

**Fundamental computational studies of the structural  
dynamical relationships of  
multifunctional enzyme type 1 (MFE-1)**

vorgelegt von

M. Sc.

**Nico Herder**

an der Fakultät II – Mathematik und Naturwissenschaften

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzende:	Prof. Dr. Franziska Heß
Gutachterin:	Prof. Dr. Maria Andrea Mroginski
Gutachter:	Dr. Patrick Scheerer
Gutachterin:	Prof. Dr. Katrina Forest

Tag der wissenschaftlichen Aussprache: 22. September 2022

Berlin 2022



## Abstract

The multifunctional enzyme type 1 (MFE-1) plays a crucial role in the fatty acid metabolism. It is a monomeric enzyme that catalyzes three reactions in the  $\beta$ -oxidation pathway: the isomerization, the hydrogenation, and the dehydrogenation. These three reactions occur at two different active sites. The MFE-1 consists of a total of five domains A, B, C, D and E. The first active site is located in the crotonase part which is formed by domain A, while domain C (in the HAD part) hosts the second active site. Isomerization and hydrogenation are carried out at the first active site, while the third reaction (dehydrogenation) occurs at the second active site. To date, the detailed mechanistic functions as well as the basic structure-dynamics relations are not known in detail. Moreover, no crystal structures of MFE-1 with embedded substrates in the second active site or intermediates from the first active site have been shown so far. This dissertation addresses these issues by the investigation of various MFE-1 constructs through molecular dynamics simulations in the form of classical and accelerated MD.

Regarding the first active site, two residues are proposed to be essential for the catalytic functionality of MFE-1. Therefore, in the first step, these two residues, the glutamates E103 and E123, were investigated. Their protonation state, which was previously unknown, was determined and their function was investigated in relation to the mechanism concerning the hydrogenation reaction. In addition to the protonation state, within the scope of this dissertation hitherto undiscovered conformational as well as distance dependencies could be identified for the first time.

Furthermore, the MFE-1 has been investigated for fundamental dynamic structure-function relations using dynamic cross correlation and normal mode analysis. Based on these domain-dynamics-structure relations, correlations of MFE-1 were revealed which possibly indicate potential substrate channeling to enable the substrate transport from the first to the second active site.

Previously no structural model with an embedded substrate in the second active site was available. For this reason, an MFE-1 model with a substrate in the second active site was generated by docking.

These calculations provide essential insights into the mechanism of MFE-1 that were undiscovered so far. Additionally, the results serve as a model and fundamental basis for future investigations on its mechanistic properties.





## Zusammenfassung

Das multifunktionale Enzym Typ 1 (MFE-1) spielt eine entscheidende Rolle im Fettsäurestoffwechsel. Es ist ein monomeres Enzym, das drei Reaktionen des  $\beta$ -Oxidationsweges katalysiert: die Isomerisierung, die Hydrierung und die Dehydrierung. Diese drei Reaktionen laufen an zwei verschiedenen aktiven Zentren ab. Das MFE-1 besteht aus insgesamt fünf Domänen A, B, C, D und E. Das erste aktive Zentrum befindet sich im Crotonase-Abschnitt, der von der Domäne A gebildet wird, während die Domäne C (im HAD-Abschnitt) das zweite aktive Zentrum beherbergt. Die Isomerisierung und Hydrierung laufen im ersten aktiven Zentrum ab, während die dritte Reaktion (Dehydrierung) im zweiten aktiven Zentrum stattfindet. Bis heute sind die detaillierten mechanistischen Funktionen sowie die grundlegenden Struktur-Dynamik-Beziehungen nicht im Detail bekannt. Außerdem wurden bisher keine Kristallstrukturen von MFE-1 mit eingebetteten Substraten im zweiten aktiven Zentrum oder Zwischenprodukten aus dem ersten aktiven Zentrum gezeigt. In dieser Dissertation werden diese Fragen durch die Untersuchung verschiedener MFE-1-Konstrukte mittels Molekulardynamiksimulationen in Form von klassischer und beschleunigter MD angegangen.

Für das erste aktive Zentrum stehen dabei zwei Aminosäuren im Fokus, die für die katalytische Funktionalität von MFE-1 essenziell sind. Daher wurden im ersten Schritt der Protonierungszustand, der bisher unbekannt war, beider Aminosäuren (Glu103 und Glu123) berechnet. Darüber hinaus wurde die Funktion der Glutamate in Bezug auf den Mechanismus hinsichtlich der Hydrierungsreaktion untersucht. Neben dem Protonierungszustand konnten im Rahmen dieser Dissertation erstmals bisher unentdeckte Konformations- sowie Abstandsabhängigkeiten identifiziert werden.

Außerdem wurde das MFE-1 mittels dynamic Cross Correlation und Normalmodenanalyse auf grundlegende dynamische Struktur-Funktions-Beziehungen untersucht. Basierend auf diesen Dynamik-Struktur-Beziehungen wurden Domänen-Korrelationen aufgedeckt, die möglicherweise auf ein potentiell Substrat-Channelling hinweisen, um den Substrattransport vom ersten zum zweiten aktiven Zentrum zu ermöglichen.

Zuvor war kein 3D-Modell mit einem eingebetteten Substrat im zweiten aktiven Zentrum verfügbar. Aus diesem Grund wurde ein MFE-1-Modell mit einem Substrat im zweiten aktiven Zentrum durch Docking erstellt.

---

Diese Berechnungen liefern wesentliche Einblicke in den Mechanismus von MFE-1, die bisher unentdeckt waren. Darüber hinaus dienen die Ergebnisse als Modell und grundlegende Basis für künftige Untersuchungen der mechanistischen Eigenschaften.

---

---

## Acknowledgement

An dieser Stelle möchte ich allen beteiligten Personen meinen großen Dank aussprechen, die mich bei der Umsetzung meiner Dissertation unterstützt haben.

Besonders danken möchte ich mich bei Prof. Dr. Maria Andrea Mroginski für die ausgezeichnete Betreuung bei der Durchführung und Umsetzung meiner gesamten Arbeit. Sowie für ihre herzliche Art bei Problemen immer ein offenes Ohr zu haben.

Dr. Patrick Scheerer und Prof. Dr. Katrina Forest spreche ich ebenfalls meinen herzlichsten Dank für die Übernahme der Aufgabe des Zweit- bzw. Drittgutachters aus.

Ebenso gilt mein Dank Prof. Dr. Franziska Heß für die Verantwortung und Durchführung des Vorsitzes im Rahmen meiner wissenschaftlichen Aussprache.

Prof. Dr. Rik Wierenga möchte ich danken für die anregenden Diskussionen und seiner fachlichen Expertise auf dem Gebiet des MFE-1 an der ich teilhaben durfte.

Prof. Dr. Matt Jacobson und den Mitarbeitern seiner AG möchte für meine Einarbeitung in die Thematik des DOCKING bedanken.

Außerdem möchte ich mich bei der gesamten AG Mroginski sowie ehemaligen Mitarbeitern wie Dr. Tillmann Utesch und Aref Takiden bedanken, mit denen ich stets in wissenschaftlichen Austausch treten konnte.

Auch möchte ich Jürgen Krauss meinen Dank aussprechen, der im Rahmen der Umsetzung dieser Dissertation für den IT-Support verantwortlich war und mit Rat und Tat sowie produktiven Gesprächen stets zur Verfügung stand.

Lars Paasche möchte ich für seine Ruhe und Ermutigungen danken.

Meinen Freunden danke ich für ihre Unterstützung.

Besonderer Dank gilt meiner Mutter sowie meiner Frau für ihre Geduld und Ermutigungen während der Arbeit an dieser Dissertation. Natürlich spreche ich meinem Bruder, meiner Schwiegermutter und Schwägerin ebenfalls unendlichen Dank aus.



## Outline

Abstract .....	3
Zusammenfassung .....	5
Acknowledgement .....	9
Outline.....	11
1 Introduction .....	15
1.1 Fatty acids .....	15
1.1.1 Properties.....	15
1.1.1.1 Acyl fatty acids .....	15
1.1.1.2 Coenzyme A .....	16
1.1.1.3 Central component in the fatty acid-synthase complex .....	18
1.1.2 Transport and activation .....	20
1.1.2.1 Transport via proteins .....	21
1.1.3 Protein and fatty acid interactions.....	24
1.1.4 Catabolism .....	25
1.2 Multifunctional enzyme type 1 (MFE-1) .....	28
1.2.1 Overall structure .....	31
1.2.2 CH <sub>2</sub> -helix.....	34
1.2.3 Anchor region and tunnel .....	34
1.2.4 Crotonase part and active site I .....	35
1.2.5 HAD part and active site II .....	37
1.3 Substrate channeling.....	39
1.3.1 Basic concept .....	39
1.3.2 Examples .....	40
1.3.3 Multifunctional enzyme type 1 (MFE-1) .....	43
2 Methods .....	45
2.1 Classical molecular dynamics simulation .....	45
2.1.1 Evolution .....	45
2.1.2 Principles of classical MD simulations .....	46
2.1.2.1 Integration algorithms .....	46
2.1.2.2 Statistics .....	47
2.1.3 Force fields .....	49
2.1.3.1 Principles of molecular mechanics force fields.....	49
2.1.3.2 Bonded potential .....	49
2.1.3.3 Non-bonded potential .....	53
2.2 Accelerated MD simulations .....	54

---

2.3 Normal mode analysis.....	58
2.3.1 Background .....	58
2.3.2 Elastic network model.....	59
2.4 Dynamic cross correlation.....	60
2.5 Principal component analysis .....	61
2.5.1 Definition and principle .....	61
2.5.2 Interpretation of eigenvalues and eigenvectors .....	64
2.5.3 Eigenvector projection .....	65
2.5.4 Choice of eigenvalues .....	66
2.6 pKa values in biological systems .....	66
2.6.1 Calculation .....	66
2.6.2 Proteins .....	68
2.6.3 Calculation of $\Delta G_c$ .....	68
2.6.4 Karlsberg2+ .....	70
2.7 Docking.....	70
2.7.1 Theory .....	70
2.7.2 Ligand protein preparation .....	71
2.7.3 Sampling methods .....	73
2.7.4 Scoring functions.....	74
2.7.5 Docking methods .....	76
2.7.6 Examples of docking applications .....	77
3 Motivation.....	81
4 Results – Active site I of MFE-1 .....	83
4.1 Introduction .....	83
4.2 Model setup and preparation .....	84
4.3 Protonation state of Glu103 and Glu123 .....	85
4.3.1 Structure stability.....	85
4.3.2 pKa calculations .....	87
4.4 Structural changes of glutamates E103 and E123 .....	90
4.4.1 MFE-1 apoprotein .....	90
4.4.2 MFE-1 complexed with 3S-hydroxydecanoyl-CoA .....	94
4.4.3 MFE-1 holoprotein .....	99
4.5 Principal component analysis .....	103
4.5.1 MFE-1 apoprotein .....	103
4.5.2 MFE-1 complexed with 3S-hydroxydecanoyl-CoA .....	106
4.5.3 MFE-1 holoprotein .....	108

---



---

4.6 Discussion.....	111
4.7 Summary .....	115
5 Results – Docking .....	117
5.1 Introduction .....	117
5.2 Modell setup and preparation .....	118
5.3 Virtual screening .....	119
5.4 Ligand conformers.....	127
5.5 Discussion.....	129
5.6 Summary .....	132
6 Results – Domain motions in MFE-1 .....	135
6.1 Introduction .....	135
6.2 Methods and model setup.....	136
6.3 Normal mode analysis.....	138
6.4 Dynamic cross correlation.....	151
6.4.1 MFE-1 apoprotein .....	151
6.4.2 MFE-1 complexed with 3S-hydroxydecanoyl-CoA .....	152
6.4.3 MFE-1 holoprotein .....	153
6.5 Discussion.....	155
6.6 Summary .....	156
7 Conclusion and outlook .....	157
8 References.....	161



## 1 Introduction

### 1.1 Fatty acids

The lipidoma located in the cell comprises more than 1000 lipid molecules, which are processed by complex metabolic pathways through transport, activation, or modification[1]. A multitude of receptors and transcriptions, which are responsible for the detection of extracellular and intracellular concentrations, is responsible for this. As a result, the cell is constantly adapting to environmental and physiological conditions by regulating the expression levels and catalytic activities of the proteins or enzymes involved in lipid metabolism. The following is an overview of important properties that are important for the characterization of lipids, fatty acids (FA), and their derivatives. Apart from that, biologically relevant systems that play a key role are presented in detail.

#### 1.1.1 Properties

In essence, lipids are understood as a class of biomolecules with a wide variety of classifications. They are differentiated into hydrophobic or amphipathic molecules. Generally, there are eight classes, with fatty acids being one of the most basic biological lipid classes. They act as storage (oxygen ester) or as reactive thioesters for relevant metabolic pathways[2].

##### 1.1.1.1 Acyl fatty acids

Fatty acids (FA) are subdivided into several subclasses, where the reduced members are mainly represented by straight-chain saturated FA with a terminal carboxyl group and a hydrophobic acyl chain. This acyl chain can be present in turn through various modifications, such as several double bonds or triple bonds. Individual positions can also be substituted by other heteroatoms such as acid or nitrogen atoms. Furthermore, there can be various branching by methyl groups or other acyl chains, which can lead to cyclic compounds. Also, functionalities like nitriles, epoxides, or alcohols are present.

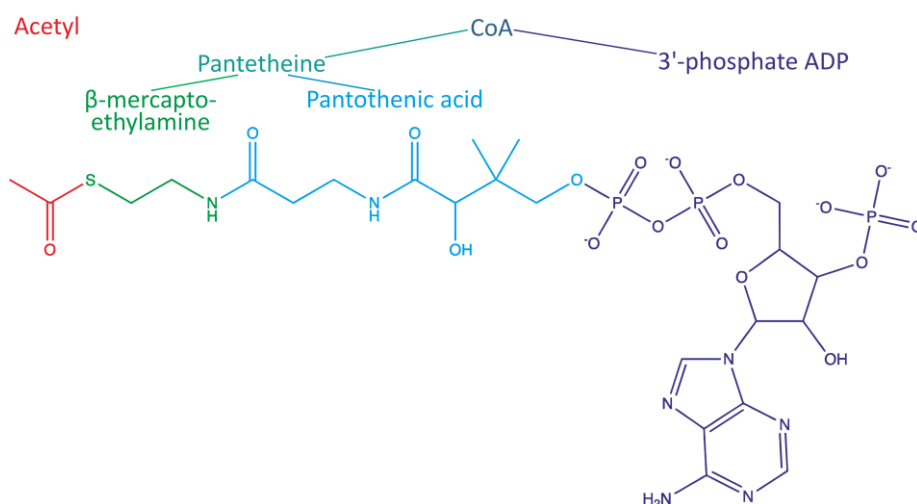
An alternative subcategory of significant biological importance is that of thioesters, in which, among others, the important coenzyme A, which is involved as a cofactor in a large number of metabolic pathways, is found. For a simpler understanding of the

biological context, FA or the straight-chain variants are divided into short (< eight carbon atoms), medium (eight to 12 carbon atoms), long-chain (13 to 18 carbon atoms), and very long-chain (>20 carbon atoms) acids. The length refers to the chain length itself. Within the chain length, the stereospecific aspects are decisive for the different functionalities. The position of a double bond or a cis/trans isomerism can decide whether FA can be processed directly by  $\beta$ -oxidation enzymes or whether co-enzymes are necessary.

