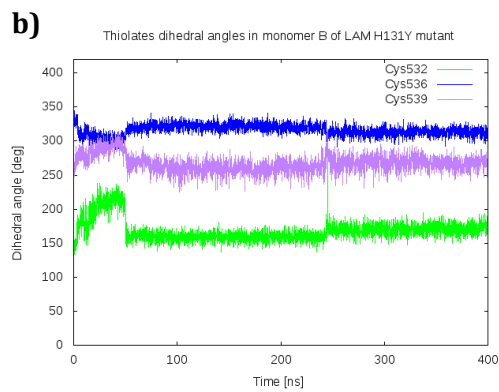
6.36   MD simulations of *C. subterminale* LAM M124Q/H131Y mutant.

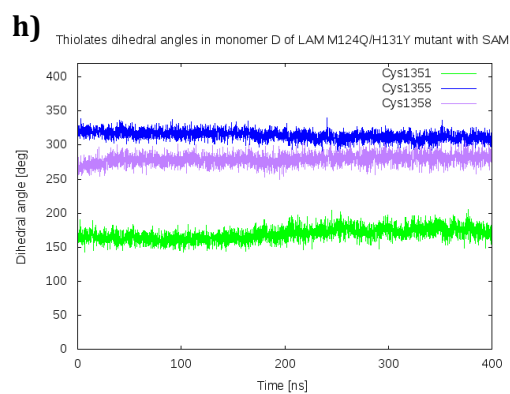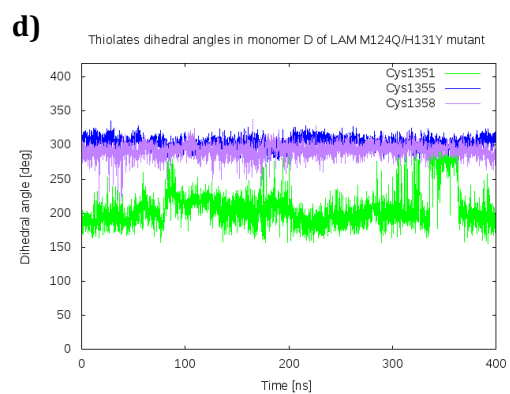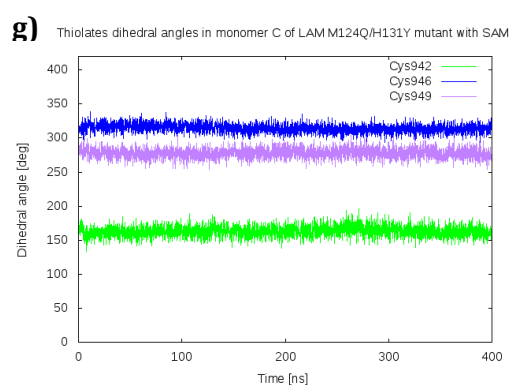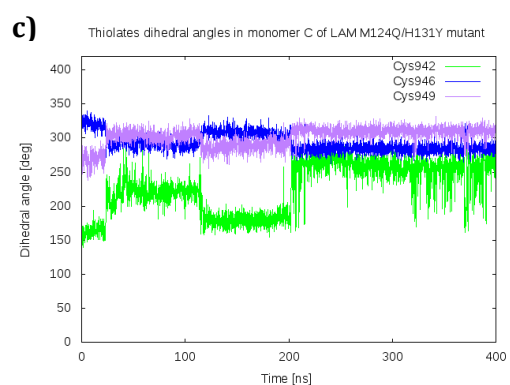RMSD, RMSF and PCA of unbound-SAM (a, c) and bound-SAM (b, d).

a)

RMSD of backbone atoms in double mutant LAM

b)

RMSD of backbone atoms in double mutant LAM with SAM

c)

RMSF of backbone atoms in double mutant LAM

d)

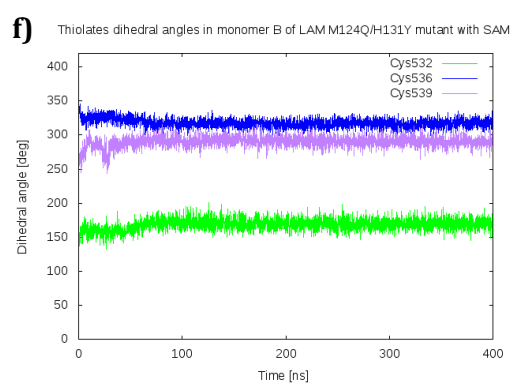RMSF of backbone atoms in double mutant LAM with SAM

[4Fe-4S]-SAM

[4Fe-4S]

Retrieved SAM distances from [Fe$_4$-S$_4$] cluster throughout the four monomers

a)

b)

c)

d)

Dihedral angles of the three [Fe$_4$-S$_4$]-binding cysteine residues in M124Q mutant throughout the four monomers.

Unbound-SAM (a, b, c, d) and bound-SAM (e, f, g, h).
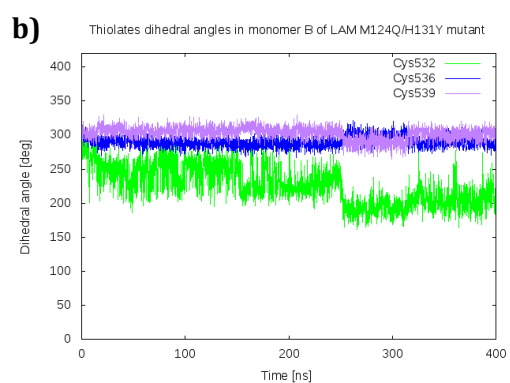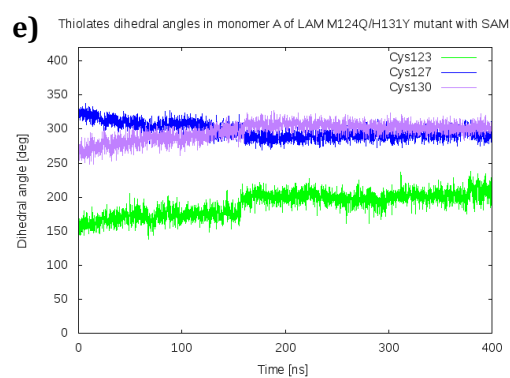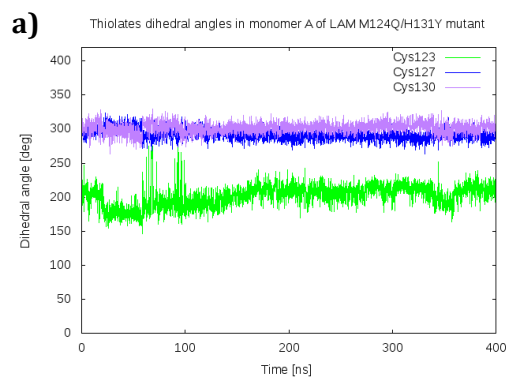
6.37 MD simulations of *C. subterminale* LAM H131Y/A138S mutant.

RMSD, RMSF and PCA of unbound-SAM (a, c) and bound-SAM (b, d).
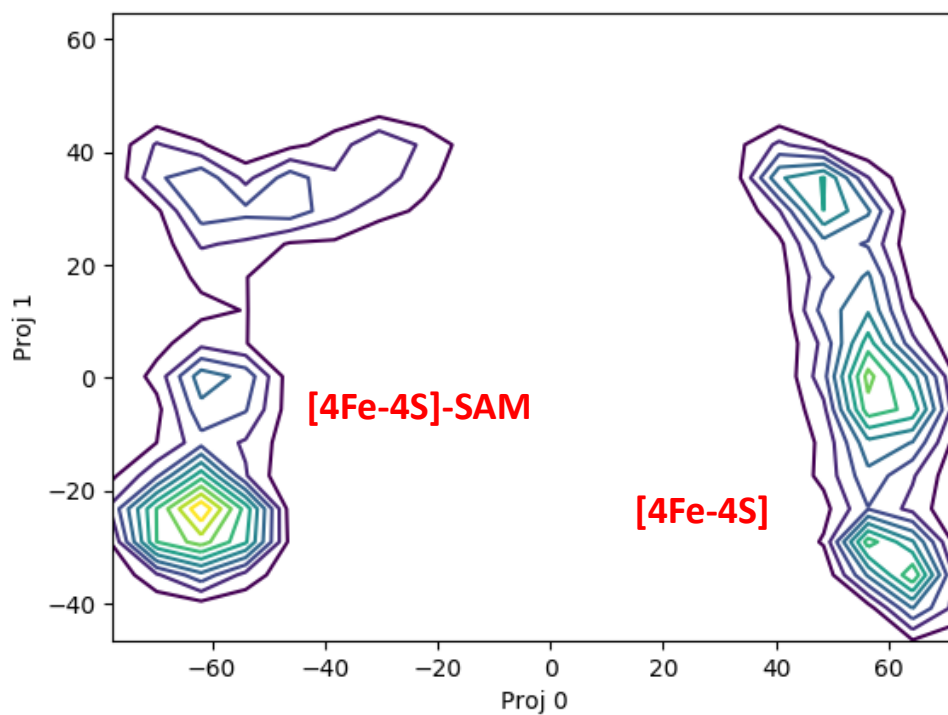
Retrieved SAM distances from [Fe_4-S_4] cluster throughout the four monomers

a)