The solubility of free FA or fatty acyl-CoA due to the chain length is accordingly insufficient. That means the longer the tail, the more dominant the hydrophobic part. The influence of the ionic strength and the presence of  $Mg^{2+}$ -ions in the medium also has a negative effect on the solubility[3]. Due to the low solubility, micelle and aggregate formation occurs in the aqueous medium, but so far only a few data on solubilities in physiological aqueous solutions are available. In the gastrointestinal tract, bile acid reduces micelle formation, which increases solubility, and in the blood, the majority of FA is bound to serum albumin while smaller amounts are bound to plasma lipoproteins[3].

#### **1.1.1.2 Coenzyme A**

Probably the best known and most important auxiliary enzyme is the co-enzyme A (CoA). First isolated in 1946, the structure at atomic level was published in 1953[4, 5]. Its natural and acyl-CoA variants are involved in a variety of primary and secondary metabolic pathways and catabolism, as well as regulatory processes[6]. Furthermore, acetyl-CoA (which contains two additional carbon atoms in the acyl chain) can be considered a key metabolite (Figure 1), as it is involved in the metabolism of biological compounds such as amino acids, carbohydrates, and fats.



*Figure 1: Chemical structure of CoA with its two structural parts pantetheine (cyan) and 3'-phosphate ADP (violet). Pantetheine is divided in two subregions represented by  $\beta$ -mercaptoethylamine (green) and pantothenic acid (blue).*

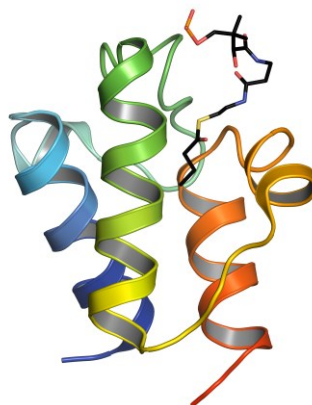
As Figure 1 shows, CoA is divided into three main parts. The first section is formed by the  $\beta$ -mercaptoethylamine and serves the thiol functionality, where the connection to a fatty acid chain takes place. The second subunit is continued by amide bonding in the form of the pantothenic acid unit of the CoA. Due to the property of this unit to be relatively rigid, the binding entropically to enzyme complexes is favored[7]. At a chiral center of the pantetheine unit (Figure 1) H- and OH-substituents are found. Furthermore, the 3'-phosphate ADP- and the  $\beta$ -mercaptoethylamine-unit are bound to these substituents. In addition, these units also contain the dimethyl groups. The idea that the methyl groups play a major role in the dynamics and conformational properties of the CoA has been repeatedly raised, but no experimental data are known so far and therefore these groups can probably only be assigned a subordinate role[7]. Studies show that interactions between the pantetheine group, adenine and the 3'-phosphate group are likely to occur[7, 8]. By means of a 5'-pyrophosphate and 3'-phosphate-adenosyldiphosphate bridge, the pantetheine unit is linked together, which in turn contains phosphoanhydrides. While the 3'- and 5'-phosphates are charged at a physiologically pH, they often interact with charged residues within binding pockets of the different enzymes by hydrogen bridge or salt bridge contacts. The adenine portion of the last subunit (Figure 1) is often involved in ligand binding in the binding pockets by aromatic

stacking with aromatic residues[9]. Although the structure of the CoA appears to be very complex, it is even more surprising that the functionality is rather basic.

CoA is responsible for the activation of the fatty acid for cellular metabolism, with the enzymatic reaction taking place at the thiol group. Other reactions concerning the thioester involve the complete thioester group and the acyl moiety. Nudix hydrolases are an exception[10]. In principle, however, in CoA metabolism the structural motif of the remaining CoA serves as a recognition feature (key-lock principle) for enzymatic binding sites. The acyl derivative is formed by acylation of thiol, which can be activated in two ways: either, by an electrophilic attack of the thioester carbonyl using a nucleophilic co-substrate or by deprotonation of the  $\alpha$ -carbon of the thioester and a subsequent reaction as nucleophile[11]. The reactivity in this molecule is determined by the presence of the sulfur atom, which instead of oxygen causes the carbonyl atom to be more susceptible to nucleophilic attack. This results in easier deprotonation at the  $\alpha$ -carbon atom. When passing through the transition state of both reactions, a negative charge is formed on the thioester oxygen, which can be stabilized by an oxyanion hole in the protein[12, 13].

#### **1.1.1.3 Central component in the fatty acid-synthase complex**

The acyl carrier protein (ACP, Figure 2) represents the essential component and thus plays a central role in the fatty acid synthesis of type II. It is the carrier of monomeric enzymes in prokaryotes and plants. In type I fatty acid synthesis, it is covalently bound in the enzyme complexes[14]. With a size of 9 kDA and acidic property, all ACPs (with the exception of pre-sequential prokaryotic ACP, which are cleavable) are structurally similar and also show a high degree of sequence similarity compared to ACPs from *E. coli*[15]. Studies have shown that soluble ACP coexist in association with a respiratory complex I-associated ACP. The latter is responsible for the stability of the whole complex[16].



*Figure 2: Overall view of Acyl-Carrier protein (PDB: 1L0I) in cartoon representation (rainbow colored). The ligand is illustrated as sticks[17].*

The ACP is important because it carries all intermediates of the Type II fatty acid synthesis that are required. The intermediates are covalently bonded to the terminal sulfhydryl of the 4'-phosphopantetheine group by a thioester bond[18, 19]. The sulfhydryl is unique because it is the only thiol group in *E. coli* ACPs that is bound to the protein by a phosphodiester bond of Ser36 of the Asp-Ser-Leu motif[20, 21].

In the absence of a direct link between different enzymes, the intermediate is transported by ACP[22]. Besides fatty acid synthesis, ACP is also involved in the synthesis of *e.g.*, bacterial glycerolipids or polyamino acid antibiotics[23, 24]. More detailed structural information for *E. coli* ACP can be found in Holak *et al.* 1988[20].

Since the first crystal structure, several derivatives of ACP have been published, which differ in loop regions, which in turn are based on different experimental methods[14, 25]. It is assumed that the interaction of ACP and the enzyme is due to helix2[26]. An exposed hydrophobic binding pocket (Ser36) provides the binding site for the ACP on the surface. NMR shifts were used to discover the binding site of acyl intermediates in butyryl ACP. While the butyryl group interacts with helix3 and helix2. The NMR data confirm the assumption that ACPs interact with acyl intermediates in the hydrophobic interstitial space[17]. Suggestions that the acyl chain as well as the prosthetic group are subject to a two-state exchange reaction were made by White *et al.* In this case, the prosthetic group is bound to the ACP surface, whereas in the other state the prosthetic group is exposed to the solvent[14]. This shifts the state equilibrium to the extent that the chain length becomes the determining factor. Therefore, as the length increases, the interaction with the hydrophobic-based protein is increasingly favored. This assumption

is supported by the fact that the acyl chain is able to counteract pH-induced denaturation[27].

### **1.1.2 Transport and activation**

The energy obtained by the  $\beta$ -oxidation requires prior activation of the fatty acids. For this, the fatty acids have to be converted into the structures of CoA derivatives, which occurs during the cancer cycle. Thus the acyl-CoA synthetase (ACS) is necessary to link the fatty acid with the coenzyme A in a kind of two-stage reaction using ATP[28]. The mechanism is not completely understood so far, although studies suggest that both ACSL (long chain ACS) and very long chain (ACSVL) synthetases can activate branched fatty acids[29, 30]. Activation of bile acid intermediates, on the other hand, occurs in liver microsomes and requires cholyl-coenzyme A synthetase and trihydroxycoprostanoyl-coenzyme A synthetase[31].

Fujino *et al.* were able to show that mitochondrial ACS is responsible for the activation of short and medium length fatty acids. There is a discrepancy to the activation ability of medium-length fatty acids of peroxisomal membrane ACSL[32, 33]. Hesler *et al.* and Aas could show that long-chain fatty acids are activated on the cytosolic surface and actively transported[34, 35]. The acyl molecule is attached to carnitine during transport (by carnitine acyl transferase I), which is located on the outer membrane of the mitochondria, and then transferred into the mitochondria via this membrane with the aid of an acylcarnitine translocase[36-38]. Once in mitochondrium, the carnitine is separated and the acyl group is bound to the coenzyme A[39]. While long- and medium-chain fatty acids are carried through the membrane by a transferase, short-chain fatty acids have barrier-free access to the mitochondrial interior and are activated there by short- and medium-chain ACS[37]. The number of acyl-CoA synthetase in the peroxisomal membrane is at least two. While one is found on the cytosolic side and activates long-chain fatty acids (for transport into the peroxisomes), the other synthetase is responsible for the activation of very long-chain fatty acids[40-42]. According to Vanhove and Wanders, the long-chain acy-CoA synthetase is responsible for the activation of 2-methyl-branched fatty acids[29, 43]. Although it may still be relatively speculative, based on experimental data the transport is barrier-free for molecules of size <1 kDa (in contrast to acyl-CoA)[39, 44]. The ACSL enzyme is identical to the



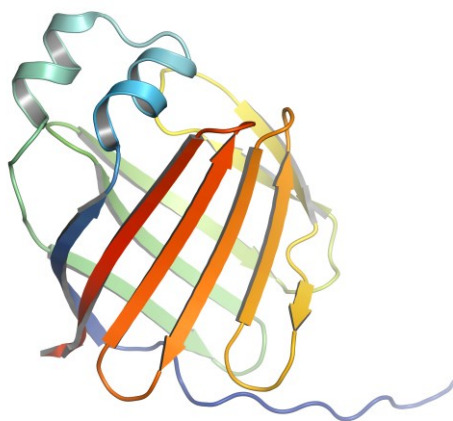
enzymes found in mitochondria and endoplasmatic reticulum (ER) on the protein level as Suzuki was able to show. In contrast, ACSL (associated with the peroxisomal membrane) activates the very long-chain fatty acids within the peroxisomes[41, 45, 46].

Within the mitochondria this enzyme is not found[45], but on the ER. This could elucidate why very long-chain fatty acids are processed exclusively by  $\beta$ -oxidation in the peroxisomes and not in the mitochondria. Comparing ACSL and ACSVL at the structural level and their biochemical and immunochemical properties, both systems are different enzymes. ACSVL are also found in human peroxisomes. In addition to their ACSL activity, they are also capable of oxidizing fatty acids with methyl branches and very long-chain fatty acids[47].

#### 1.1.2.1 Transport via proteins

Due to their low solubility in aqueous environments, fatty acids are embedded in cellular-based membrane systems. This ensures a specific mechanism in the metabolism to transport the fatty acids to the required location. Several proteins are involved, such as the fatty acid binding protein (FABP), the sterol carrier protein type 2 (SCP-2) and the acyl-CoA binding protein (ACBP).

**Fatty acid binding proteins (FABP)** are also called lipid-binding proteins and control the lipid levels of various tissues (Figure 3). The size is about 14-16 kDa and they bind variants of amphiphilic molecules such as fatty acids, bile acids and retinoids[48, 49].



*Figure 3: Overall view of FABP (PDB: 2F73) in cartoon representation (rainbow colored)[50, 51].*

The outstanding feature of the FABP family is that it has more than one billion years old precursors, which are available in as apo- and holoproteins by X-ray and NMR structure elucidation[52]. Due to the property to bind many essential lipids, they are assigned several possibilities of the functionality. On the one hand they are passive "buffers" to increase the solubility of fatty acids and other lipids. On the other hand, they are a kind of active chaperones, which facilitate the transport of lipids to specific targets (membranes or other proteins)[53, 54]. The expression of fatty acid binding proteins takes place in tissues with a high fat metabolism, *e.g.*, skeletal muscle, liver, intestine or also generally fatty tissue. The basic content of FABP in cells is reached in adipocytes with about 3% or 8% in other tissues[55, 56].

The specific assignment for certain lipids of the different FABPs is probably based on electrostatic properties, which are linked to the surface potential. As a consequence, there would be a different approach speed to the binding pocket in the respective FABP[57]. Consequently, the ligand affinity would be due to structural changes on the surface pattern, which would strengthen or weaken and suppress the protein-ligand interactions.

Another possibility is based on folding stability, which affects the protein stability itself and thus influences the half-life of the cell[58]. While other lipid binding proteins show a helix fold[59-61], FABP show  $\beta$ -strands in the form of ten antiparallel forms, which are a kind of  $\beta$ -clam. The cave formed by the  $\beta$ -strands is closed by means of short  $\alpha$ -helices and terminated with hydrophobic residues. The place where the ligand essentially occurs is the region around the beta C $\beta$ -D $\beta$  turn and the  $\alpha$ -helix II. Inside the barrel (shape of the cave) is characterized by hydrophobic as well as charged residues. The ligand itself fills about 40% of the binding pocket, while the rest is occupied by solvent molecules[62, 63]. The subdivision of the FABP is based on the amino acid sequence or binding properties of different ligands. Subunit I binds vitamin A derivatives. The head group of the ligand is always involved in non-covalent interactions with its carboxyl group in the binding pocket[64].

In contrast to subunit I, subunit II also binds bulky lipids. The type of binding is different: type II is also able to bind two long-chain fatty acids, in contrast to type I, where this is not the case[65]. In subunit III, the intestinal type FABP is the only representative. This protein binds at most one long-chain fatty acid. If required, it is available in a rather curved non-linear form, similar to subunit I but not identical[66]. Subunit IV forms the

last form of the fatty acid binding protein family tree. It is also the most diverse one. The characteristic of this FABP is the way the ligands or fatty acids are bound. In contrast to the simple curvature, the bound ligands are bound in a U-shape up to the spiral form of polyunsaturated fatty acids with a very long chain[67].

**Sterol carrier protein (SCP)** is responsible for the distribution of cholesterol between the aqueous phase and membranes (Figure 4). Since it is very poorly soluble, the SCP ensures the specific transport of the cholesterol to its destination[68]. One of the best elucidated sterol carriers is SCP-2, which prevents the hydrolysis of various coenzyme A-esters by peroxisomal thioesterases. It also serves as an antioxidant to protect against superoxide peroxidation of fatty acyl-CoA compounds. Structurally it was elucidated by Szyperski *et al.* and Garcia *et al.*[69, 70]. Based on NMR structures the binding of ligands and their recognition of a hydrophobic tunnel could be identified. SCP-2 can also be found as a part of large enzyme complexes, *e.g.*, in MFE-2 as a separate domain[71, 72]. The structure of SCP-2 like domain could be shown by Haapalainen *et al.*[73].

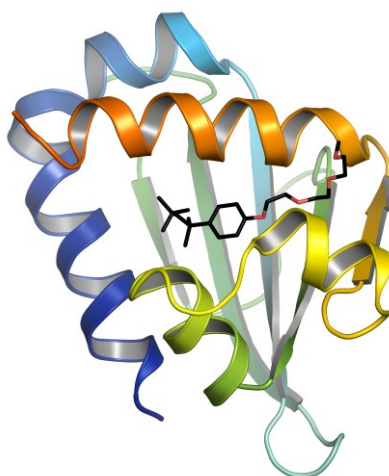
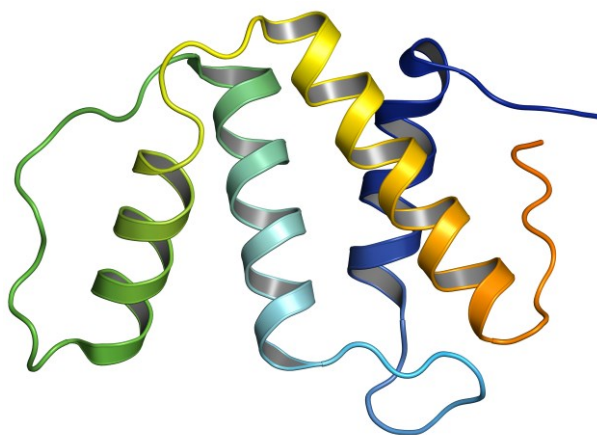


Figure 4: Overall view of SCP (PDB: 1IKT) in cartoon representation (rainbow colored). Ligand is represented as sticks[73].

Cavities with hydrophobic orientation as in SCP-2 are also present in plants in the form of lipids. Although they are smaller and do not show sequence similarity, the other properties such as width, shape, hydrophobicity, and length are very similar. Therefore, a consistent function can be assumed[74].

**Acyl-Coenzyme-A binding proteins (ACBP)** are responsible for the specific 1:1 stoichiometrically important separation of long-chain fatty acids (Figure 5). Their size is 10 kDa and they are present in various multi-domain proteins, such as membrane-bound (N-terminal ACBP) proteins, multifunctional  $\beta$ -oxidation enzymes (ACBP and  $\Delta 3$ - $\Delta 2$  enoyl-CoA isomerase) or proteins with an ACBP domain with an ankyrin repeat[75-77].



*Figure 5: Overall view of ACBP (PDB: 5IJM) in cartoon representation (rainbow colored)[78].*

In addition to NMR and X-ray structures were also published[6, 79, 80]. In the structures, the ligand or a part of it is embedded in a kind of groove which has hydrophobic character, while the acyl chain is protected by the end of the coenzyme A (by spiral folding over the chain).

Apart from lipid metabolism, ACBP serves as a diazepam binding inhibitor (DBI). Due to the binding of ACBP to benzodiazepine receptors (PBR) on the mitochondrial membranes, ACBP is used as an allosteric modulator on steroid synthesis. Thus, it belongs to a large family of regulatory active peptides of this function[81].

### 1.1.3 Protein and fatty acid interactions

For interactions to occur and lead to binding and reaction in an enzymatic complex, hydrophobic as well as hydrophilic regions are required within the binding pockets of enzymes. Due to the polarity of the acyl-CoA molecule, these conditions are essential. Furthermore, the poor solubility of the acyl chain of the acyl-CoA in the aqueous phase

requires efficient localization of the environment. The free binding energy ( $\Delta G$ ), which is important for ligand binding, can be divided into two terms: van der Waals interactions essentially between surfaces, i.e. ( $\Delta H$ ); and the entropy ( $\Delta S$ ), which is often the driving force in chemical reactions. Entropy is associated with water loss in the binding pocket caused by ligand binding[82]. Structural complex studies show that the CoA molecule can occupy various conformations in a protein-molecule complex. However, certain conformations seem to be preferred, which could be shown by NMR data[7].

Based on studies and NMR data, many conformations have been discovered, ranging from a strongly bent conformer as in 2-enoyl-CoA-hydratase (ECH) to an extremely extended form as in thiolase[9, 83]. It is noticeable in the behavior of CoA that the 3'-phosphate basically points towards the solvent and the adenine part is oriented towards the protein[9]. Experimental investigations showed that although the 3'-phosphate of coenzyme A points towards the solvent, in some cases it seems to be coordinated by conserved residues. In the example of ACP it could be shown that affinity is reduced and thus the binding energy significantly decreased when derivatives without this group were embedded[84].

#### **1.1.4 Catabolism**

Fatty acid metabolism occurs in the cells in three different essential ways.  $\alpha$ -Oxidation,  $\omega$ -Oxidation and  $\beta$ -Oxidation are oxidations in the cell organism whereby the fatty acids are split, and smaller metabolites are formed. The  $\alpha$ -oxidation is necessary as soon as a methyl group is substituted at the C3-position. This prevents the 3-hydroxyacyl-CoA dehydrogenase from degradation into a variant of the 3-keto form according to the  $\beta$ -oxidation[29, 85]. After successful  $\alpha$ -oxidation, a fatty acid with an asymmetric 2-methyl substitution at the carbon atom is obtained at the end of a four-step conversion. The  $\omega$ -oxidation is cytochrome P450 dependent and results in a monocarboxylated fatty acid at  $\omega$  or  $\omega$ -1 carbon[86]. In mammals, the main degradation path of fatty acids is the  $\beta$ -oxidation which can be divided into two areas: Peroxisomes as well as mitochondria. The  $\beta$ -oxidation occurs in a kind of complete cycle either in the peroxisomal plane or the mitochondrial plane. Finally, the fatty acid is shortened by two carbon atoms. Both subplanes of the cell are analogous in the execution of the  $\beta$ -oxidation and consist of

four substeps: I) oxidation/dehydrogenation, II) hydrogenation, III) NAD<sup>+</sup>-dependent dehydrogenation/oxidation, and IV) thiolitic cleavage[87].

Although it is basically the same reaction for peroxisomes and mitochondria, different enzymes have developed through evolution for the respective intermediate steps[37]. The distinction is made on a functional level. The  $\beta$ -oxidation in the mitochondria is based on electron transfer chains and oxidative phosphorylation. This in turn serves as energy production for the citric acid cycle in the cells. In contrast, peroxisomal  $\beta$ -oxidation ensures the release of the released electrons to molecular oxygen. However, this does not release energy that can be used further.

The degradation of the majority of short-chain, medium-chain, long-chain and very long-chain fatty acids occurs in mammals in the mitochondria by their  $\beta$ -oxidation enzymes[37]. Basically, peroxisomal  $\beta$ -oxidation enzymes metabolize more complex substrates, including (poly)unsaturated fatty acids, that are not or only slowly metabolized by mitochondrial enzymes. After completion of the  $\beta$ -oxidation in the peroxisomes, the products are transported to the mitochondria where they are further processed for the mitochondrial oxidation[37, 88, 89].

Mammalian multifunctional enzymes within the mitochondria and peroxisomes of some eukaryotes responsible for the  $\beta$ -oxidation are functionally similar but structurally different in detail compared to bacterial enzyme complexes. It has not yet been clarified why in biological systems enzymes act as multifunctional complexes while the same reactions are also carried out separately by different enzymes. However, the reason for the subdivision seems to be justified by the fact that certain reactions could be more efficiently and better controlled. This is due to the fact that these enzymatic complexes have a certain proximity of the active sites.

The **trifunctional enzyme (TFE) or multifunctional complex (MTP)** is embedded in the mitochondrial part. It shows 3(S)-hydroxyacyl-CoA dehydrogenase and 3-ketoacyl-CoA thiolase activity in the  $\alpha$ -subunit and  $\beta$ -subunit, respectively[90, 91]. The structure itself is formed by a  $\alpha_4\beta_4$ -complex, which in turn is made up of a dimer with two  $\alpha_2\beta_2$ -complexes (Figure 6). The catalytic activity with highest affinity shows the enzyme at chain lengths of 12 carbon atoms upwards. Whereby C10-C16 chain length is basically metabolized[91]. Compared with the peroxisomal counterpart i.e., the multifunctional enzyme type 1 (MFE-1) has a sequence identity of 65%[92], which was also

demonstrated by the preservation of the hydratase and dehydrogenase activity after dissociation of the  $\alpha$ -subunit from the complex. Interesting is however that for the preservation of the thiolase functionality of the  $\beta$ -subunit, the whole enzyme complex must be present[93]. Contradictory results were nevertheless obtained that the thiolase activity can also occur without the complete complex[94]. By sequence matching of remarkable 35% with the *Pseudomonas fragi* of the human MTP  $\alpha 4\beta 4$ -complex a homology model could be created. It could be shown that positively charged residues are found in the center of the complex, which are responsible for the attraction of negatively charged ligands (due to their binding to coenzyme A)[95]. Due to the lack of detection of  $\beta$ -oxidized intermediates in the extracellular fluid substrate channeling is assumed[96, 97]. The counterpart with high sequence identity as described above is located in the peroxisomal region. In mammals, this includes in particular two multifunctional enzymes: multifunctional enzyme type 1 (MFE-1) and multifunctional enzyme type 2 (MFE-2).

Both enzymes reflect the uniqueness of being the enzyme that catalyze the second and third stages of the  $\beta$ -oxidation of fatty acids. In the following chapter the structural properties of MFE-1 in particular are shown and described.

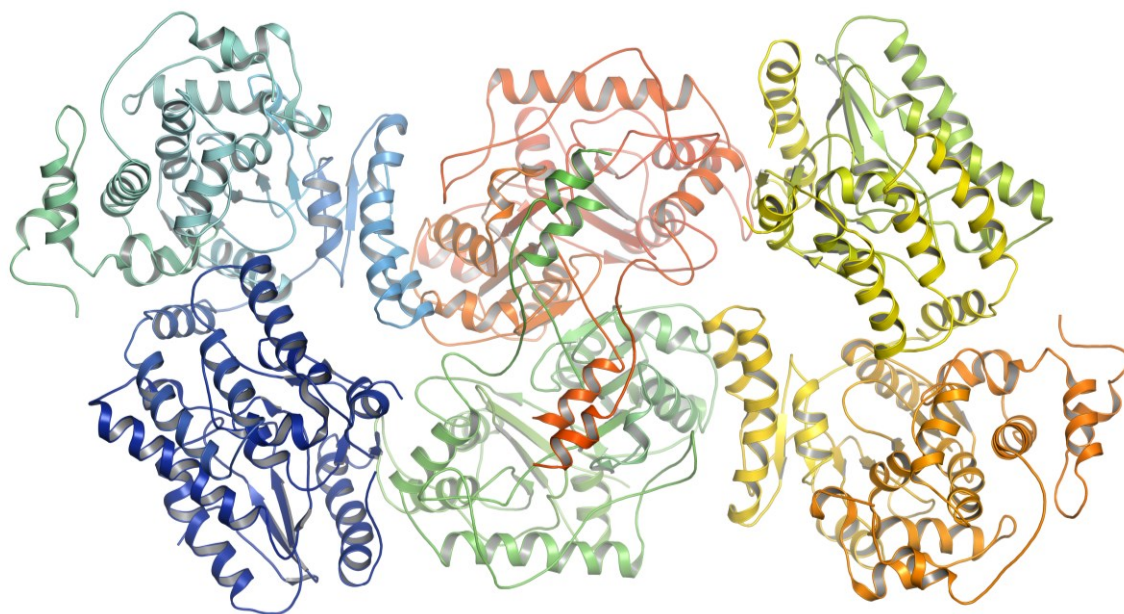


Figure 6: Overall view of TFE (PDB: 5ZQZ)[98] in cartoon representation (rainbow colored).

## 1.2 Multifunctional enzyme type 1 (MFE-1)

The multifunctional enzyme type 1 (MFE-1) belongs to the mammalian proteins that is abundant in the peroxisomes. Its catalytic activities as  $\Delta^3, \Delta^2$ -enoyl-CoA isomerase,  $\Delta^2$ -enoyl-CoA-hydratase-1, as well as (3S)-hydroxyacyl-CoA-dehydrogenase (Figure 7), makes it one of the most important enzymes in the  $\beta$ -oxidation of fatty acids[87, 99]. The first step of the MFE-1, *i.e.*, the isomerization, concerns the rearrangement of the double bond of the 3E-enoyl-CoA and the 3Z-enoyl-CoA substrate towards 2E-enoyl-CoA. This is followed by the hydrogenation which results in the product of the first active site, the 3S-hydroxyacyl-CoA molecule. Once 3S-hydroxyacyl-CoA is present in the second active site, oxidation to the 3-ketoacyl-CoA molecule finally takes place using NAD<sup>+</sup>. This is the final product of the MFE-1 reactions.

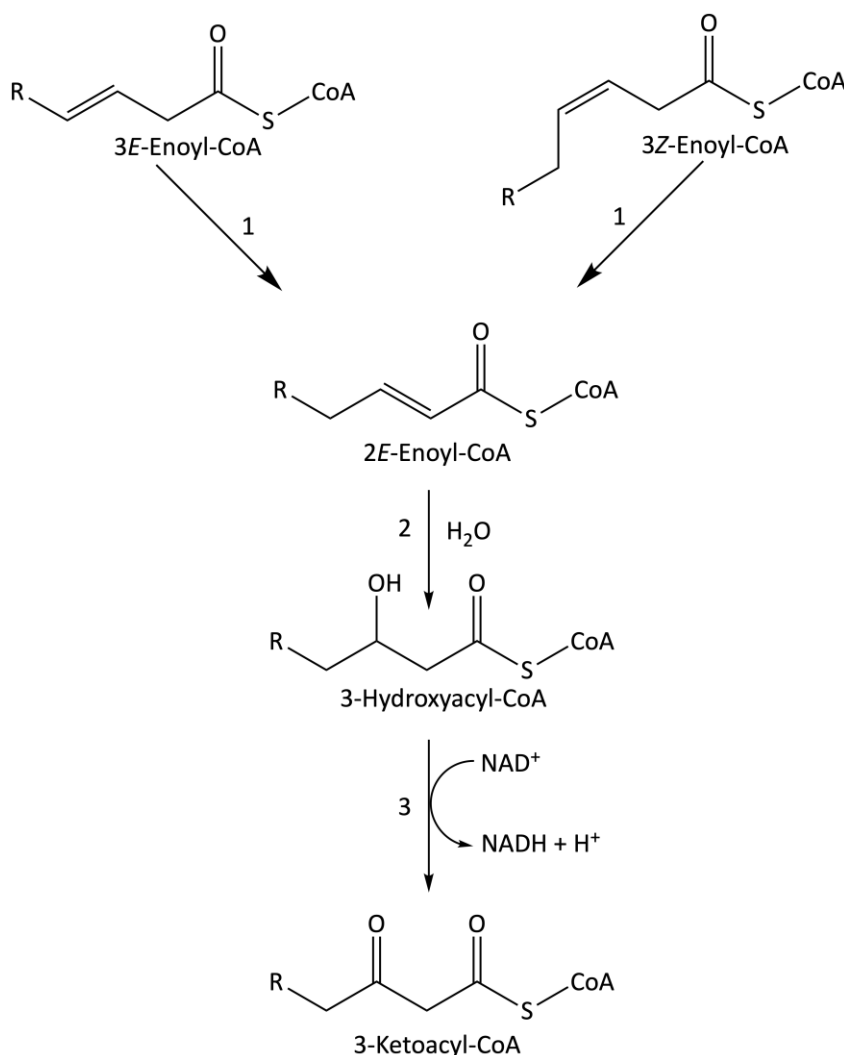


Figure 7: The reaction scheme of the three steps of the MFE-1. First the isomerization, second the hydrogenation and third the dehydrogenation.

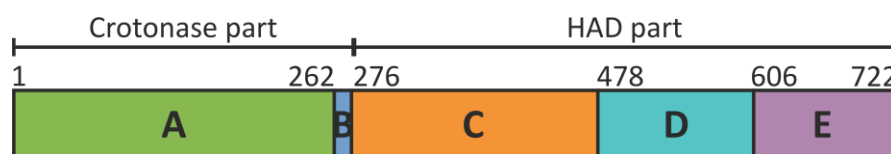


In contrast to its other multifunctional enzymes and representatives of fatty acid degradation, rat peroxisomal as well as mammalian peroxisomal MFE-1 are present as monomeric enzymes. They consist of 722 residues in total, with a basic overhang resulting in a pI of 9.3[99, 100]. The enzyme MFE-1 has two active sites. The first functions as an isomerase-hydratase combination (H/I) and is located in the N-terminal section. The second active site shows the functionality of a dehydrogenase and is located in the C-terminal part of the enzyme. Structural investigations of corresponding derivatives of the monofunctional enzymes such as isomerase, hydratase (also known as crotonases) or dehydrogenase showed in the sequence analyses a descent of two different superfamilies. The superfamilies can be divided into the crotonase superfamily and HAD superfamily, with the N-terminal part of MFE-1 belonging to the former and the C-terminal to the latter superfamily[101, 102].