Distance between cluster and SAM bound in monomer A

b)

Distance between cluster and SAM bound in monomer B

c)

Distance between cluster and SAM bound in monomer C

d)

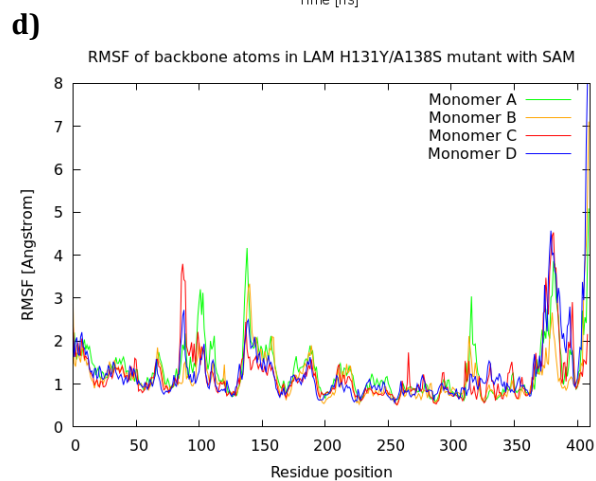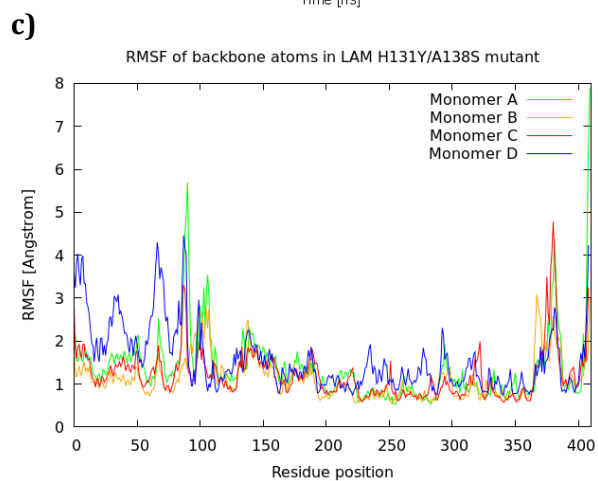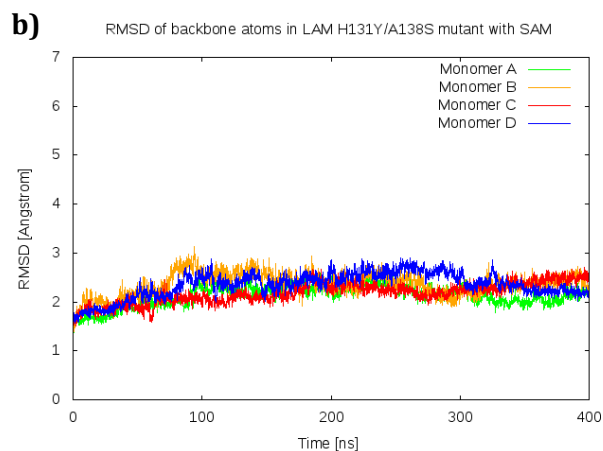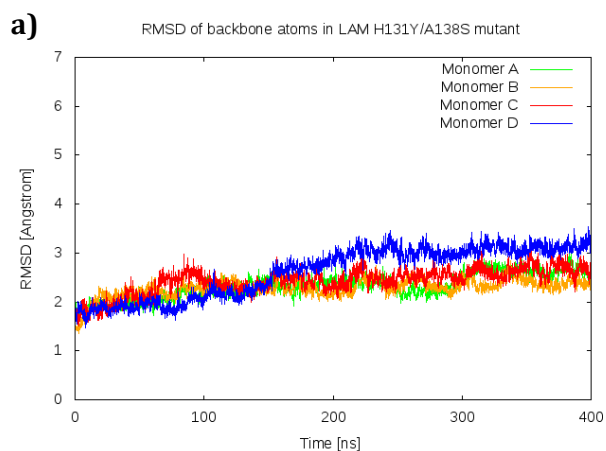Distance between cluster and SAM bound in monomer D