The exact physiological role of MFE-1 has not yet been clarified. This is due to the fact that MFE-1 has an overlapping substrate and catalysis specificity with the multifunctional enzyme type 2 (MFE-2). However, it is structurally different in comparison[13, 87]. One reason for the physiological significance is the participation in an  $\alpha$ -methylacyl-CoA racemase-MFE-2 independent synthesis, in which bile acids are formed from 24S-hydroxyoxysterols[103]. Experimental investigations and data provide results that reveal MFE-1 to be involved in the  $\beta$ -oxidation of long-chain dicarboxylic acids. These results are consistent with the in vitro substrate specificity of MFE-1, showing that coenzyme A derivatives of linear as well as 2-methyl branched fatty acids including bile acids are processed with steroid units[104-107], whereby a high promiscuity makes high demands on the uniqueness of structural features of the different active sites. The N-terminal section of MFE-1 is of great interest, as the ability to catalyze two reactions in catalytic form successively in a single active site (hydratase and isomerase activity) can be considered unique[108]. In vitro the isomerase activity of rat MFE-1 could be shown and characterized[109]. By in vivo expression of rpMFE-1 in the  $\Delta$ eci-knock-out strain, the isomerase activity is functional[110] could be determined. Another outstanding aspect is that all other derivatives of the crotonase superfamily perform their function as monofunctional enzymes, while MFE-1 is multifunctional. However, it is important to note that the isomerase activity of monofunctional hydratase is about 1000 times lower compared to the monofunctional isomerase[111, 112]. In contrast, the isomerase activity of MFE-1 is about 10 times lower compared to its hydratase reaction[111, 112].

Probably, promiscuous properties related to the specificity of MFE-1 and its H/I activity were conserved during the evolution of MFE-1, whereas monofunctional derivatives such as 2-enoyl-CoA hydratase and 3,2-enoyl-CoA isomerase have only one conserved function either as hydratase or isomerase.

Data on kinetic experimental investigations of the MFE-1 indicate substrate channeling from the first active site to the second active site, since the intermediate is not released to the solvent[112, 113]. A good agreement of this assumption could be shown for structurally elucidated enzymes with multifunctional function. They show that their active sites seem to be connected by a tunnel system, through which a substrate is transported from one active site to the other by a kind of diffusion dependent mechanism, including solvent shielding[114-116]. Well-known examples for substrate channeling are tryptophan synthase and aldolase dehydrogenase. Both enzymes, transport smaller intermediates such as indole and acetaldehyde between the active sites via channeling[117, 118]. In comparison, the intermediate to be transported in MFE-1 is relatively large, polar and negatively charged. In general, the channel mechanism is still poorly understood, especially for this type of intermediates and enzymes. Only a small number of studies with *e.g.*, the dihydrofolate reductase thymidylate or enzymes of the  $\alpha_2\beta_2$  complex for the  $\beta$ -oxidation of fatty acids exist. The possibility of electrostatic interactions related to the type of tunnel system is also proposed. This is reflected by the bifunctional enzyme of Dihydrofolate reductase-thymidylate synthase, where a substrate-dependent diffusion for bacterial fatty acid oxidation complexes is assumed[95, 119, 120]. Previous studies on the structure of MFE-1 describe of five domains A to E (Figure 8).



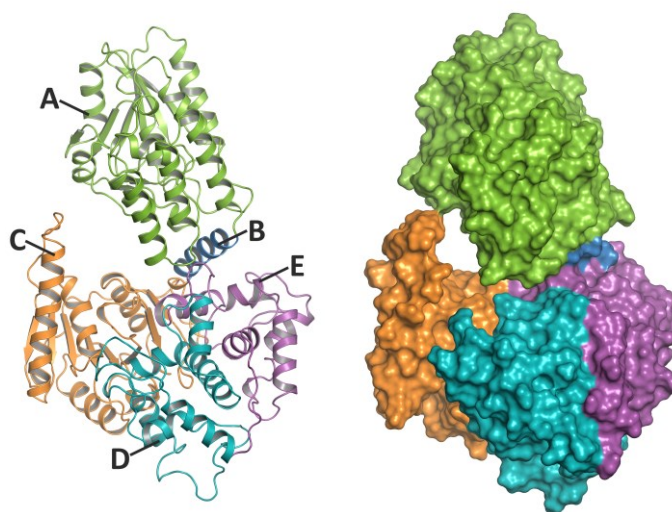
*Figure 8: The domain organization of the MFE1 protein. The domains A and B form the crotonase part. The B-domain is the helix-10/linker helix. Structurally it belongs to the HAD part, which includes the C, D and E domains.*

Figure 8 shows the modular structure of the MFE-1, which enables the possibility to divide the whole system into two evolutionary units (crotonase part and HAD part) and five structural units (domains A to E). In the N-terminal section of the MFE-1 are the domains A and B and in the C-terminal part are the domains C to E. Domain B is called linker helix, which is functionally bound to the N-terminal part of MFE-1, but structurally belongs to the C-terminal part. In general, however, the MFE-1 can be divided into three main subunits as shown in Figure 8: 1. domain A, 2. domain C and 3. domains B, D and E. The latter are connected by more flexible linkers. Structures of the entire MFE-1 crystal structure have already been described[121, 122]. Domains A and B form the crotonase part due to the crotonase folding. Furthermore, various crystal structures of crotonase enzymes have been published, such as  $\Delta^3, \Delta^2$ -enoyl-CoA isomerase as well as  $\Delta^2$ -enoyl-CoA-hydratase-1 and  $\Delta^{3,5}, \Delta^{2,4}$ -dienoyl-CoA isomerase with isomerase or hydratase activities[101, 123]. The C-terminal section of the MFE-1 belongs to the so-called HAD superfamily, where the HAD fold corresponds to a Rossmann-like structural fold, which additionally binds NAD (domain C). The domain D is a dimerization domain. The basis of this superfamily is the human mitochondrial 3S-hydroxyacyl-CoA-dehydrogenase enzyme, which is also called hmHAD[124]. Also known is the homologue of the porcine[125]. Of particular interest is that the crotonase fold (domain A) as well as the HAD fold (domains C and D) are also contained in the bacterial *Pseudomonas fragile* fatty acid oxidation multienzyme  $\alpha_2\beta_2$  complex (pfFOM)[95]. Due to the origin and the property to be a homolog of pfFOM, the  $\alpha$ -chains of MFE-1 and pfFOM have the same domain structure (Figure 8). In the pfFOM complex, the  $\alpha$ -chains form an  $\alpha_2$  dimer while the  $\beta$ -chains form an  $\alpha_2$  thiolase dimer. Structural similarity with MFE-1 and pfFOM can also be seen in the C-terminal section. Extended by domain E in MFE-1, both (C-terminal pfFOM and MFE-1) correspond to the HAD section. The domains D and E are directly related to each other. The evolutionary connection between the two domains has already been described[99, 102].

### 1.2.1 Overall structure

As described in the introduction 1.3, the MFE-1 structure is essentially composed of three structural units. Residue 1 to 261 refer to domain A followed by domain B (residues 262 to 276). Domain C is located from residue 277 to 477. The last two

domains D and E form the C-terminal part of MFE-1. The former ranges from residues 478 to 605 and the latter from 606 to 722 (Figure 9).



*Figure 9: Overall view of the MFE-1 (PDB: 3ZW8) in cartoon (left) and surface (right) presentation. Domain A is colored green, domain B blue, domain C orange, domain D cyan and domain E magenta.*

The secondary structure is illustrated by the amino acid sequence for the respective enzyme sections as shown by Wierenga *et al.*[121]. They show the designation of regions in MFE-1 and domain B, which acts as a linker helix and associated to domain E, is also marked.

It is also reported that the regions are bound to each other with different degrees of flexibility (Figure 10). This could lead to dynamic changes in the position of the domains in relation to each other[121].

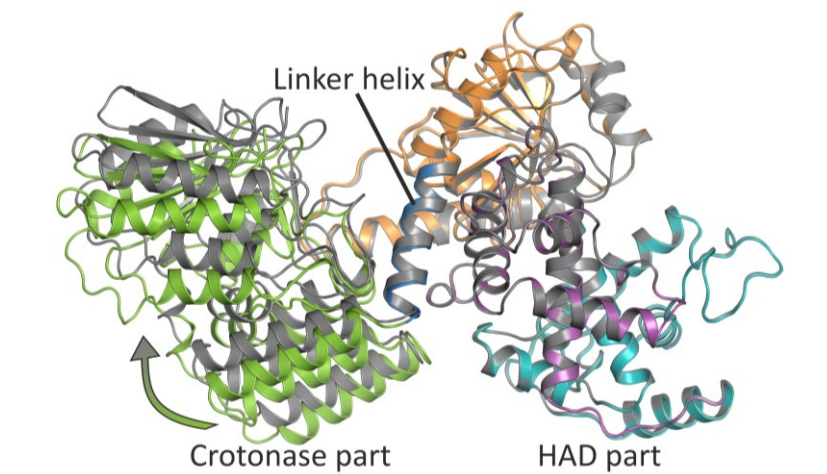


Figure 10: Overall view of MFE-1 (PDB: 3ZW8) in cartoon representation. Two arrangements of domain A (colored in green and gray) are shown (adapted from [99, 121]). Both conformations are facilitated through the hinge region (N-terminal part of domain B).

The N-terminal part of B domain is connected to the H/I section by interactions with the loops of the strands B2/B4. The C-terminal part of the linker helix is connected to the domain E by interactions with the EH2-helix and the loop between the EH2 and EH3 helices. The crystal structure analysis by Kasaragod *et al.*[121] showed that the domains have different relative conformations to each other. The location of the flexible connection between the crotonase folding and the HAD part is shown in Figure 10. The comparison between an Apo rpMFE-1-DH and rpMFE-1 also reveal a similar structural behavior[99, 121].

Further comparative crystal structure analyses by Kasaragod *et al.*[121] describe the main difference between the two structures as a joint movement of domain A in relation to domain C (Figure 10). The N-terminal part moves in a certain kind of rotation compared to the C-terminal part.

Thus, Kasaragod *et al.*[121] were able to uncover three main structural features of rpMFE-1: 1. a protruding CH2-helix, 2. a hinge region connecting the N- and C-terminal part, and 3. a kind of tunnel between active site I and II (Figure 9).

### 1.2.2 CH2-helix

The CH2-helix is located in the C domain of the HAD part. It is assumed that it has an interaction with domain A due to its protrusion. Because of this structural property, it contributes to the formation of the tunnel system within MFE-1. In pfFOM and hmHAD, however, this helix motif is a helix-turn-helix motif and not an extension loop ending in the bulk of the protein[121]. Due to the very high B-factor in this region of the MFE-1, a complete resolution of this position is difficult. However, further crystal structure analyses by Kasaragod *et al.*[121] reveal that this region or the tip of the helix shifts regularly but remains the same and is not subject to any conformational change. The structural differences that occur in connection with the CH2-helix are caused by the general change in position of the N- and C-terminal parts in relation to each other of the MFE-1.

### 1.2.3 Anchor region and tunnel

Based on crystal structure experiments and molecular studies of different constructs of MFE-1, Kasaragod *et al.*[121] showed that especially A domain of MFE-1 seems to occupy various orientations probably due to conformational changes of MFE-1. The N-terminal part represented by domain A approaches the C-terminal part (domains B to E) of the MFE-1. Finally, there are two different positions of domain A: close to the C-terminal part or remote from the C-terminal part. The motion is illustrated by the arrow shown in Figure 10. The rotational movement of domain A leads to a structural mechanism of both structural elements of MFE-1, which can be assumed as an opening and closing motion comparable to a scissor.

The coenzyme A is embedded within the N-terminus and carries the fatty acid. It is in contact with domain A and B. These contacts are probably salt bridges. Since B domain is also within the functional part of the crotonase folding, it is also called helix-10. The reason for this functional property is based on structural studies of homologous monofunctional enzymes of the crotonase superfamily. Helix-10 was also found to have salt bridges and hydrophobic van der Waals contacts with the various coenzyme A molecules[126]. Basically, the crystal structure analyses also showed that the apoprotein seems to be much more flexible, because so far MFE-1 can only be crystallized with a

ligand in the first active site. However, the relevance of this high flexibility could not be clarified yet[121].

The tunnel is essentially formed by the CH2-helix and four regions within the crotonase part (see Kasaragod *et al.*[121] and the linker helix. The part of the dehydrogenase region (C-terminus) for the tunnel is formed by the EH2-helix and different loop regions of the HAD part (Kasaragod *et al.*[121]). The tunnel consists of positively charged side chains facing inwards. Due to the positive charge excess, substrate diffusion along the charged surface from one active site to the other is assumed instead of a release to the solvent.

#### 1.2.4 Crotonase part and active site I

The characteristic feature of the crotonase family is the spiral folding of its units. Structurally, this results in a  $\beta$ - $\beta$ - $\alpha$  unit at the N-terminal part of MFE-1, followed by a spiral part and two crotonase  $\beta$ -strands[121]. As shown for human mitochondrial isomerase[126], the C-terminus of the classical crotonase fold consists of a total of four consecutive helices, the latter forming the helix-10 (Figure 11). In some members of the superfamily of crotonases, the helix-10 always forms the C-terminus[127]. Moreover, the helix forms important intramolecular interactions with the bound coenzyme A molecule and has a stabilizing effect[126].

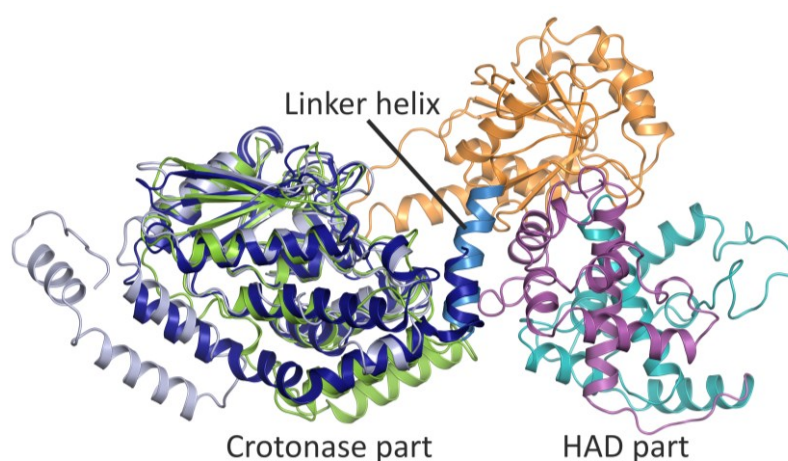


Figure 11: Structural comparison of the crotonase part of MFE-1 (PDB: 3ZW8, multicolored) with the superimposed isomerase (PDB: 1SG4, colored dark blue) and hydratase (PDB: 1DUB, colored light blue).

Although MFE-1 and its N-terminal section belong to the crotonase family, three major structural differences to classical crotonases have been shown by Kasaragod *et al.*[121], which are topological in nature with respect to the first differences. A significant difference concerns the B domain of MFE-1, which corresponds to helix-10. In comparable crotonase structures, the helix-10 is always poorly resolved due to a very high B-factor and the location at C-terminus.

In MFE-1, however, this helix is rather rigid and seems to interact with the E domain[121]. While trimerization is possible in the classical crotonase folding enzyme, this is also not possible in MFE-1 due to topological structural differences. Therefore, the crotonase part of the MFE-1 corresponds to a monomer instead of an oligomer like all other crotonase family members. Kasaragod *et al.*[121] also show a comparison between representatives of isomerases and hydratases by sequence analysis. The MFE-1 and its crotonase part has a sequential agreement of 32% with a rat hydratase while the H/I part of the pfFOM complex still has an agreement of 27%. Sequence analysis showed that for human mitochondrial isomerase and for yeast peroxisomal isomerase the agreement is less than 20%.

Looking into the detail, the sequence alignment reveals two conserved residues within the first active site, which are described to be important for the isomerization and hydrogenation reaction. Figure 12 shows these residues (two glutamates) which are important for the reactions in the first active site[121]. Both residues are catalytically important amino acids called glutamate 123 and glutamate 103, and the ligand is generally embedded and bound in a kind of ox-anion hole.



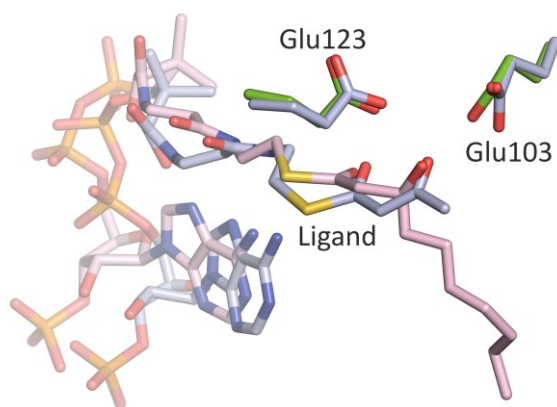


Figure 12: View of the first active site of MFE-1 with CoA and acetoacetyl fatty acid bound (represented as sticks) with the hydratase superimposed (PDB: 1DUB, colored light blue). Also shown are the two glutamates 103 and 123 as stick models.

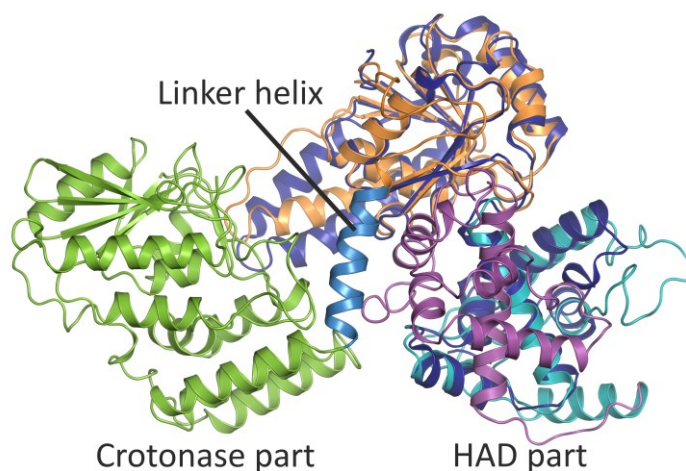
Furthermore, the crystal structure has a glycerol hydroxyl group between the two glutamates. As Kasaragod *et al.*[121] pointed out, this group represents a placeholder, which under physiological conditions should correspond to a water molecule. The water molecule in turn should serve as a catalyst and initiate the hydrogenation reaction (Figure 7). But it is not yet known how the active site I of MFE-1 carries out the isomerization and hydrogenation reactions in detail. However, smallest structural differences in the type of binding and interaction between 3-enoyl-CoA (isomerase reaction) and 2-enoyl-CoA (hydratase reaction) could be decisive[121]. So far it is assumed that glutamate 123 is responsible for the deprotonation during the isomerase reaction. While the interaction of the water molecule between both glutamates (103 and 123) seems to play a crucial role[128, 129].

The region for the so-called ox-anion hole can be found in loop2. However, structural differences can be found in the crotonase superfamily region in this loop, which leads to different substrate affinity and specificity[130].

### 1.2.5 HAD part and active site II

As already described for the hmHAD, the structural part of the domains C to E of the MFE-1 represents the C-terminus and the HAD fold[124]. Kasaragod *et al.*[121] also show for this structural part of the MFE-1 an agreement of 30% with hmHAD and pfFOM by

sequence analysis, while the highest agreement of 33% was found for rpMFE-1 with the monofunctional hmHAD.



*Figure 13: Structural comparison of the HAD part of MFE-1 (PDB: 3ZW8, multicolored) with the dehydrogenase (PDB: 1F14, colored dark blue). The dehydrogenase is superimposed with the HAD part of MFE-1.*

Structurally, the catalytic residues are highly conserved in all sequences. The dimerization unit of the hmHAD dimer and the rpMFE-1 monomer is also very similar, both show the same fold, whereas in the MFE-1 the domains D and E correspond to the dimerization domain of the hmHAD due to their high structural similarity (Figure 13).

As crystal structure analyses[124] have shown, the hmHAD structure basically occupies two main conformations. One corresponds to a kind of "open" form, which was shown for an unligated variant, while the other describes a "closed" form. The latter, however, has only been shown in a ternary bound (NAD<sup>+</sup> and acetoacetyl CoA bound) complex[124]. As in the N-terminal part of the MFE-1, the hinge region is located in the C-terminal. It is located between the domains C and D. Structural analyses by Kasragod *et al.*[121] could show, in accordance with the different conformations of the HAD part of the hmHAD, that the HAD part of the MFE-1 also seems to adopt "open" and "close" states. This is probably reinforced by the presence of NAD<sup>+</sup>[121, 124]. Kasaragod *et al.* could also show that the DH2-helix and its conformation ("open" or "closed") seems to play a major role in the substrate specificity.

### 1.3 Substrate channeling

#### 1.3.1 Basic concept

Tunneling or channeling is the transport of intermediates or substrates from one part (active center) of an enzyme (E1) after the first reaction without dissociation or further reactions up to another second active center of the same enzyme or of another enzyme (E2). It is an essential mechanistic process occurring in nature without releasing the intermediate into the solvent[131-134]. This means that the reaction products are translocated from one to another active center either within an enzyme or intermolecularly (*i.e.*, from one enzyme to another) (Figure 14).



*Figure 14: General process during substrate channeling. At first, the substrate is locked into the first active center of an enzyme (E1) and converted into an intermediate. Then via substrate channeling it is transported to the second enzyme or second active center of the same molecule (E2). There the final reaction product is released.*

This type of mechanistic process allows tunneling and channeling to occur within different multifunctional enzymes or multi-enzyme complexes closely related to their type. Tunneling has several advantages over the free diffusion of the reactants within the solvent, which the enzymatic systems thereby provide. The transport speed from one reaction site to another is significantly increased[135, 136]. Susceptible intermediates (chemically labile) are thus protected from side reactions with an aqueous environment[137]. Equilibria that are very unfavorable or could be reached are avoided and intermediates of competitive transformations of other enzymatic reactions are prevented[138-140].

Several forms of substrate tunneling for biological systems have been reported in the literature[131, 141-147]. These include purine, pyrimidine biosynthesis, and amino acid metabolism. They also deal with lipid metabolism, glycolysis, the tricarboxylic acid cycle, DNA replication, RNA synthesis and protein biosynthesis. However, the direct detection

of substrate channeling by experimental data is still difficult for biological enzymatic complexes. Some kinetic approaches to tunneling of intermediates of multifunctional enzymes and their complexes were investigated[148]. This also includes measurements of the associated competitive reactions, transfer times from one active center to the other or transient states kinetics. More detailed analyses and explanations can be found in the literature[131, 132].

However, the challenge to obtain experimental data therefore leaves many fundamental questions unanswered for most systems. Nevertheless, examples in the literature such as tryptophan synthetase or other members of the well-known fatty acid  $\beta$ -oxidation multienzyme (FOM) family, which are shown in the following section.

### 1.3.2 Examples

One of the most well-known representatives of enzymes in the field of substrate channeling is the tryptophan synthase. The tryptophan synthase (Figure 15) catalyzes two reactions of the biosynthesis of L-tryptophan[149-153]. This enzyme exists in bacteria as an  $(\alpha\beta)_2$  complex, where each  $\alpha$  unit is responsible for the cleavage of indole-3-glycerolphosphate. The  $\beta$  unit harbors a pyridoxal phosphate moiety responsible for the formation of L-tryptophan[149, 151, 153, 154].

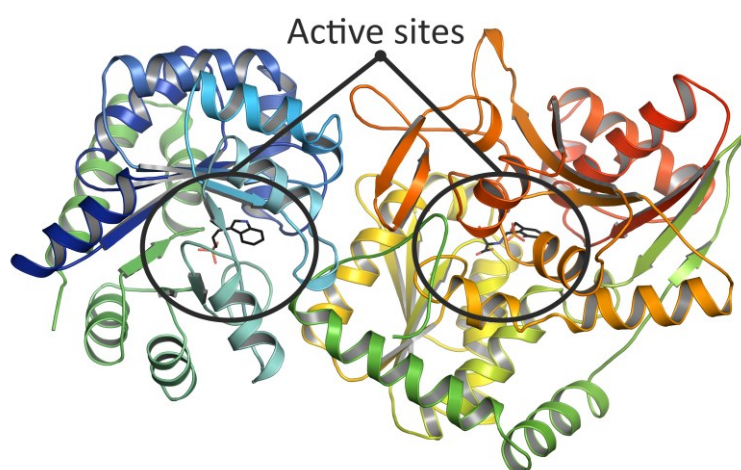


Figure 15: Overall view of a heterodimer of tryptophan synthase (PDB: 2TRS, cartoon representation, rainbow colored). Ligands are represented as black sticks. The two active sites are highlighted with black circles[155].

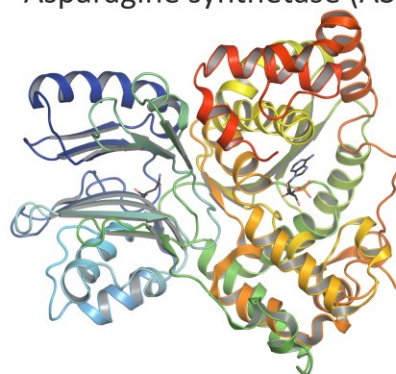
To date, experimental data show that during L-tryptophan synthesis the intermediate indole is not released into the solvent but is transported between the two active centers via tunnels. The distance between the two active centers is 25 Å, which is a considerable length[117, 156-159] (Figure 15). Additional investigations show that the tunnel between the two centers is mainly hydrophobic and has a dimension to transport several indole molecules. Furthermore, subsequent structural elucidation showed that any conformational changes are induced by ligands and are associated with allosteric communication between the two active centers for channeling. Therefore, the tryptophan synthase is one of the best-known representatives in the field of substrate channeling[155, 160-163].

In the past, tryptophane synthase was considered as a unique example of these functional systems[114], but until today more representatives could be elucidated. Another example is the carbamoyl phosphate synthetase (CPS, Figure 16), which is one of the biologically most important multifunctional enzymes. CPS is responsible for the catalysis or formation of carbamoyl phosphate from bicarbonate, glutamine and two molecules of magnesium ATP[164-169]. Like the tryptophan synthase, CPS has a remarkable structural property, that of the distance of the active sites to each other within the heterodimer. The active site of the small subunit is about 45 Å away from the active site in the carboxyphosphate domain, which forms the large subunit and is about 35 Å away from the carbamoyl phosphate motif's active center. In total, this results in a tunnel with a length of about 80 Å. Since the intermediates are unstable, they are transported via tunneling within the whole protein complex[170-174].

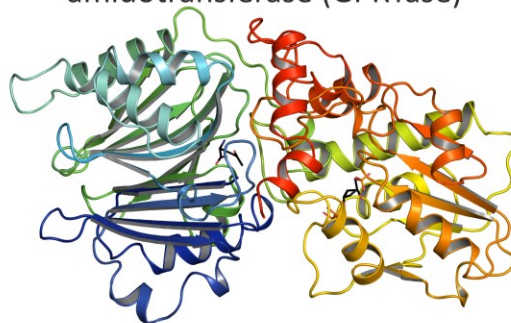
Carbamoyl phosphate synthetase (CPS)



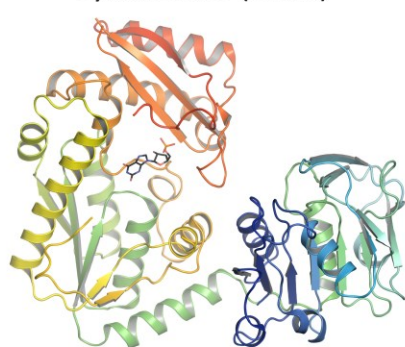
Asparagine synthetase (ASB)



Glutamine phosphoribosylpyrophosphate amidotransferase (GPRTase)



Guanosine monophosphate synthetase (GMP)



Thymidylate synthase-dihydrofolate reductase (TS-DHFR)

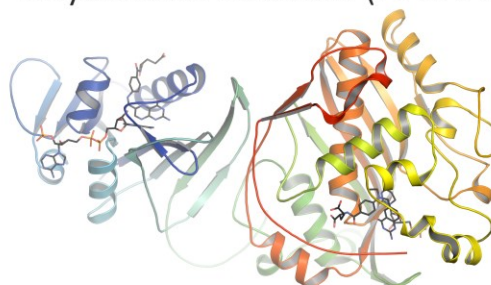


Figure 16: Overall view of selected representative enzymes for substrate channeling. CPS (PDB: 1JDB)[171], GMP (PDB: 2YWC)[175], ASB (PDB: 1CT9)[176], GPRTase (PDB: 1ECC)[177, 178] and TS-DHFR (PDB: 1QZF)[179] are rainbow colored and represented as cartoon. Ligands are shown as black sticks.

Further examples for substrate channeling are bifunctional enzymes represented by guanosine monophosphate (GMP) synthetase[175, 180-182], formiminotransferase cyclodeaminase (FT-CD)[183-188] and the asparagine synthase B (ASB)[176, 182, 189-193] (Figure 16). Additionally, enzymes of the de novo synthesis like the glutamine phosphoribosylpyrophosphate amidotransferase (GPRTase)[177, 194-204] and the combination of thymidylate synthase (TS) and dihydrofolate reductase (DHFR)[186, 205-216] are also representatives for substrate channeling (Figure 16). Moreover, other



known examples are enzyme complexes such as pyruvate dehydrogenase, 2-oxo-glutarate dehydrogenase, lumazine synthase, riboflavin synthase complex belonging to the 2-oxo acid dehydrogenases[217-231] (Figure 16).

### 1.3.3 Multifunctional enzyme type 1 (MFE-1)

A special case of multifunctional biological systems is the multifunctional enzyme type 1 (MFE-1). Like other representatives of the group it consists of an ( $\alpha\beta$ ) unit[232]. However, unlike the other representatives MFE-1 is not present as an oligomer, but as a monomer with two active sites and still catalyzing three reactions in parallel[102]. Both active sites are 40 Å apart, which are connected by a tunnel (Figure 9). Along the tunnel are several positively charged amino side chains (lysine and arginine) present. Due to the negative charges on the co-substrate, the co-enzyme A, and the structure of the MFE-1 itself, it is assumed that a kind of substrate crawling is facilitated and preferred by the interaction of the positive and negative charges[115, 173, 233]. In addition to that, kinetic experimental data show that substrate channeling occurs between the two active sites[112, 113].

Considering that, based on previous studies Wierenga *et al.*[121] propose that the domain positioning of MFE-1 could play a supporting role for substrate channeling. Thus, in particular, the relative position of the crotonase part (domain A) to the dehydrogenase part (Figure 10) in MFE-1 seems to be crucial. Further X-ray structure analyses comparing the crystal structure of MFE-1 to related enzymes[95, 234] show dynamic domain rearrangements. The observed domain flexibility is therefore understood to be an essential key feature of these related enzymes for substrate channeling[95, 234]. Due to the structural similarity of these enzymes to MFE-1[121], this flexibility is assumed to occur accordingly in MFE-1 for instance facilitated by the hinge region (between domain A and B). Hence, this flexibility is considered a possible key feature for substrate channeling in MFE-1.

Despite this knowledge, substrate channeling, its trigger and mechanism is still rarely understood at the molecular level[235].





## 2 Methods

### 2.1 Classical molecular dynamics simulation

#### 2.1.1 Evolution

Molecular Dynamic Simulations (MD) first described in literature[236] is a computer assisted technique which studies the movement of molecules and atoms in a physical manner. For a certain period of time the dynamic behavior of a whole system is represented by interactions and motions of the atoms of the system. For the trajectories of N particles Newton's equations of motion are solved.

$$F = m \cdot a \quad (1)$$

where F is force, m is mass and a is acceleration.