247

Dihedral angles of the three [Fe$_4$-S$_4$]-binding cysteine residues in M124Q mutant throughout the four monomers.
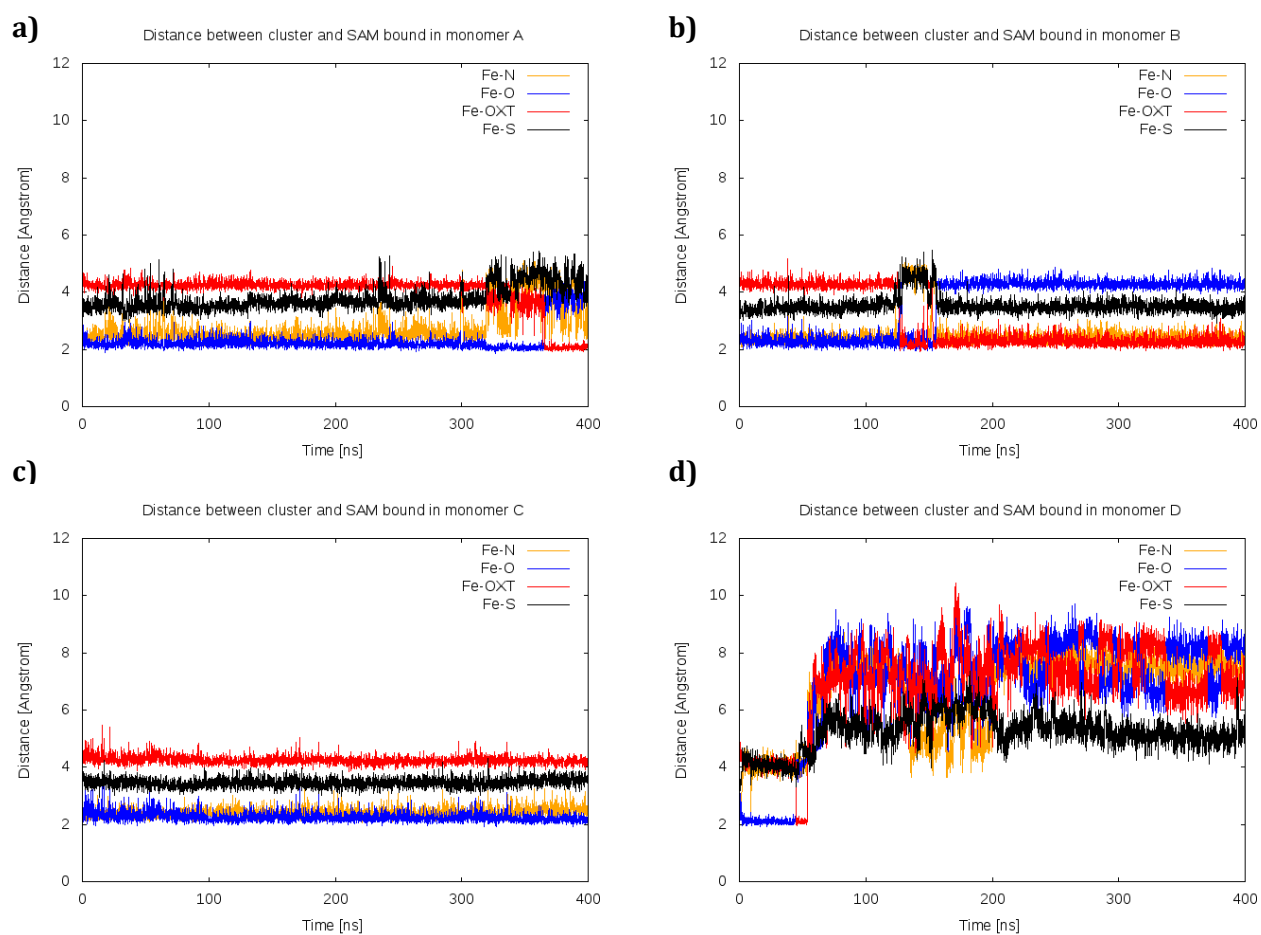
Unbound-SAM (a, b, c, d) and bound-SAM (e, f, g, h).
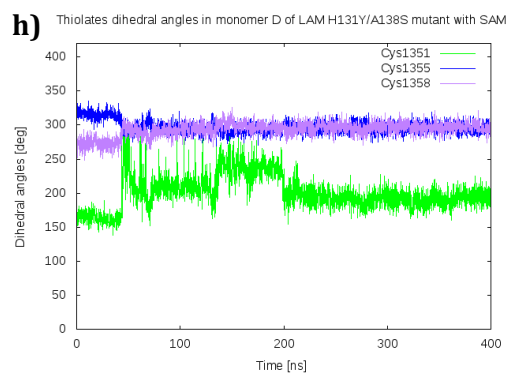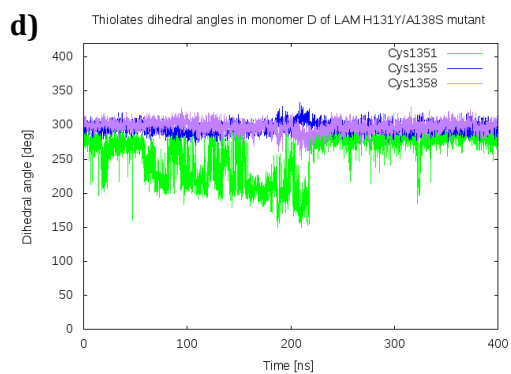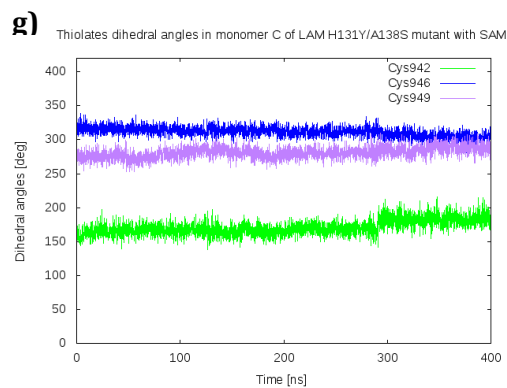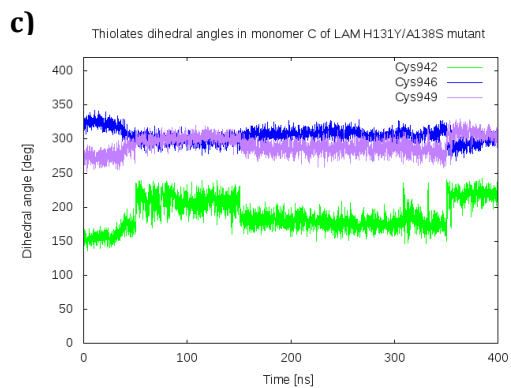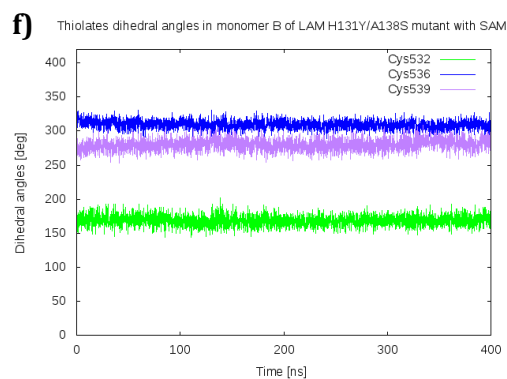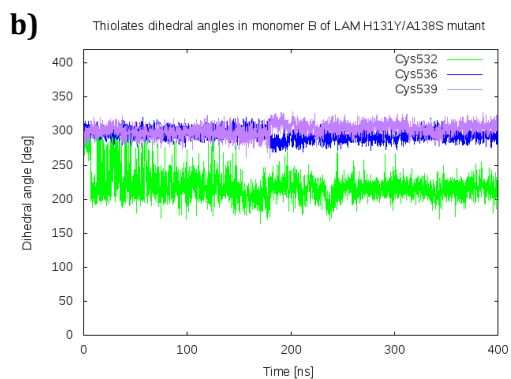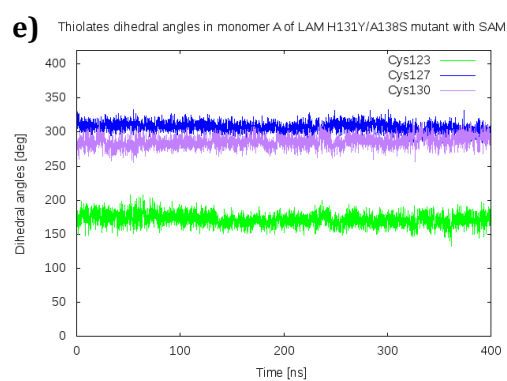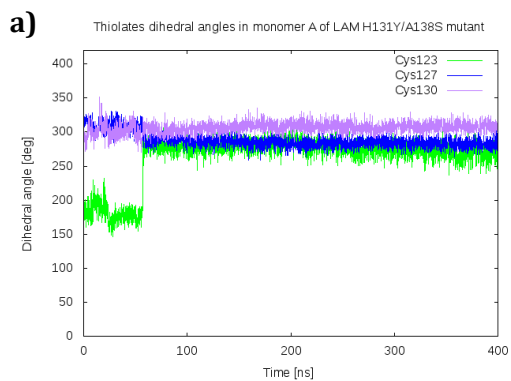
6.38   LAM sequence network analysed.

The following legend helps in the interpretation of the retrieved SSN to which the network was rendered by type of life.
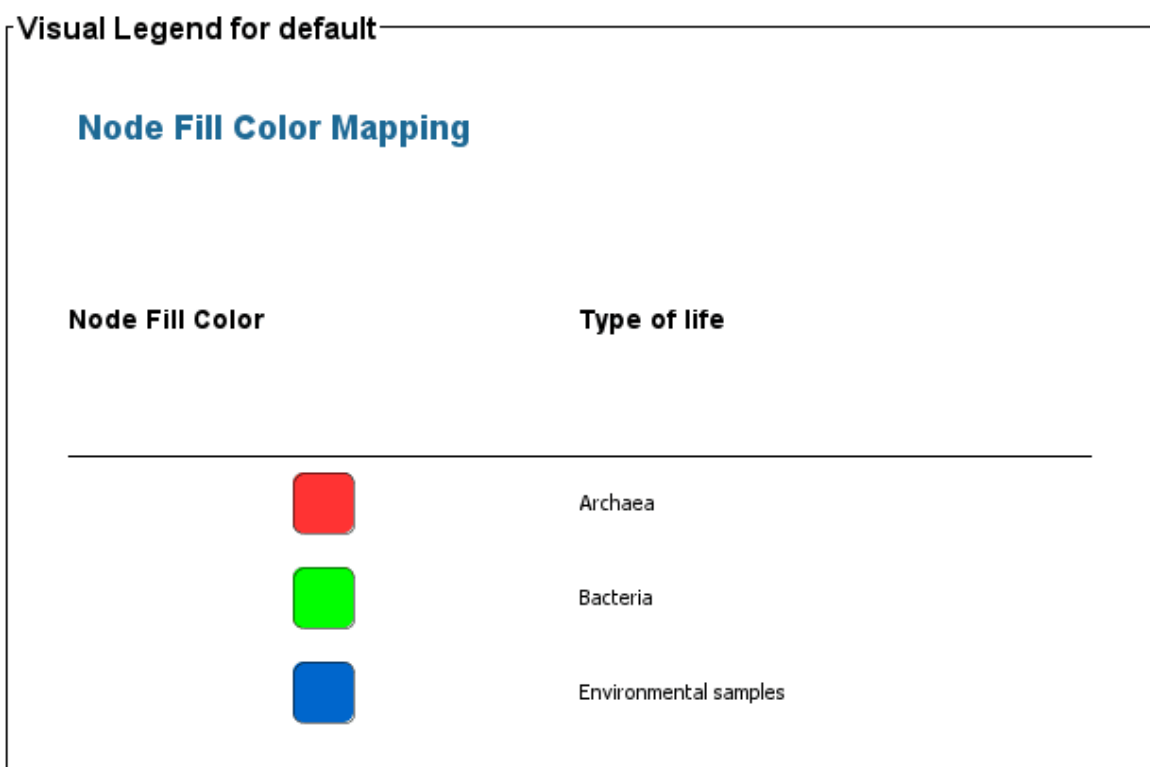
References

1. Kennedy, J.; Figueiredo, Z. Dordick, J. S., Ed. Plenum Press, New York: 1992,.

2. Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B., *Nature,* **2001,,** *409* (6817), 258.