Over the years, the MD simulations have been improved to be used for biological systems such as proteins or even systems embedded in membranes[237]. This was shown in 1977 by McCammon *et al.*[238] who performed one of the first simulations for a biological molecule (pancreatic inhibitor). They used an X-ray crystal structure of a biological system and performed dynamics calculations. They calculated the fluctuations of the average structure. The molecule consisted of 500 atoms and the simulation was done in vacuum with a total length of 9.2 ps.

This fundamental work helped to shift the view away from rigid structures to dynamical and more flexible systems. Since then, the research field and the method has been enhanced over years. Current simulations usually range from several hundreds of atoms to large macromolecular systems (>100,000 atoms) with biological relevance[239-242].

Simulations of DNA were first done in 1986[243] contributing to elucidate DNA via NMR structure comparison. While 1,200 water molecules are common in the present, it was a huge development in the work of modelling biomolecules. Another milestone provided by Duan *et al.* was the first 1  $\mu$ s simulation of a system in explicit solvent[244]. With this they have looked into detail of protein folding of Villin, an Actin binding protein[245]. Because of the usage of parallel architecture, simulations of ca. 500  $\mu$ s with explicit

solvent could be done. With this, it was demonstrated that a very detailed view of protein folding can be obtained from Molecular Dynamics Simulation. Another milestone was published in 2006: Freddolino *et al.* [246] have firstly studied the whole viral system of the Satellite Tobacco Mosaic virus. Thereby they showed that in the absence of RNA the protein capsid of the virus is much less stable. These were an agreement of previous results concerning the capsid.

## 2.1.2 Principles of classical MD simulations

### 2.1.2.1 Integration algorithms

All classical molecular dynamics algorithms are based on Newton's equation of motion (1). The software packages like CHARMM, Amber, Gromacs and NAMD integrate this equation forecasting the position of an atom in future point of time [247-250]. Solving this equation is not trivial and several algorithms are used to overcome this problem. It is assumed that the position of an atom at a certain timestep can be approximated via a Taylor's series expansion (2), (3), and (4) [251]:

$$x(t + \Delta t) = x(t) + \Delta t \cdot v(t) + \frac{1}{2} \cdot \Delta t^2 \cdot a(t) + \dots \quad (2)$$

$$v(t + \Delta t) = v(t) + \Delta t \cdot a(t) + \frac{1}{2} \cdot \Delta t^2 \cdot b(t) + \dots \quad (3)$$

$$a(t + \Delta t) = a(t) + \Delta t \cdot b(t) + \frac{1}{2} \cdot \Delta t^2 \cdot c(t) + \dots \quad (4)$$

$x$ ,  $v$  and  $a$  represents the positions, velocities and the accelerations of each particle of the system. The Taylor's expansion is truncated after the third term. There are several used integration algorithms *e.g.*, Verlet-Algorithm and the Leap-Frog-Algorithm. The commonly used integration form of molecular dynamics is the Velocity-Verlet-Algorithm [252]. It is derived with the sum of the Taylor expressions for the coordinates at certain time steps ( $t + \Delta t$ ). The advantage of this algorithm is the calculation of position, velocities and accelerations at the certain time step:

$$x(t + \Delta t) = x(t) + \Delta t \cdot v(t) + \frac{1}{2} \cdot \Delta t^2 \cdot a(t) \quad (5)$$

$$v(t + \Delta t) = v(t) + \frac{1}{2} \cdot \Delta t^2 \cdot [a(t) + a(t + \Delta t)] \quad (6)$$

The calculation of the velocities  $v(t+\Delta t)$  needs acceleration values at  $t$  and  $t+\Delta t$ , therefore two calculations have to be done. This means, first the positions at  $t+\Delta t$  are calculated and then the velocities at  $t+\Delta t$  using:

$$v(t + \Delta t) = v(t + \frac{1}{2} \Delta t) + \frac{1}{2} \cdot \Delta t \cdot a(t + \Delta t) \quad (7)$$

#### 2.1.2.2 Statistics

Molecular dynamics provide much information on a microscopic level. Therefore, an important goal is to convert this into a macroscopic level. For this, the consideration of temperature and pressure during molecular dynamics simulations is needed. For this, time independent statistical averages are used. They are defined as follows:

Thermodynamic/macroscopic state:

Defined by parameters which describe all properties thermodynamically of the system. For example, the temperature  $T$ , pressure  $P$  and number of atoms  $N$ . The rest of the properties are derived from other thermodynamic equations.

Ensemble:

Ensembles are a set of macroscopic observables (e.g. temperature, internal energy, volume, or pressure). Frequently used ensembles for molecular dynamics are NVE, NVT and NPT (with  $E$  = total energy,  $V$  = volume,  $P$  = pressure,  $T$  = temperature).

The idea to get the conversion from the microscopic to the macroscopic level bases on the Ergodic Hypothesis[253]. It states that the average of time is equivalent to the ensemble average.

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} \quad (8)$$

with  $A$  representing the measured observable.

This means, the goal is to conduct a long simulation so that the system can go through a huge number of conformations which are then treated like a subset of a specific ensemble.

One of the simplest types of ensembles is the NVE. The number of atoms (N), the volume (V) and the energy (E) is constant through the whole simulation. This ensemble is known as a microcanonical ensemble. The pressure and temperature are allowed to vary during the molecular dynamics run.

In contrast to that, in case the number of atoms, the volume and the temperature is kept constant, this results in a NVT ensemble. These ensembles use thermostats to keep the temperature constant. A method to maintain the temperature is scaling the velocities of the atoms. A known thermostat is the Berendsen[254] which scales the velocities whereas the rate of temperature caused changes stays proportional to temperature differences. This keeps the system at a specific temperature rather than the starting temperature. But an improved thermostat of NVT ensembles is the so-called Nosé-Hoover algorithm[255, 256] modelling heat exchange with the system and a bath. By adding degree of freedom, the velocities and temperature can be reduced. Another example for this kind of algorithm was developed by Andersen[257] where the velocity of a random atom is replaced by one of Maxwell-Boltzmann distribution.

A generally used ensemble is the NPT. It is an isothermic-isobaric ensemble keeping the system at a constant pressure and temperature. For this, similar to the NVT ensemble a thermostat is needed in addition to a barostat. Several barostats are known to work similar to the thermostats. Size scaling of the cell and the atomic coordinates is part of the Berendsen barostat[254] to keep a constant pressure. Different systems use the Nosé-Hoover thermostat to have effects on the pressure. Another commonly used barostat and thermostat for the NPT ensemble is the Langevin-Thermostat/-barostat. Langevin method simulates interactions with a solvent. For this, two terms are added to Newton's equation of motion (friction term for the movement through a solvent and a random term for the energy from collisions/interactions with solvent molecules)[258].

### 2.1.3 Force fields

#### 2.1.3.1 Principles of molecular mechanics force fields

The energy of a system treated with molecular mechanics is calculated on the nuclei motion while the electrons are not explicitly regarded. This bases on the Born-Oppenheimer[259] approximation which also neglects the motion of the electrons causing a remarkable difference in the mass of the system.

This means, when the electronic properties are not crucial to consider, molecular mechanic methods are commonly used to calculate molecular dynamics. In general, a force field consists of several different terms which are calculated differently depending on the software packages which use different force fields (e.g., CHARMM or AMBER).

The principal equation of the force field can be separated into two parts:

$$V_{total} = V_{bonded} + V_{nonbonded} \quad (9)$$

with  $V$  representing the total potential energy which is calculated as a sum of bonded and non-bonded interaction energies.

The bonded potential energy  $V_{bonded}$  consists of three terms:

$$V_{bonded} = V_{bonds} + V_{angles} + V_{torsion} \quad (10)$$

with  $V_{bonds}$  refers to the bond stretching,  $V_{angles}$  is the bond angle potential and  $V_{torsion}$  represents the bond torsion potential.

#### 2.1.3.2 Bonded potential

##### **Bond length:**

In general, the Morse potential (Figure 17) shows the curve of a typical shape of a bond energy.

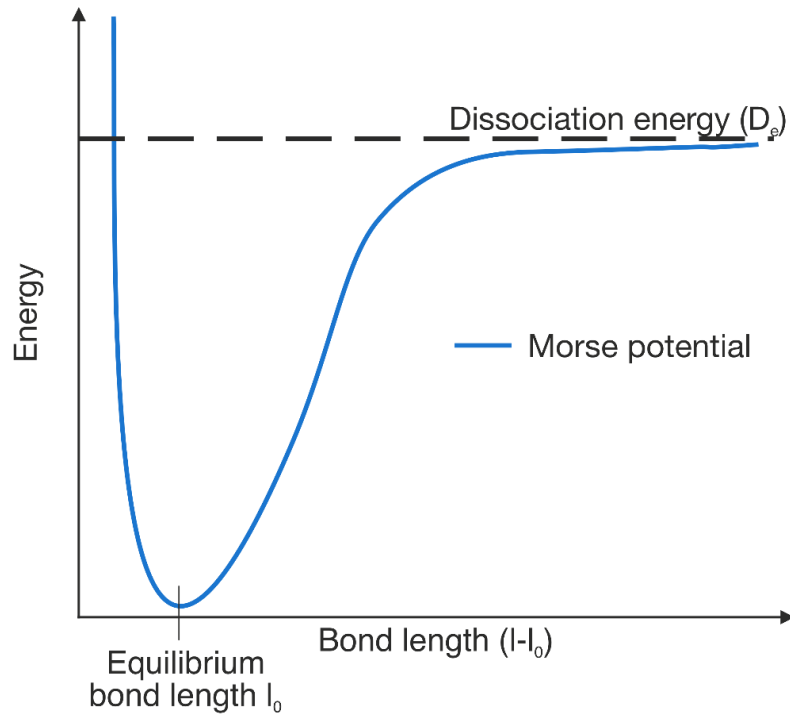


Figure 17: Schematic representation of the Morse potential (blue). The dependence of an arbitrary electronic potential on the nuclear bond distance of an indeterminate molecule is shown. The equilibrium distance  $l_0$  in the potential well and the dissociation energy ( $D_e$ ) represented by the dashed line are plotted.

The Morse potential describes the dependence of the potential energy of a diatomic molecule upon the bond length ( $l$ ) of the considered atoms. Here  $l_0$  is the equilibrium distance between the two atoms and  $D_e$  represents the dissociation energy at which the bond between the atoms is broken. The harmonic potential approximation (Figure 18, (11)) is used which is a consequence of Hooke's law.

$$V_b(l) = \frac{1}{2} \cdot k_l \cdot (l - l_0)^2 \quad (11)$$

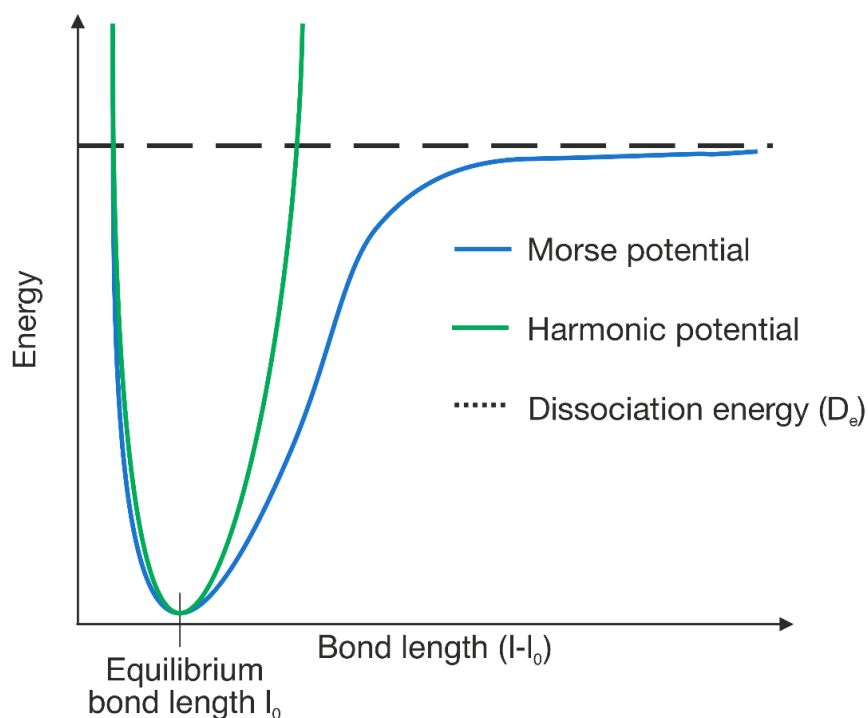


Figure 18: Schematic representation of the Morse potential (blue) compared with the harmonic oscillator potential (green). The equilibrium distance  $l_0$  in the potential well and the dissociation energy ( $D_e$ ) represented by the dashed line are plotted also.

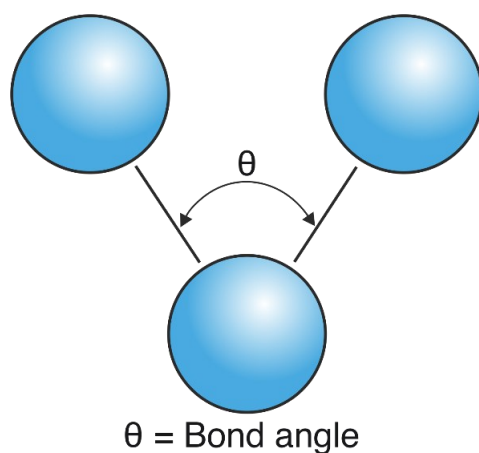
Figure 18 shows that the approximation to the shape of the Morse potential is in good agreement at low energies but shows anharmonicities for higher energies.

### Bond angles:

The bond angle of a triatomic molecule is illustrated in Figure 19. The angle is described by the harmonic potential  $V_{\text{Angle}}$ :

$$V_{\text{Angle}}(\Theta) = \frac{1}{2} \cdot k_{\Theta} \cdot (\Theta - \Theta_0)^2 \quad (12)$$

with  $k_{\Theta}$  representing the force constant,  $\Theta$  the angle between the atoms and  $\Theta_0$  the equilibrium angle. The equilibrium angle depends on the chemical properties of the compound.



*Figure 19: Schematic representation of the bond angle  $\theta$  of a triatomic arbitrary molecule. The blue spheres represent the three atoms of the molecule that are at an angle (bond angle) to each other. The angular size varies depending on the type of molecular composition.*

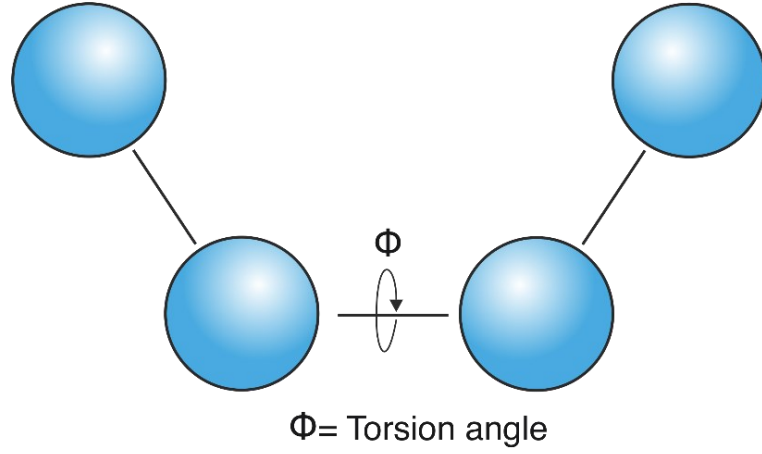
#### **Bond torsions:**

Torsional angles are defined as the plane formed by four covalently connected atoms[260] (Figure 20). These torsions are given by a cosine series expansion in the following form:

$$V_{Torsion}(\Phi) = \frac{1}{2} \cdot k_{\Phi} \cdot [1 + \cos(n\Phi - \delta)] \quad (13)$$

with  $k_{\Phi}$  representing the force constant,  $n$  the periodicity and  $\delta$  the phase. The periodicity and the force constant are varied to gain minima positions.