3. Jäger, C. M.; Croft, A. K., *ChemBioEng Rev.;* **2018,** *5* (3), 143.

4. Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E., *Nuc. Acid. Res.,* **2001,** *29* (5), 1097.

5. Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M., *Chem. Rev.,* **2014,** *114* (8), 4229.

6. Chirpich, T. P.; Zappia, V.; Costilow, R. N.; Barker, H. A., *J. Biol. Chem.,* **1970,** *245*, 1778.

7. Gerlt, J. A., *Biochemistry,* **2017,** *56* (33), 4293.

8. Frey, P. A.; Hegeman, A. D.; Ruzicka, F. J., *Crit. Rev. Biochem. Mol. Biol.,* **2008,** *43* (1), 63.

9. Buckel, W.; Golding, B. T., *Annu. Rev. Microbiol.,* **2006,** *60*, 27.

10. Layer, G.; Kervio, E.; Morlock, G.; Heinz, D. W.; Jahn, D.; Retey, J.; Schubert, W. D., *Biol. Chem.,* **2005,** *386* (10), 971.

11. Esberg, B.; Leung, H. C.; Tsui, H. C.; Bjork, G. R.; Winkler, M. E., *J. Bacteriol.* **1999,** *181*, 7256.

12. Arragain, S.; Garcia-Serres, R.; Blondin, G.; Douki, T.; Clemancey, M.; Latour, J. M.; Forouhar, F.; Neely, H.; Montelione, G. T.; Hunt, J. F.; Mulliez, E.; Fontecave, M.; Atta, M., *J. Biol. Chem.,* **2010,** *285* (8), 5792.

13. Baker, J. J.; Stadtman, T. C., *B12: Biochemistry and Medicine,* **1982,** *2*, 203.

14. Walsby, C. J.; Ortillo, D. O.; Broderick, W. E.; Broderick, J. B.; Hoffman, B. M., *J. Am. Chem. Soc.,* **2002,** *124*, 11270.

15. Wang, J.; Woldring, R. P.; Roman-Melendez, G. D.; McClain, A. M.; Alzua, B. R.; Marsh, E. N., *ACS Chem. Biol.,* **2014,** *9* (9), 1929.

16. Duschene, K. S.; Veneziano, S. E.; Silver, S. C.; Broderick, J. B., *Curr. Opin. Chem. Biol.,* **2009,** *13* (1), 74.

17. Vey, J. L.; Drennan, C. L., *Chem. Rev.,* **2011,** *111* (4), 2487.

18. Jarrett, J. T.; Farrar, C. E., RadicalS-Adenosylmethionine (SAM) Superfamily,. *eLS. John Wiley &*

Son, **2014**.

19.   Sharma, P. K.; Chu, Z. T.; Olsson, M. H.; Warshel, A., *Proc. Natl. Acad. Sci. U. S. A.,* **2007,** *104*
(23), 9661.

20.   Vey, J. L.; Yang, J.; Li, M.; Broderick, W. E.; Broderick, J. B.; Drennan, C. L., *Proc. Natl. Acad. Sci.
U. S. A.,* **2008,** *105* (42), 16137.

21.   Farrar, C. E.; Jarrett, J. T., *Biochemistry,* **2009,** *48* ((11)), 2448.

22.   Frey, P. A., *Annu. Rev. Biochem.,* **2001,** *70*, 121.

23.   Frey, P. A.; Ballinger, M. D.; Reed, G. H., *Biochem. Soc. Trans.,* **1998,** *26*, 304.

24.   Frey, P. A., *Chem. Rev.,* **1990,** *90*, 1343.

25.   Frey, P. A., *FASEB J.,* **1993,** *7*, 662.

26.   Baraniak, J.; Moss, M. L.; Frey, P. A., *J. Biol. Chem.,* **1989,** *264*, 1357.

27.   Moss, M. L.; Frey, P. A., *J. Biol. Chem.,* **1987,** *262* (31), 14859.

28.   Walsby, C. J.; Ortillo, D. O.; Yang, J.; Nnyepi, M. R.; Broderick, W. E.; Hoffman, B. M.; Broderick,
J. B., *Inorg. Chem.,* **2005,** *44*, 727.

29.   Henshaw, T. F.; Cheek, J.; Broderick, J. B., *J. Am. Chem. Soc.,* **2000,** *122*, 8331.

30.   Fontecave, M.; Ollagnier-de-Choudens, S.; Mulliez, E., *Chem. Rev.,* **2003,** *103* (6), 2149.

31.   Booker, S. J.; Cicchillo, R. M.; Grove, T. L., *Curr. Opin. Chem. Biol.,* **2007,** *11* (5), 543.

32.   Fugate, C. J.; Jarrett, J. T., *Biochim. Biophys. Acta Proteins Proteom.,* **2012,** *1824* (11), 1213.

33.   Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B., *Nature,* **2001, ,** *409*
258.

34.   Zelder, O.; Jeong, W. K.; Klopprogge, C.; Herold, A.; Schroder, H., *US patent application,* **2009,**
us 2009/0029425.

35.   Frey, P. A.; Ruzicka, F. J., *Wisconsin Alumni Research Foundation,* **2008**, US patent 7,456,271
B2.

36.   Grammel, N.; Pankevich, K.; Demydchuk, J.; Lambrecht, K.; Saluz, H.-P.; Keller, U.; Krugel, H.,
Eur. J. Biochem., **2002,** *269*, 347.

37.   Carter, J. H.; Du Bus, R. H.; Dyer, J. R.; Floyd, J. C.; Rice, K. C.; Shaw, P. D., *Biochemistry,* **1974,** *13*
(6), 1227.

38.   Thiruvengadam, T. K.; Gould, S. J.; Aberhart, D. J.; Lin, H. J., *Journal of the American Chemical
Society,* **1983,** *105* (16), 5470.

39.   Barrett, G. C., Chemistry and biochemistry of the amino acids. *Biochemical Education* **1985,** *16* (2), 118.

40.   Smith, M., Methods of Non-a-Amino Acid Synthesis. **1995,** (Boca Raton: CRC Press).

41.   Behshad, E.; Ruzicka, F. J.; Mansoorabadi, S. O.; Chen, D.; Reed, G. H.; Frey, P. A., *Biochemistry,* **2006,** *45* (42), 12639.

42.   Chatterjee, R. e. a., *US patent application* **2010**, No. US2010/0099143A1.

43.   Liao, H. H.; Gokarn, R. R.; Gort, S. J.; Jessen, H. J.; Selifonova, O., *Novozymes A/S,* **2010**, US patent 7,655,451 B2.

44.   Jang, S.; Imlay, J. A., *J. Biol. Chem.,* **2007,** *282* (2), 929.

45.   Imlay, J. A., *Mol. Microbiol.,* **2006,** *59* (4), 1073.

46.   Crack, J. C.; Green, J.; Cheesman, M. R.; Le Brun, N. E.; Thomson, A. J., *Proc. Natl. Acad. Sci. U. S. A.,* **2007,** *104* (7), 2092.

47.   Flint, D. H.; Tuminello, J. F.; Emptage, M. H., *J. Biol. Chem.,* **1993,** *268*, 22369.

48.   Kuo, C. F.; Mashino, T.; Fridovich, I., *J. Biol. Chem.,* **1987,** *262*, 4724.

49.   Chen, G.; Ruzicka, F. J.; Frey, P. A., *Biochem. J.,* **2000,** *348*, 539.

50.   Saichana, N.; Tanizawa, K.; Pechousek, J.; Novak, P.; Yakushi, T.; Toyama, H.; Frebortova, J., *J. Biochem.,* **2016,** *159* (1), 87.

51.   Barr, I.; Stich, T. A.; Gizzi, A. S.; Grove, T. L.; Bonanno, J. B.; Latham, J. A.; Chung, T.; Wilmot, C. M.; Britt, R. D.; Almo, S. C.; Klinman, J. P., *Biochemistry,* **2018,** *57* (8), 1306.

52.   Yang, J.; Naik, S. G.; Ortillo, D. O.; Garcia-Serres, R.; Li, M.; Broderick, W. E.; Huynh, B. H.; Broderick, J. B., *Biochemistry,* **2009,** *48* (39), 9234.