*Figure 20: Schematic representation of the torsion angle  $\Phi$  of a four-atom molecule. The blue spheres represent the four atoms of the molecule that are at an angle (torsion angle) to each other. The angular size varies depending on the type of molecular composition.*

### 2.1.3.3 Non-bonded potential

#### Electrostatic Interactions:

The electrostatic interactions are described via the Coulomb potential:

$$V_{Coulomb}(i, j) = \frac{q_i \cdot q_j}{4 \cdot \pi \cdot \epsilon_0 r_0} \quad (14)$$

with  $r_0$  being the distance between the charges  $q$  of particles  $i$  and  $j$ .  $\epsilon_0$  is the electrical permittivity.

#### Van der Waals Interactions:

The van der Waals interactions represented by attractive and repulsive interactions of the corresponding atoms  $i$  and  $j$  are defined with the following equation:

$$V_{vdW}(r) = 4\epsilon \cdot \left[ \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] \quad (15)$$

where  $\sigma_{i,j}$  is the equilibrium distance between the atoms  $i$  and  $j$  (the atoms do not attract or repel each other),  $r_{i,j}$  is the distance between the atoms  $i$  and  $j$  and  $\epsilon$  is the depth of the potential well.

This is the common form of the Lennard-Jones potential.

All these terms together (equations (11) – (15)), the general force field formula which is used for classical molecular dynamics runs can be written as:

$$\begin{aligned}
 V_{total} = & \sum \frac{1}{2} \cdot k_l \cdot (l - l_0)^2 \\
 & + \sum \frac{1}{2} \cdot k_\theta \cdot (\theta - \theta_0)^2 \\
 & + \sum \frac{1}{2} \cdot k_\phi \cdot [1 + \cos(n\phi - \delta)] \\
 & + \sum \sum \left( 4\epsilon \cdot \left[ \frac{\sigma_{i,j}}{r_{i,j}} \right]^{12} - \left[ \frac{\sigma_{i,j}}{r_{i,j}} \right]^6 \right) + \left( \frac{q_i q_j}{4\pi\epsilon_0 r_0} \right)
 \end{aligned} \tag{16}$$

## 2.2 Accelerated MD simulations

Classical molecular dynamics is used to sample the conformational space of a system by integration of Newton's equation of motion. A complete sampling of the potential energy landscape is needed to calculate the thermodynamic and kinetic properties in an appropriate manner. For most biological systems the timescale is limited to a nanosecond timescale. Since biological systems have multiple landscape energy minima the system is trapped in energy wells due to high free energy barriers during the simulation. Therefore, for adequate sampling a transition from one state to another and thus to overcome the energy barrier has to be facilitated. For this, different computational approaches such as replica exchange[261], self-guided MD[262], umbrella sampling[263] and others are used.

Early work of Voter[264, 265], Steiner *et al.*[266] and Rahman and Tully[267] showed an approach which does not demand the knowledge of the location of either the energy barriers or the corresponding energy wells. They also altered the computational time of the biological system spending in the energy minima well.

Bases on this work and by adding a bias potential  $\Delta V$  to the true potential, Hamelberg *et al.*[268] developed and implemented a new promising and efficient approach, the accelerated Molecular Dynamics (aMD).

The general idea is given by Figure 21.

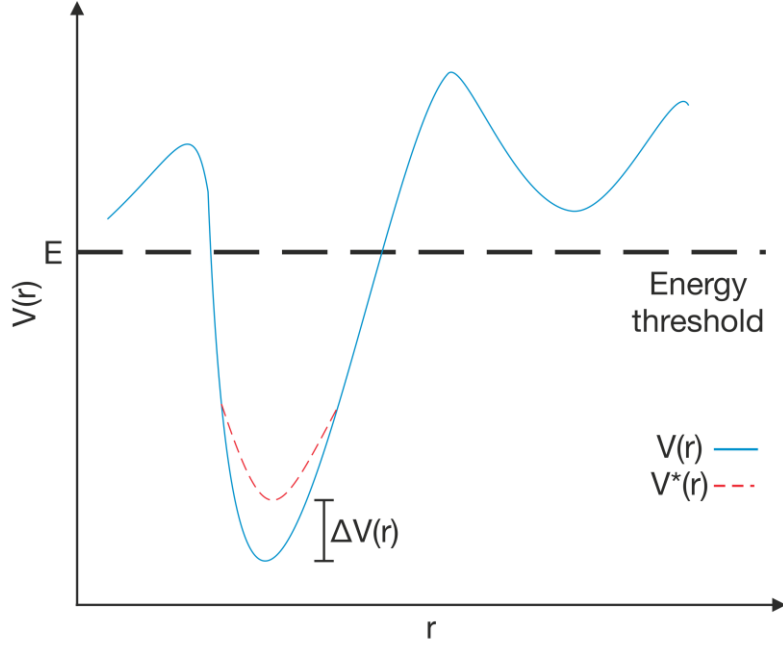


Figure 21: Schematic representation of an arbitrary chosen normal potential  $V(r)$  (blue curve), the modified potential  $V^*(r)$  (red dashed line) after adding a boost (biased) potential  $\Delta V(r)$  and the energy threshold (black dashed line).

The boost potential  $\Delta V(r)$  is defined in a way that it is applied when the original potential is below the certain threshold  $E$ . The modified potential is then described via:

$$V^*(r) = V(r) + \Delta V(r) \quad (17)$$

If the modified potential is larger than  $E$ , it is set equal to the true potential. This means that the new potential becomes:

$$V^*(r) = \begin{cases} V(r), & V(r) \geq E \\ V(r) + \Delta V(r), & V(r) < E \end{cases} \quad (18)$$

$$(19)$$

Conventional molecular dynamics extensively samples conformations around a local minimum but is inadequate elsewhere in the energy landscape. The modified potential facilitates the capability of escape from potential wells. The higher escape rate leads to a non-linear timescale:

$$\Delta t_i^* = \Delta t e^{[\beta \Delta V(r(t_i))]} \quad (20)$$

with  $\Delta t$  representing the unmodified timestep of the normal simulation time and  $\beta$  is the product of the Boltzmann constant  $k_B$  and the temperature  $T$ . This equation allows to modify the time scale at each step as a function of  $\Delta V(r)$ . The total statistical simulation time is thus given by:

$$t^* = \sum_i^N \Delta t_i^* = t \langle e^{[\beta \Delta V(r(t_i))]} \rangle \quad (21)$$

with  $N$  representing the total number of steps during an MD simulation. This means if  $\Delta V$  gets high then the boost factor gets very high also which would cause less efficient simulated energy wells.

For describing the boost potential, several approaches were developed[**262, 263, 269, 270**] and Hamelberg *et al.*[**268**] defined the boost potential as follows:

$$\Delta V = \frac{E - V(r)^2}{\alpha + (E - V(r))} \quad (22)$$

with  $\alpha$  representing a tuning parameter for the boost. This determines how deep the wells would be. It also shows that  $\alpha$  has to be chosen in an appropriate way because if  $\alpha$  is zero the modified potential would be even to the true potential. Moreover, the choice of  $E$  and  $\alpha$  is crucial for the accelerated molecular dynamics approach (Figure 22).

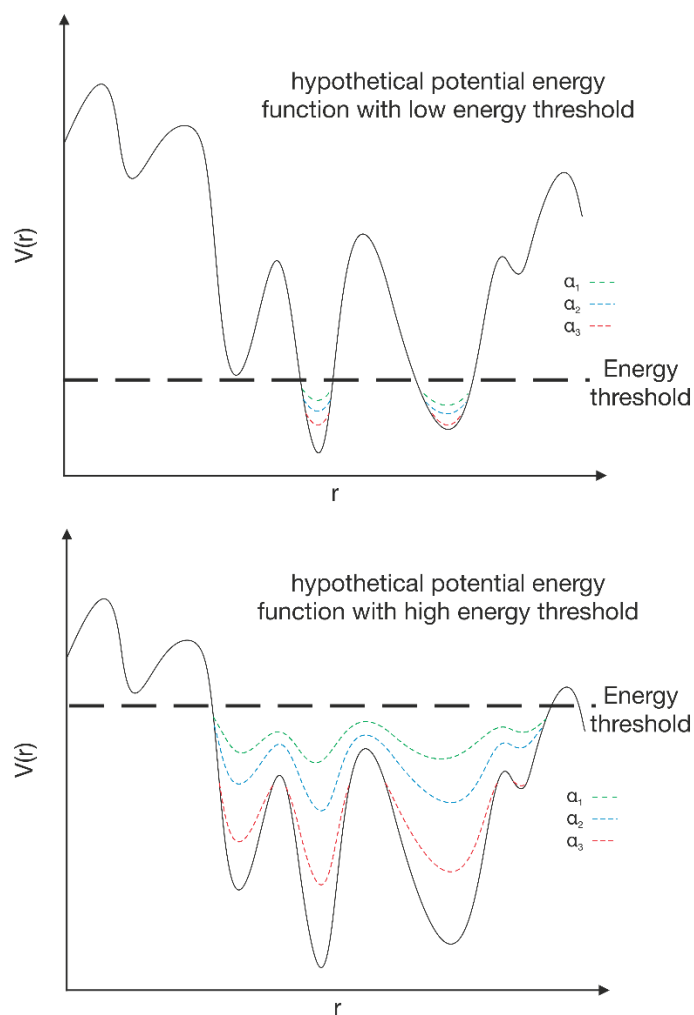


Figure 22: **Top:** Schematic representation of a hypothetical potential energy function (black curve) with a low energy threshold (dashed black line). Different bias potentials are plotted representing different values of  $\alpha$  (dashed green, blue and red curves). **Bottom:** Schematic representation of a hypothetical potential energy function (black curve) with a high energy threshold (dashed black line). Different bias potentials are plotted representing different values of  $\alpha$  (represented with colored dashed lines:  $\alpha_1$  is green,  $\alpha_2$  is blue and  $\alpha_3$  is red).

Estimation of  $E$  is essential: First, a classical conventional molecular dynamics simulation has to be conducted and out of this, an average of the total potential energy is calculated. This average energy represents the energy threshold  $E$ . Just arbitrary chosen  $E$  can significantly raise the chance that  $E$  is less than  $V_{\min}$  and the simulation run would proceed like a normal molecular dynamics simulation run. Thus, the choice of  $E$  should always be  $>V_{\min}$  with the magnitude depending on the desired acceleration.

The approach of accelerated molecular dynamics is done via all-atom force field. The fact that conformational changes in proteins involve mostly the torsion term of the force field, the accelerated molecular dynamics approach is applied to the sum of the dihedral torsions instead. Therefore, it is given by:

$$\Delta V_{dihedral}(r) = \frac{(E_{dihedral} - V_{dihedral}(r))^2}{\alpha} + (E_{dihedral} - V_{dihedral}(r)) \quad (23)$$

with

$$V^*(r) = \begin{cases} V_0(r) + V_{dihedral}(r), & V_{dihedral}(r) \geq E \\ V_0(r) + V_{dihedral}(r) + \Delta V_{dihedral}(r), & E < V(r) \end{cases} \quad (24)$$

$$(25)$$

with  $V_0(r)$  representing the sum of interaction potential without dihedral and the non-bonded potential energies.

## 2.3 Normal mode analysis

### 2.3.1 Background

Modeling of biological systems require powerful approaches and tools for dynamical calculations. The dynamics is defined as a time-dependent change of the structure, including fluctuations governing biological functions of a protein[271]. This structural flexibility was used to successfully describe protein functions at all levels of the protein. Starting at the level of an amino acid side chain and ending at a reorientation of a large domain.

During the last years numerous studies have been performed to describe structures qualitatively and their link to the functionality of a system. To complement experimental data with dynamical models, computational approaches (*e.g.*, Molecular Dynamics) were developed.

Although approaches (*e.g.*, aMD) are developed, performing classical molecular dynamics is still computational expensive for a sufficient sampling of large motions on an

appropriate timescale. Therefore, an efficient method to calculate large motions is the Normal Mode Analysis (NMA).

In general, normal modes can be described as a pattern of motion. Within this motion all parts of the biological system move with the same frequency and phase relation.

Normal mode analysis uses the Elastic Network Model (ENM). It models atomic motions, with each mode associated to frequencies, as a coupled harmonic oscillator[272-275]. It is shown that modes of low frequencies correlate to the protein functionality[276-279].

In general, normal mode analysis is in good agreement with molecular dynamics simulations[275, 280-284].

### 2.3.2 Elastic network model

The Elastic Network Model is an efficient way to predict and characterize slow motions of proteins. This model represents the protein as a spring network in which the atoms are individually or collectively connected by the spring. Interactions between atoms are described by a harmonic potential and the protein model is basically represented by a reduced number of atoms at the C $\alpha$  positions of each residue[285, 286]. This simplified model is a well-proved reduction in dimension which is nearly independent of whole-residue motion from side-chain motion as well[287].

The relation between dynamics and structural functionalities of proteins is analyzed by comparing protein homologues. This comparative analysis is used to validate ENM-based NMA[288-296]. Therefore, this comparative analysis helps to check if properties of certain biological systems are intrinsic and depend on the shape of the protein fold in similar systems. Successful examples are described in literature[297-305].

The efficiency of elastic network model drove the successful development for online tools. These allow normal modes to be calculated for protein structures[295, 306-311]. A well-known interface is the WEBnm@ interface of the Reuther group[312]. It uses the C $\alpha$  force field onto the elastic network model developed by Hinsen *et al.*[287]. The force field can be written as:

$$U(R) = \sum_{\text{all pairs } i,j} U_{i,j}(R) \quad (26)$$

with  $R$  representing a certain conformation of the elastic network model and  $U$  the total corresponding potential energy. The normal modes are the eigenvectors of the mass weighted matrix of the second partial derivatives of  $U$ . They describe intrinsic deformations.

## 2.4 Dynamic cross correlation

To understand the biological function of proteins and enzymes the knowledge of the atomic motions and their correlated character is essential[313]. This information can usually be obtained from a molecular dynamics simulations of a protein.

For years, positional fluctuations and correlations between atoms and their corresponding pairs have been observed and analyzed in molecular dynamics simulations[314, 315]. Many dynamical properties were investigated so far, including the understanding of the related protein functionalities which contain *e.g.*, the identification of atom groups moving concerted, models of binding, dynamics differences between free and bound protein[314, 315].

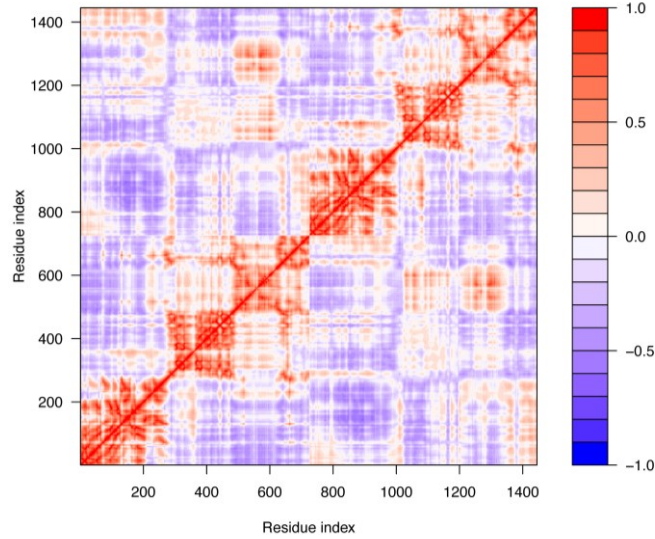
Positional fluctuations and their cross correlation analysis appeals for several reasons. Because of the possibility that fluctuations could be related to crystallographic B-factors, they can be used to be compared with experimental data[314, 315]. Another reason is broadening the scope of classical molecular dynamics simulations via extrapolation of the motions along the direction even on a larger timescale (*e.g.*, used in accelerated molecular dynamics). With this, it is theoretically possible to gain information for substrate binding or other important binding properties[314, 315].