53.   Berkovitch, F.; Nicolet, Y.; Wan, J. T.; Jarrett, J. T.; Drennan, C. L., *Science* **2004,** *303* (5654), 76.

54.   Wolthers, K. R.; Levy, C.; Scrutton, N. S.; Leys, D., *J. Biol. Chem.,* **2010,** *285* (18), 13942-50.

55.   Overton, T.; Reid, E. G. F.; Foxall, R.; Smith, H.; Busby, S. J. W.; Cole, J. A., *Journal of Bacteriology,* **2003,** *185* (16), 4734.

56.   Singh, B. B.; Curdt, I.; Jakobs, C.; Shomburg, D.; Bisen, P. S.; Bhöme, H., *Biochimica et Biophysica Acta - Bioenergetics,* **1999,** *1412* (3), 288.

57.   Dementin, S.; Leroux, F.; Cournac, L.; de Lacey, A. L.; Volbeda, A. L., C.; ; Burlat, B.; Martinez, N.; Champ, S.; Martin, L.; Sanganas, O.; Haumann, M.; Fernandez, V. M.; Guigliarelli, B.; Fontecilla-Camps, J.; Rousset, M., ;, *J. Am. Chem. Soc.,* **2009,** *131*, 10156.

58. Jervis, A. J.; Crack, J. C.; White, G.; Artymiuk, P. J.; Cheesman, M. R.; Thomson, A. J.; Le Brun, N. E.; Green, J., *Proc. Natl. Acad. Sci. U. S. A.,* **2009,** *106* (12), 4659.

59. Singh., B. B.; Curdt, I.; Shomburg, D.; Bisen, P. S.; Bhöme, H., *Mol. Cell. Biochem.,* **2001,** *217,* 137.

60. Leroux, F.; Dementin, S.; Burlat, B.; Cournac, L.; Volbeda, A.; Champ, S.; Martin, L.; Guigliarelli, B.; Bertrand, P.; Fontecilla-Camps, J.; Rousset, M.; Leger, C., *Proc. Natl. Acad. Sci. U. S. A.,* **2008,** *105* (32), 11188.

61. Volbeda, A.; Montet, Y.; Vernede, X.; Hatchikian, C. E.; Fontecilla-Camps, J., *Int. J. Hydrogen Energy,* **2002,** *27,* 1449.

62. Buhrke, T.; Lenz, O.; Krauss, N.; Friedrich, B., *J. Biol. Chem.,* **2005,** *280* (25), 23791.

63. Stadtman, E. R.; Moskovitz, J.; Berlett, B. S.; Levine, R. L., *Mol. Cell. Biochem.,* **2002,** *234,* 4.

64. Pal, D.; Chakrabarti, P., *J. Biomol. Struct. Dyn.,* **2001,** *19* (1), 115.

65. Stadtman, E. R.; Van Remmen, H.; Richardson, A.; Wehr, N. B.; Levine, R. L., *Biochim. Biophys. Acta,* **2005,** *1703* (2), 135.

66. Tilley, G. J.; Camba, R.; Burgess, B. K.; Armstrong, F. A., *Biochem J.,* **2001,** *15* (360), 717.

67. Chen, K.; Tilley, G. J.; Sridhar, V.; Prasad, G. S.; Stout, C. D.; Armstrong, F. A.; Burgess, B. K., *J. Biol. Chem.,* **1999,** *274,* 36479.

68. Jung, Y. S.; Bonagura, C. A.; Tilley, G. J.; Gao-Sheridan, H. S.; Armstrong, F. A.; Stout, C. D.; Burgess, B. K., *J. Biol. Chem.,* **2000,** *275* (47), 36974.

69. Finnegan, M. G.; Conover, R. C.; Park, J.-B.; Zhou, Z. H.; Adams, M. W. W.; Johnson, M., *Inorg. Chem.,* **1995,** *34,* 5358.

70. Boll, M.; Fuchs, G.; Tilley, G. J.; Armstrong, F. A.; Lowe, D. J., *Biochemistry,* **2000,** *39,* 4929.

71. Camba, R.; Armstrong, F. A., *Biochemistry,* **2000,** *39,* 10587.

72. Carter, C. W.; Kraut, J.; Freer, S. T.; Alden, R. A.; Sieker, L. C.; Adman, E.; Jensen, L. H., *Proc. Nat. Acad. Sci. U. S. A.,*
 1972, *69* (12), 3526.

73. Cammack, R., In *Advances in Inorganic Chemistry,*, Academic Press: 1992; Vol. 38, p 281.

74. Hosseinzadeh, P.; Lu, Y., *Biochim. Biophys. Acta,* **2016,** *1857* (5), 557.

75. Sheridan, R. P.; Leland, C. A.; Carter, C. W., *j. Biol. Chem.,* **1981,** *256* (10), 5052.

76. Torres, R. A.; Lovell, T.; Noodleman, L.; Case, D. A., *J. Am. Chem. Soc.,* **2003,** *125,* 1923.

77.   Carvalho, A. T.; Swart, M., *J. Chem. Inf. Model.,* **2014,** *54,* (2), 613.

78.   Stephens, P. J.; Jollie, D. R.; Warshel, A., *Chem. Rev.,* **1996,** *96*, 2491.

79.   Swartz, P. D.; Beck, B. W.; Ichiye, T., *Biophys. J.,* **1996,** *71*, 2958.

80.   Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P.; Sprik, M., *J. Phys. Chem. B,* **2007,** *111*, 3969.

81.   Denke, E.; Merbitz-Zahradnik, T.; Hatzfeld, O. M.; Snyder, C. H.; Link, T. A.; Trumpower, B. L., *J. Biol. Chem.,* **1998,** *273*, 9085.

82.   Schuetz, M.; Zirngibl, S.; le Coutre, J.; Buettner, M.; Xie, D.-L.; Nelson, N.; Deutzmann, R.; Hauska, G., *Photosynth. Res.,* **1994,** *39*, 163.

83.   Riedel, A.; Kellner, E.; Grodzitzki, D.; Liebl, U.; Hauska, G.; Mueller, A.; Rutherford, A. W.; Nitschke, W., *Biochim. Biophys. Acta,* **1993,** *1183* (2), 263.

84.   Barker, J. J.; Stadman, T. C., *John Wiley & Sons, New York,* **1982,** *2*, 203.

85.   Chen, D.; Frey, P. A., *Biochemistry,* **2001,** *40*, 596.

86.   Hanzelmann, P.; Schindelin, H., *Proc. Natl. Acad. Sci. U. S. A.,* **2006,** *103* (18), 6829.

87.   Brazeau, B. J.; Gort, S. J.; Jessen, H. J.; Andrew, A. J.; Liao, H. H., *Appl. Environ. Microbiol.,* **2006,** *72* (9), 6402.

88.   Layer, G.; Moser, J.; Heinz, D. W.; Jahn, D.; Schubert, W. D., *EMBO J.,* **2003,** *22* (23), 6214.

89.   Ballinger, M. D.; Reed, G. H.; Frey, P. A., *Biochemistry,* **1992,** *31* (4), 949.

90.   Aberhart, D. J.; Lin, H. J.; Weiller, B. H., *J. Am. Chem. Soc.,* **1981,** *103* (22), 6750.

91.   Chen, D.; Frey, P. A.; Lepore, B. W.; Ringe, D.; Ruzicka, F. J., *Biochemistry,* **2006,,** *45*, 12647.

92.   Lepore, B. W.; Ruzicka, F. J.; Frey, P. A.; Ringe, D., *Proc. Natl. Acad. Sci. U. S. A.,* **2005,** *102* (39), 13819.

93.   Hinckley, G. T.; Frey, P. A., *Biochemistry,* **2006,** *45* (10), 3219.

94.   Wang, S. C.; Frey, P. A., *Biochemistry,* **2007,** *46* (45), 12889.

95.   Nicolet, Y.; Amara, P.; Mouesca, J.-M.; Fontecilla-Camps, J. C., *Proc. Nat. Acad. Sci. U. S. A.,* **2009,** *106* (35), 14867.

96.   Lieder, K.; Booker, S.; Ruzicka, F. J.; Beinhert, H.; Reed, G. H.; Frey, P. A., *Biochemistry,* **1998,** *37*, 2578.

97.   Hung, C. C.; Lai, M. C., *J. Microbiol. Immunol. Infect.,* **2013,** *46* (1), 1.

98.   Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C., *Nuc. Acids Res.,* **2003,** *31,* (13), 3381.

99. McCammon, J. A.; Gelin, B. R.; Karplus, M., *Nature,* **1977,** *267* (5612), 585.

100. Alder, B. J.; Wainwright, T. E., *J. Chem. Phys.,* **1957,** *27* (5), 1208.

101. Stillinger, F. H.; Rahman, A., *J. Chem. Phys.,* **1974,** *60* (4), 1545.

102. Karplus, M., *Acc. Chem. Res.,* **2002,** *35* (6), 321.

103. Ermer, O., In *Bonding forces. Structure and Bonding*, Springer, B., Heidelberg, Ed. 1976; Vol. 27.

104. Hagler, A. T., In *Conformation in Biology and Drug Design*, Hruby, V. J., Ed. Academic Press: 1985; Vol. 7, p 213.

105. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., *The Journal of Physical Chemistry B,* **1998,** *102* (18), 3586.

106. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., *J. Am. Chem. Soc.,* **1995,** *117* (19), 5179.

107. Hermans, J.; Berendsen, H. J. C.; Van Gunsteren, W. F.; Postma, J. P. M., *Biopolymers,* **1984,** *23* (8), 1513.

108. van Gunsteren, W. F., Biomos ; Zürich: 1996.

109. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.,* **1983,** *79* (2), 926.

110. Berman, H. M.; Westbrook, J.; Z., F.; G., G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic. Acids Res.,* **2000,** *28* (1), 235.

111. Marc, A.; Martí-Renom, A. C.; Stuart, A.; Fiser, R.; Sánchez, F. M.; Šali, A., *Annu. Rev. Biophys. Biomol. Struct.,* **2000,** *29*, 291.

112. Chothia, C.; Lesk, A. M., *EMBO J.,* **1986,** *5* (4), 823.

113. Rost B., *Protein Eng.,* **1999,** *12* (2), 85.

114. Madden, T., *The NCBI Handbook,* **2002**, 13.

115. Raman, S.; Vernon, R.; Thompson, J.; Tyka, M.; Sadreyev, R.; Pei, J.; Kim, D.; Kellogg, E.; DiMaio, F.; Lange, O.; Kinch, L.; Sheffler, W.; Kim, B.-H.; Das, R.; Grishin, N. V.; Baker, D., *Proteins,* **2009,** *77* (9), 89.

116.   Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y., *Nat. Methods.,* **2015,** *12* (1), 7.

117.   Marco Biasini, S. B., Andrew Waterhouse, Konstantin Arnold, Gabriel Studer, Tobias Schmidt, Florian Kiefer, Tiziano Gallo Cassarino, Martino Bertoni, Lorenza Bordoli, Torsten Schwede. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research; (1 July 2014) 42 (W1): W252-W258; doi: 10.1093/nar/gku340. Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37, D387-D392. Guex, N., Peitsch, M.C., Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis, 30(S1), S162-S173.

118.   Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V., *J. Mol. Biol.,* **1963,** *7* (1), 95.

119.   Karplus, P. A., *Prot. Sci.,* **1996,** *5*, 1406.

120.   Baruch-Shpigler Y.; Wang H.; Tuvi-Arad I.; Avnir D., *Biochemistry,* **2017,** *56* (42), 5635.

121.   Smart O. S.; Neduvelil J. G.; Wang X.; Wallace B. A.; Sansom M. S., *J. Mol. Graph.,* **1996,** *14* (6), 354.

122.   Sehnal, D.; Svobodová V. R.; Berka, K.; Pravda, L.; Navrátilová, V.; Banáš, P.; Ionescu, C.-M.; Otyepka, M.; Koča, J., *Journal of Cheminformatics,* **2013,** *5* (1), 39.

123.   Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J., *PLoS Comput. Biol.,* **2012,** *8* (10), e1002708.

124.   Yaffe, E.; Fishelovitch, D.; Wolfson, H. J.; Halperin, D.; Nussinov, R., *Nucleic Acids Res.,* **2008,** *36* (Web Server issue), W210-5.

125.   Simons, J., An Introduction To Theoretical Chemistry,. *Cambridge, UK: Cambridge University Press,* **2003,** *Chapter 6.*

126.   Slater, J. C., *Phys. Rev.,* **1928,** *32* (3), 339.

127.   M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria,; M. A. Robb, J. R. C., Montgomery, Jr., J. A., T. Vreven,; K. N. Kudin, J. C. B., J. M. Millam, S. S. Iyengar, J. Tomasi,; V. Barone, B. M., M. Cossi, G. Scalmani, N. Rega,; G. A. Petersson, H. N., M. Hada, M. Ehara, K. Toyota,; R. Fukuda, J. H., M. Ishida, T. Nakajima, Y. Honda, O. Kitao,; H. Nakai, M. K., X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross,; V. Bakken, C.

A., J. Jaramillo, R. Gomperts, R. E. Stratmann,; O. Yazyev, A. J. A., R. Cammi, C. Pomelli, J. W. Ochterski,; P. Y. Ayala, K. M., G. A. Voth, P. Salvador, J. J. Dannenberg,; V. G. Zakrzewski, S. D., A. D. Daniels, M. C. Strain,; O. Farkas, D. K. M., A. D. Rabuck, K. Raghavachari,; J. B. Foresman, J. V. O., Q. Cui, A. G. Baboul, S. Clifford,; J. Cioslowski, B. B. S., G. Liu, A. Liashenko, P. Piskorz,; I. Komaromi, R. L. M., D. J. Fox, T. Keith, M. A. Al-Laham,; C. Y. Peng, A. N., M. Challacombe, P. M. W. Gill,; B. Johnson, W. C., M. W. Wong, C. Gonzalez, J. A. Pople, *Gaussian, Inc.,*. Wallingford, CT,: 2004.

128.   Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A., *Int. J. Quantum Chemistry,* **2013,** *113* (18), 2110.

129.   F. Neese, *Comput. Mol. Sci.,* **2012,** *2*, 73.

130.   Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. J.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., *J. Comput. Chem.,* **2005,** *26* (16), 1668.

131.   Gaussian 03, R. C., Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A., Gaussian, Inc., Wallingford CT. 2004.

132.   Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J., *BMC Bioinformatics,* **2006,** *7* (2), 316.

133.   Markham, G. D.; Norrby, P. O.; Bock, C. W., *Biochemistry,* **2002,** *41*, 7636.

134.   Kemege, K. E.; Hickey, J. M.; Lovell, S.; Battaile, K. P.; Zhang, Y.; Hefty, P. S., *J. Bacteriol.,* **2011,** *193* (23), 6517.

135.   Moult, J.; Pedersen, J. T.; Judson, R.; Fidelis, K., *Proteins,* **1995,** *23* (3), ii.

136.   Wu, S.; Zhang, Y., *Nuc. Acid. Res.,* **2007,** *35* (10), 3375.

137.   Zhang, Y., *BMC Bioinformatics,* **2008,** *9*, 40.

138.   Maestro, S., *LLC, New York, NY.* **2016**.

139.   Schrödinger Suite, Induced Fit Docking protocol; Glide version 7.0,. *LLC, New York, NY.* **2016**.

140.   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., *J. Med. Chem.,* **2004,** *47* (7), 1739.

141.   Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., *J. Med. Chem.,* **2004,** *47* (7), 1750.

142.   D.A. Case; R.M. Betz; D.S. Cerutti; T.E. Cheatham; T.A. Darden; R.E. Duke; T.J. Giese; H. Gohlke; A.W. Goetz; N. Homeyer; S. Izadi; P. Janowski; J. Kaus; A. Kovalenko; T.S. Lee; S. LeGrand; P. Li; C. Lin; T. Luchko; R. Luo; B. Madej; D. Mermelstein; K.M. Merz; G. Monard; H. Nguyen; H.T. Nguyen; I. Omelyan; A. Onufriev; D.R. Roe; A. Roitberg; C. Sagui; C.L. Simmerling; W.M. Botello-Smith; J. Swails; R.C. Walker; J. Wang; R.M. Wolf; X. Wu; Xiao, L.; Kollman, P. A., AMBER 2016,. *University of California, San Francisco.* **2016**.

143.   Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., *Journal of Chemical Theory and Computation* **2015,** *11* (8), 3696.

144.   Wang, J.; Wolf, R. M.; Caldwell, J. V.; A., K. P.; Case, D. A., *J. Comput. Chem.,* **2004,** (25), 1157.

145.   Bame, J.; Hoeck, C.; Carrington, M. J.; Butts, C. P.; Jäger, C. M.; Croft, A. K., *Physical Chemistry Chemical Physics,* **2018,** *20* (11), 7523.

146.   Saez, D. A.; Vohringer-Martinez, E., *J. Comput. Aided Mol. Des.,* **2015,** *29,* (10), 951.

147.   Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., *J. Chem. Theory Comput.,* **2013,** *9* (9), 3878.

148.   Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *J. Phys. Chem.;* **1987,** *91* (24), 6269.

149.   Darden, T.; York, D.; Pedersen, L., *J. Chem. Physics,* **1993,** *98* (12), 10089.

150.   Pastor, R. W.; Brooks, B. R.; Szabo, A., *Molecular Physics,* **1988,** *65* (6), 1409.

151.   Pettersen E. F.; Goddard T. D.; Huang C. C.; Couch G. S.; Greenblatt D. M.; Meng E. C.; Ferrin T. E., *J. Comput. Chem.,* **2004,** *25* (13), 1605.

152.   Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V., *Nucleic. Acids Res.,* **2012,** *40* (Web Server issue), W537.

153.   Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A., *Proteins: Structure, Function, and*

Bioinformatics, **2006,** *63* (4), 928.

154.   Karami, Y.; Bitard-Feildel, T.; Laine, E.; Carbone, A., *Sci. Rep.,* **2018,** *8* (1), 16126.

155.   Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E., *Curr. Opin. Struct. Biol.,* **2009,** *19* (2), 120.

156.   Roe, D. R.; Cheatham, T. E., *J. Chem. Theory Comput.,* **2013,** *9* (7), 3084.

157.   Humphrey, W.; Dalke, A.; Schulten, K., *J. Molec. Graphics,* **1996,** *14* (1), 33.

158.   Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., *J. Appl. Cryst.,* **1993,** *26* (2), 283.

159.   Laskowski, R. A.; Moss, D. S.; Thornton, J. M., *J. Mol. Biol.,* **1993,** *231* (4), 1049.

160.   Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C., *Proteins: Structure, Function, and Bioinformatics* **1993,** *17* (4), 412.

161.   Aalten, D. M. F. v.; Findlay, J. B. C.; Amadei, A.; Berendsen, H. J. C., *Protein Eng.,* **1995,** *8* (11), 1129.

162.   Amadei, A.; Linssen, A. B. M.; de Groot, B. L.; van Aalten, D. M. F.; Berendsen, H. J. C., *J. Biomol. Struct. Dyn.,* **1996,** *13* (4), 615.

163.   Teeter, M. M.; Case, D. A., *J. Phys. Chem.,* **1990,** *94* (21), 8091.

164.   Kato, K.; Nakayoshi, T.; Fukuyoshi, S.; Kurimoto, E.; Oda, A., *Molecules,* **2017,** *22* (10), 1716.

165.   Shkurti, A.; Goni, R.; Andrio, P.; Breitmoser, E.; Bethune, I.; Orozco, M.; Laughton, C. A., *SoftwareX* **2016,** *5*, 44.

166.   Kluyver, T.; Ragan-Kelley, B.; Pérez, P.; Granger, B.; Bussonnier, M.; Frederic, J.; Willing, C., *Positioning and Power in Academic Publishing: Players, Agents and Agendas,* **2016**, 87.

167.   Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J., *PLoS. Comput. Biol.;* **2012,** *8* (10), e1002708.

168.   Yaffe, E.; Fishelovitch, D.; Wolfson, H. J.; Halperin, D.; Nussinov, R., *Nucleic Acids Res.,* **2008,** *36*, W210-5.

169.   M. J. Frisch; G. W. Trucks; H. B. Schlegel; G. E. Scuseria; M. A. Robb; J. R. Cheeseman; G. Scalmani; V. Barone; G. A. Petersson; H. Nakatsuji; X. Li; M. Caricato; A. Marenich; J. Bloino; B. G. Janesko; R. Gomperts; B. Mennucci; H. P. Hratchian; J. V. Ortiz; A. F. Izmaylov; J. L. Sonnenberg; D. Williams-Young; F. Ding; F. Lipparini; F. Egidi; J. Goings; B. Peng; A. Petrone; T. Henderson; D.

Ranasinghe; V. G. Zakrzewski; J. Gao; N. Rega; G. Zheng; W. Liang; M. Hada; M. Ehara; K. Toyota; R. Fukuda; J. Hasegawa; M. Ishida; T. Nakajima; Y. Honda; O. Kitao; H. Nakai; T. Vreven; K. Throssell; J. A. Montgomery, J.; J. E. Peralta; F. Ogliaro; M. Bearpark; J. J. Heyd; E. Brothers; K. N. Kudin; V. N. Staroverov; T. Keith; R. Kobayashi; J. Normand; K. Raghavachari; A. Rendell; J. C. Burant; S. S. Iyengar; J. Tomasi; M. Cossi; J. M. Millam; M. Klene; C. Adamo; R. Cammi; J. W. Ochterski; R. L. Martin; K. Morokuma; O. Farkas; J. B. Foresman; Fox;, D. J., *Gaussian, Inc., Wallingford CT.* **2016**.

170.   Hehre, W. J.; Stewart, R. F.; Pople, J. A., *J. Chem. Phys.,* **1969,** *51* (6), 2657.

171.   Wachters, A. J. H., *J. Chem. Phys.,* **1970,** *52* (3), 1033.

172.   Hay, P. J., *J. Chem. Phys.,* **1977,** *66* (10), 4377.

173.   Frisch, M. J.; Pople, J. A.; Binkley, J. S.,

. *J. Chem. Phys.,* **1984,** *80* (7), 3265.

174.   Becke, A. D., *Phys. Rev. A,* **1988,** *38* (6), 3098.

175.   Perdew, J. P., *Phys. Rev. B,* **1986,** *33* (12), 8822.

176.   Harris, T. V.; Szilagyi, R. K., *J. Comp. Chem.,* **2016,** *37* (18), 1681.

177.   Szilagyi, R. K.; Winslow, M. A., *J. Comput. Chem.,* **2006,** *27* (12), 1385.

178.   Harris, T. V.; Szilagyi, R. K., *J. Comput. Chem.,* **2014,** *35* (7), 540.

179.   Zanello, P., Inorganic electrochemistry: theory, practice and application;. *Royal Society of Chemistry, Padstow, Cornwall, UK.* **2013**.

180.   Backes, G.; Mino, Y.; Loehr, T. M.; Meyer, T. E.; Cusanovich, M. A.; Sweeney, W. V.; Adman, E. T.; Sanders-Loehr, J., *J. Am. Chem. Soc.,* **1991,** *113* (6), 2055.

181.   Mouesca, J.-M.; Chen, J. L.; Noodleman, L.; Bashford, D.; Case, D. A., *J. Am. Chem. Soc.,* **1994,** *116* (26), 11898.

182.   Kuznetsov, A. M.; Zueva, E. M.; Masliy, A. N.; Krishtalik, L. I., *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* **2010,** *1797* (3), 347.

183.   Sigfridsson, E.; Olsson, M. H. M.; Ryde, U., *Inorg. Chem.,* **2001,** *40* (11), 2509.

184.   Jensen, G. M.; Warshel, A.; Stephens, P. J., *Biochemistry,* **1994,** *33* (36), 10911.

185.   D.A. Case, R. M. B., W. Botello-Smith, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, J. Swails, R.C.

Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York, P.A. Kollman, AMBER 2016. **2016,**.

186. Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., *J. Chem. Theory. Comput.,* **2013,** *9* (9), 3878.

187. Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., *J. Chem. Theory. Comput.,* **2012,** *8* (5), 1542.

188. Le Grand, S.; Götz, A. W.; Walker, R. C., *Computer Physics Communications,* **2013,** *184,* (2), 374.

189. Humphrey W.; Dalke A.; Schulten K. J., *Molec. Graphics.,* **1996,** *14*, 33.

190. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., *J. Comput. Chem.,* **2004,** *25* (13), 1605.

191. Williams, T.; Kelley, C., Gnuplot 4.5: an interactive plotting program. **2011**.

192. Madhavi Sastry, G.; Adzhigirey, M.; Day, T., *J. Comput. Aided Mol. Des.,* **2013,** *27*, 221.

193. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., *J. Med. Chem.,* **2006,** *49* (21), 6177.

194. Maisuradze, G. G.; Liwo, A.; Scheraga, H. A., *J. Chem. Theory Comput.,* **2010,** *6* (2), 583.

195. David, C. C.; Jacobs, D. J., *J. Mol. Graph. Model.,* **2011,** *31*, 41.

196. Carvalho, H. F.; Roque, A. C.; Iranzo, O.; Branco, R. J., *PLoS One,* **2015,** *10* (9), e0138118.

197. Lepore B. W.; Ruzicka F. J.; Frey P. A.; Ringe D., *Proc. Natl. Acad. Sci. U. S. A.,* **2005,** *102* (39), 13819.

198. Roy, A.; Kucukural, A.; Zhang, Y., *Nat. Protoc.,* **2010,** *5* (4), 725.

199. Hooft, R. W. W.; Sander, C.; Vriend, G., *Bioinformatics,* **1997,** *13* (4), 425.

200. S. C. Lovell; I. W. Davis; W. B. Arendall; P. I. W. de Bakker; J. M. Word; M. G. Prisant; J. S. Richardson; Richardson, D. C., *PROTEINS: Structure, Function, and Genetics,* **2003,** *50*, 437.

201. Pandey, B.; Grover, A.; Sharma, P., *BMC Genomics,* **2018,** *19* (1), 132.

202. Ullah, M.; Hira, J.; Ghosh, T.; Ishaque, N.; Absar, N., *J. Young. Pharm.,* **2012,** *4* (4), 261.

203. Benkert, P.; Biasini, M.; Schwede, T., Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **2011,** *27* (3), 343-50.

204. N. Dhurigai; R. R. Daniel; L. R. Auxilia, *Int. J. Curr. Microbiol. App. Sci.,* **2014,** *3* (10).

205. J. C. Evans; D. P. Huddler; M. T. Hilgers; G. Romanchuk; R. G. Matthews; M. L. Ludwig, *Proc. Natl. Acad. Sci. U.S.A.,* **2004,** *101* (11), 3729.

206. Grove, T. L.; Himes, P. M.; Hwang, S.; Yumerefendi, H.; Bonanno, J. B.; Kuhlman, B.; Almo, S. C.;

Bowers, A. A., *J. Am. Chem. Soc.,* **2017,** *139*, 11734.

207.   Ereno-Orbea, J.; Majtan, T.; Oyenarte, I.; Kraus, J. P.; Martinez-Cruz, L. A., *Proc. Natl. Acad. Sci. USA,* **2014,** *111* (37), E3845.

208.   Mas-Droux, C.; Biou, V.; Dumas, R., *J. Biol. Chem.,* **2006,** *281* (8), 5188.

209.   Guja, K. E.; Venkataraman, K.; Yakubovskaya, E.; Shi, H.; Mejia, E.; Hambardjieva, E.; Karzai, A. W.; Garcia-Diaz, M., *Nucleic. Acids. Res.,* **2013,** *41* (16), 7947.

210.   Davis, K. M.; Schramma, K. R.; Hansen, W. A.; Bacik, J. P.; Khare, S. D.; Seyedsayamdost, M. R.; Ando, N., *Proc. Natl. Acad. Sci. U. S. A.,* **2017,** *114* (39), 10420.

211.   G., M. S.; Adzhigirey, M.; Day, T., *J. Comput. Aided Mol. Des.,* **2013,** *27*, 221.

212.   Repasky, M. P.; Murphy, R. B.; Banks, J. L., *J. Comput. Aided Mol. Des.,* **2012,** *26*, 787.

213.   Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., *J. Chem. Theory Comput.,* **2016,** *12* (1), 281.

214.   Zhao, Y.; Zeng, C.; Massiah, M. A., *PLOS ONE* **2015,** *10* (4), e0124377.

215.   Volbeda, A.; Montet, Y.; Vernède, X.; Hatchikian, E. C.; Fontecilla-Camps, J. C., *Int. J. Hydrogen Energy,* **2002,** *27* (11), 1449.

216.   Cano, M.; Volbeda, A.; Guedeney, G.; Aubert-Jousset, E.; Richaud, P.; Peltier, G.; Cournac, L., *Int. J. Hydrogen Energy,* **2014,** *39* (30), 16872.

217.   Bin, P.; Huang, R.; Zhou, X., *BioMed Res. Int.,* **2017,** *2017*, 6.

218.   Collazo, L.; Klinman, J. P., Control of the Position of Oxygen Delivery in Soybean Lipoxygenase-1 by Amino Acid Side Chains within a Gas Migration Channel. *J Biol Chem* **2016,** *291* (17), 9052-9.

219.   Pandelia, M. E.; Nitschke, W.; Infossi, P.; Giudici-Orticoni, M. T.; Bill, E.; Lubitz, W., *Proc. Natl. Acad. Sci. U. S. A.,* **2011,** *108* (15), 6097.

220.   Kassner, R. J.; Yang, W., *J. Am. Chem. Soc.,* **1977,** *99* (13), 4351.

221.   Czernuszewicz, R. S.; Macor, K. A.; Johnson, M. K.; Gewirth, A.; Spiro, T. G., *J. Am. Chem. Soc.,* **1987,** *109* (23), 7178.

222.   Chen, D.; Frey, P. A.; Lepore, B. W.; Ringe, D.; Ruzicka, F. J., *Biochemistry,* **2006,** *45* (42), 12647.

223.   Perrin, B. S., Jr.; Niu, S.; Ichiye, T., *J. Comput. Chem.,* **2013,** *34,* (7), 576.

224.   Tran, K. N.; Niu, S.; Ichiye, T., *J. Comput. Chem.,* **2019,** *40* (12), 1248.

225.   Link, T. A., In *Adv. Inorg. Chem.,*, 1999; Vol. 47, p 83.

226.   Liu, J.; Chakraborty, S.; Hosseinzadeh, P.; Yu, Y.; Tian, S.; Petrik, I.; Bhagi, A.; Lu, Y., *Chem. Rev.,* **2014,** *114* (8), 4366.

227.   Saichana, N.; Tanizawa, K.; Pechoušek, J.; Novák, P.; Yakushi, T.; Toyama, H.; Frébortová, J., *J. Biochem.,* **2016,** *159*, 87.

228.   Saichana, N.; Tanizawa, K.; Ueno, H.; Pechoušek, J.; Novák, P.; Frébortová, J., *FEBS Open Bio.,* **2017,** *7*, 1864.

229.   A. Warshel; P. K. Sharma; M. Kato; Y. Xiang; H. Liu; Olsson, M. H. M., *Chem. Rev.,* **2006,** *106*, 3210.

230.   Cohen, J.; Kim, K.; King, P.; Seibert, M.; Schulten, K., Finding gas diffusion pathways in proteins: application to O2 and H2 transport in CpI [FeFe]-hydrogenase and the role of packing defects. *Structure* **2005,** *13* (9), 1321-9.

231.   Bergonzo, C.; Cheatham, T. E., 3rd, *J. Chem. Theory Comput.,* **2015,** *11* (9), 3969.

232.   Sharma, S.; Sivalingam, K.; Neese, F.; Chan, G. K., *Nat. Chem.,* **2014,** *6* (10), 927.

233.   Hennemann, M.; Clark, T., *J. Mol. Model.,* **2014,** *20*, 2331.

234.   Fricker, M. D., *Antioxidants & Redox Signaling,* **2015,** *24* (13), 752.

235.   Stuyver, T.; Huang, J.; Mallick, D.; Danovich, D.; Shaik, S., *J. Comput. Chem.,* **2019,** *9999*, 1.

236.   Prah, A.; Frančišković, E.; Mavri, J.; Stare, J., *ACS Catalysis,* **2019,** *9* (2), 1231.

237.   Fried, S. D.; Bagchi, S.; Boxer, S. G., *Biophysics,* **2014,** *346* (6216), 1510.

238.   Turker, L., *Scientific World Journal,* **2012**, 985958.

239.   Klinska, M.; Smith, L. M.; Gryn'ova, G.; Banwell, M. G.; Coote, M. L., *Chem. Sci.,* **2015,** *6,* (10), 5623.

240.   Akiva, E.; Brown, S.; Almonacid, D. E.; Barber, A. E.; Custer, A. F.; Hicks, M. A.; Huang, C. C.; Lauck, F.; Mashiyama, S. T.; Meng, E. C.; Mischel, D.; Morris, J. H.; Ojha, S.; Schnoes, A. M.; Stryke, D.; Yunes, J. M.; Ferrin, T. E.; Holliday, G. L.; Babbitt, P. C., *Nucleic. Acids. Res.,* **2014,** *42* (Database issue), D521.

241.   Jiménez-González, C.; Poechlauer, P.; Broxterman, Q. B.; Yang, B.-S.; am Ende, D.; Baird, J.; Bertsch, C.; Hannah, R. E.; Dell'Orco, P.; Noorman, H.; Yee, S.; Reintjens, R.; Wells, A.; Massonneau, V.; Manley, J., *Organic Process Research & Development, Am. Chem. Soc.,* **2011,** *15* (4), 900.