A common computational method is the so-called Dynamic Cross Correlation (DCC)[316]. This method is used to analyze trajectories of certain molecular dynamics simulations. It can be difficult to detect correlated motions but DCC facilitates the elucidation of dynamical and structural relations of particular motions. As a benefit of that, the correlated motions between certain atoms can be quantified[317]. The DCC of atomic positions between atom  $i$  and  $j$  can be described as:

$$DCC_{i,j} = \frac{\langle \Delta r_i(t) \cdot \Delta r_j(t) \rangle}{\sqrt{(\|\Delta r_i(t)\|^2) \cdot (\|\Delta r_j(t)\|^2)}} \quad (27)$$



with  $r_i(t)$  and  $r_j(t)$  representing the vector of the corresponding atom's coordinates (atom  $i$  and  $j$ ) as a time depending function of  $t$ .  $\langle \rangle$  is the time average and  $\Delta r(t)$  is the difference between the starting and end coordinates after the whole simulation. Finally the values are calculated and plotted as the dynamic cross correlation matrix (refers to a covariance matrix)[316]:



*Figure 23: Schematic cross correlation matrix. Red areas represent positive correlations between the positions of the residues. Blue areas represent negative (anti) correlation between the positions of the residues. White areas are not correlated residues.*

## 2.5 Principal component analysis

### 2.5.1 Definition and principle

The main goal of PCA is to reduce a multidimensional data set to a smaller data set. The new data set of unrelated attributes should be able to describe the majority of the variance of the data set as a whole. The original data set or its data matrix with  $n$  number of samples that belong together with  $m$  number of measurements are defined as new matrix  $X$ .

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \quad (28)$$

Subsequently, a deviating matrix  $Y$  is built by averaging the data of matrix  $X$ . Subsequently, a deviating matrix  $D$  is built by averaging the data of matrix  $X$ . The mean value centering is performed by subtracting the mean value of the data from each individual data point. The goal is to remove arbitrary distortion from measurements.

$$Y = \begin{bmatrix} x_{11} - \bar{X}_1 & \cdots & x_{1m} - \bar{X}_m \\ \vdots & \ddots & \vdots \\ x_{m1} - \bar{X}_1 & \cdots & x_{nm} - \bar{X}_m \end{bmatrix} \quad (29)$$

This is followed by the calculation of the covariance matrix  $Z$  of the data set with

$$Y = \frac{Y \cdot Y^T}{n} \quad (30)$$

results in

$$Z = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix} \quad (31)$$

with

$$a = \frac{1}{n} \{ (x_i - \bar{X}_i)(x_j - \bar{X}_j) \} (i, j = 1, 2, \dots, m) \quad (32)$$

The variables of variance and covariance are only absolute values. If the values of the covariance matrix are not measured in the same units, they lose comparability. This means that the variables with larger values automatically produce larger variances and the variables with small values produce small variances.

To avoid this dependency, the normalization of the data and division of each matrix element by its standard deviation is used.

The normalized matrix element  $A_{ij}$ ,

$$A_{ij} = \frac{a_{ij}}{\sqrt{\text{Var}(i) \cdot \text{Var}(j)}} (i, j = 1, 2, \dots, m) \quad (33)$$

The variance of the  $i/j$ -th element is given by  $\text{Var}(i,j)$ . The correlation between both variables,  $A_{ij}$ , can never exceed  $\sqrt{\text{Var}(i) \cdot \text{Var}(j)}$  and thus never become higher than 1 or lower than -1. Therefore, if two variables are uncorrelated and the covariance is zero ( $A_{ij} = A_{ji} = 0$ ).

The reduction of dimensions is the main goal of a PCA. At the same time, the maximum variation of the original data set should be preserved.

Since a covariance matrix defines both scatter (variance) and orientation (covariance), an  $M \times M$  covariance matrix in combination with the mean values can explain a normally distributed  $M$ -dimensional data set.

A process known in mathematics for converting a square matrix into a diagonal matrix is called matrix diagonalization. This matrix has the same basic properties as the original matrix. The diagonalization is also a transformation of the original variable into a special new set of variables, but now with canonical form. This is virtually the same way of calculating eigenvalues and eigenvectors as with a square matrix. However, the eigenvalues lie on the diagonal and the associated eigenvectors form the variables of the new diagonal matrix.

To obtain the eigenvalues and vectors from the correlation matrix  $Z$ , it must first be diagonalized.

$$Z\vec{v}_i = \lambda_i \vec{v}_i \quad (i = 1, 2, \dots, m) \quad (34)$$

with  $\lambda_i$  represents the eigenvalue (scalar value) and  $\vec{v}_i$  is the corresponding eigenvector of the matrix  $Z$ .

$$Z\vec{v}_i - \lambda_i \vec{v}_i = 0 \quad (35)$$

$$(Z - \lambda_i I)\vec{v}_i = 0 \quad (36)$$

with  $I$  represents the identity matrix with the same dimensions of  $Z$ . As long as  $\vec{v}_i$  is not null-vector, the equation (36) is only defined if  $(Z - \lambda_i I)$  is not invertible.

When a square matrix itself is invertible the determinant is zero.

$$\det(Z - \lambda_i I) = 0 \quad (37)$$

The solution of equation (36) yields the required eigenvalues and the corresponding eigenvectors.

### 2.5.2 Interpretation of eigenvalues and eigenvectors

Equation (34) applies to each eigenvalue eigenvector pair. The same equation is written as a matrix form.

$$ZV - LV = 0 \quad (38)$$

with V representing a MxM matrix with eigenvectors of Z are the columns and L represents the diagonal matrix. The non-zero elements of L are the eigenvalues of Z. Being an eigenvalue, L and equation (38) can be written as,

$$ZV - VL = 0 \quad (39)$$

Multiplying equation (39) with  $V^{-1}$  delivers,

$$Z = VLV^{-1} \quad (40)$$

Equation (40) represents the correlation matrix as a function of its eigenvalues and -vectors. It describes the so-called eigen-decomposition of the correlation matrix.

Here the correlation matrix is decomposed by diagonalization into a sequence of rotation and scaling operations. The resulting rotation matrix (V) is defined by the eigenvectors of the correlation matrix. While the so-called scaling matrix(L) is given by the diagonal matrix. Its elements or non-zero elements are the corresponding eigenvalues.

These in turn represent a variance and the corresponding eigenvector explains the direction of this.

The resulting sequence or arrangement of the eigenvalues (from highest to lowest or vice versa) thus results in an ordered basis. This base is in the order from the highest variance to the lowest.

### 2.5.3 Eigenvector projection

The eigen-decomposition obtained in chapter 2.5.3 and the original  $m$  correlated variables  $(X_1, X_2, \dots, X_m)$  are converted into new  $m$  uncorrelated variables  $(S_1, S_2, \dots, S_m)$ . This transformed matrix is given as,

$$V = [v_1 \dots v_m] \quad (41)$$

Thus, the eigenvectors are defined by an  $M \times M$  matrix. Thus, the original matrix can be transformed into the matrix  $S$  with the help of this eigenvector matrix.

$$S = V^T(x - \bar{X}) \quad (42)$$

$x$  and  $\bar{X}$  are  $m \times l$  vectors of observations on the original variables and their means.  $S$  the new transformed matrix represents principal components (PCs).

The  $i$ -th PC is described by equation (43). It has the mean value zero with a variance of  $\lambda_i$ . Basically, the most significant PCs are centered at zero. This means that basically all original variables are subtracted by the corresponding mean values.

$$s_{ni} = v_{n1}(x_{1i} - \bar{X}_1) + v_{n2}(x_{2i} - \bar{X}_2) + \dots + v_{nm}(x_{mi} - \bar{X}_m) \quad (n \leq m) \quad (43)$$

The reduction to a smaller number of dimensions of the original data set while maintaining the maximum variance is the main goal of PCA. If all main components are

retained, the variance is automatically 100%. The practical benefit is the dimensional reduction while maintaining the defined variance range.

If only a certain number  $n$  of principal components from the original data set with  $m$  variables are used, a variance results which is described in equation (44).

$$a = \frac{\sum_{i=1}^n \lambda_i}{\sum_{j=1}^m \lambda_j} \quad (44)$$

#### 2.5.4 Choice of eigenvalues

The principal components of significant importance are determined by their corresponding eigenvalues. The rule of thumb is to choose the main components whose eigenvalues have a significant value ( $\geq 1$ ). However, this rule must be evaluated individually from system to system and may differ.

A second rule of thumb is to use a scree plot[318]. In this plot, the calculated eigenvalues are plotted against the main components and simply used up to a certain point of the function graph, the main components. Both methods can be used individually or even in combination. This depends on the system to be examined.

### 2.6 pKa values in biological systems

In almost all biological systems, knowledge of the protonation state of the amino acids is very important. To estimate the protonation state of a particular amino acid, the pka value can be calculated.

#### 2.6.1 Calculation

The acid HA decays into an  $A^-$  anion and an  $H^+$  proton. This dissociation is described in equation (45).



The equilibrium constant  $K_A$  on which this is based is defined as,

$$K_A = \frac{[A^-][H^+]}{[HA]} \quad (46)$$

In addition, the pH value is defined as the negative decadic logarithm of the proton concentration. Adding the logarithm of the acid constant results in the so-called Henderson-Hasselbach equation (47),

$$pK_A = pH - \log \left( \frac{[A^-]}{[HA]} \right) \quad (47)$$

Replacing the logarithmic pattern in equation (47) with the protonation probability, equation (48) is obtained,

$$pK_A = pH - \log \left( \frac{1 - \langle \chi \rangle}{\langle \chi \rangle} \right) \quad (48)$$

To calculate the protonation probability, equation (48) is rearranged and the following expression is obtained.

$$\chi = \frac{e^{-\ln(10)[pH - pK_A]}}{1 + e^{-\ln(10)[pH - pK_A]}} \quad (49)$$

The free enthalpy of reaction necessary for the pKa calculation for the in this case deprotonation of the titratable group to be calculated is given by equation (49) to the following expression,

$$\Delta G_{Acid} = -\ln(10) \cdot RT \cdot [pH - pK_A] \quad (50)$$

with R representing the universal gas constant and with T representing the temperature.

### 2.6.2 Proteins

The equations described in chapter 2.6.1 are basically applicable to all biological systems in aqueous solution. However, for proteins, an additional term must be considered. This is the term of the protein environment, which leads to the following expression

$$\Delta G_{Acid} = -\ln(10) \cdot RT \cdot [pH - pK_A] + \Delta G_c \quad (51)$$

with  $\Delta G_c$  representing the free reaction enthalpy term of the environment of the titratable residue within the protein.

Proteins or enzymes pass through various confirmations due to the change in pH value. A change in the protonation state of the side chains of the amino acids is more likely. Furthermore, individual changes in the protonation state can have far-reaching consequences on other or neighboring amino acids and thus also change them. To ensure that this possibility is not neglected, the correction term is regarded as a function of all titratable residues in the protein. It thus forms a protonation pattern.

The basic pattern is described as the so-called protonation vector  $c(pH)$ . It is now assumed that all titratable residues change the protonation states from a start vector  $c_r$ (=reference) to any protonation state  $c_k$  (=new vector). The resulting energy is thus  $\Delta G_c(c_r \rightarrow c_k, pH)$ . The choice of the reference vector can be arbitrary or based on earlier work[319, 320]. The calculation of  $\Delta G_c$  is described in the following chapter.

### 2.6.3 Calculation of $\Delta G_c$

The equation (52) shows the probability  $\chi(c_k)$  of a certain protonation vector at given pH,

$$\chi(c_k, pH) = \frac{e^{-\frac{\Delta G_c(c_k, pH)}{k_B T}}}{\sum_i^{N_c} e^{-\frac{\Delta G_c(c_i, pH)}{k_B T}}} \quad (52)$$



with  $k_B$  representing the Boltzmann constant,  $T$  the temperature and  $N_c$  the number of protonation states.

The most important information about the protonation state is at physiological pH (usually pH 7). Ideally, all theoretically possible protonation states should be considered at this pH value.

Since the demands on computer-based resources would be immensely high, methods (for example Karlsberg2+) have been developed over time to increase the efficiency of pKa value calculation. Karlsberg2+[320, 321] uses the Monte Carlo algorithm to determine the probability of a certain protonation state. The functionality of Karlsberg2+ is essentially described in chapter 2.6.4.

In principle, equation (49) can be used as a good approximation for the probability of protonation. However, the pH value is often a complex function where the aqueous environment has to be taken into account, which can lead to a shift in the pKa value. Equation (51) is used to determine the pKa values.  $\Delta G_c$  can be described with four individual energy terms[322-324].

$$\begin{aligned} \Delta G_c(c_r \rightarrow c_k, \lambda, pH) &= \sum_{\mu=0}^N (x_{\mu}^k - x_{\mu}^r) \cdot RT \ln(10) \cdot (pH - pK_{A,ref}^{\mu}) \\ &+ \sum_{\mu=0}^N \Delta \Delta G_{desolv, \mu}^{\lambda}(s_{\mu}) + \sum_{\mu}^N \sum_{\eta, \mu \neq \eta}^N W_{\mu, \eta}^{\lambda}(s_{\mu}, s_{\eta}) + \Delta G_{conf}^{\lambda} \end{aligned} \quad (53)$$

with the first term represents equation (51) and  $x_{\mu}^k$  can be either zero (residue is protonated) or one (residue is deprotonated).  $x_{\mu}^r$  differentiate between acid (= 0) or alkaline (= 1).  $\Delta \Delta G_{desolv}^{\lambda}$  describes the energetic interactions of the titratable group in the protonation state  $s$  with other titratable groups. The other groups must be in the reference state. Due to this assumption, the interactions with the other titratable groups are missing but are described in the third term  $W$  ( $W$ -matrix). The last term represented with  $\Delta G_{conf}^{\lambda}$  describes the conformational change or its energetic change. All titratable residues are in their reference states.

## 2.6.4 Karlsberg2+

Originally published as Karlsberg+[319], this program was modified and optimized and re-released as Karlsberg2+[320]. This modern and due to modifications much more efficient computer aided method yields very good pKa-value calculations. Here, the basic procedure of Karlsberg2+ is explained. The commonly used source of supply for crystal structures is the Protein Database (PDB)[325]. Unfortunately, the identified structures show an important deficiency: in principle, the hydrogen atoms are missing in all structures.

The hydrogen atoms and their (if present) hydrogen bonding networks usually provide important information for possible electrostatic interactions in biological systems. By knowing this, a solid pKa value calculation, as described in the previous chapters 2.6.1 – 2.6.3, is possible with programs like Karlsberg2+.

An essential feature for the calculation of pKa values with Karlsberg2+ is the application of pH-adapted conformations (PCA)[319]. In other words, the program models a protonation pattern for each selected pH value.

Throughout the development stages it could be shown that essentially three pH values are sufficient to calculate pKa values. For Karlsberg2+ these are the pH values -10, 7 and 20.

At a value of seven, all residues are ionized and promoted to -10. Acids are thus neutrally charged and alkalines positively charged. The same applies to the pH value (deprotonated and acids negative and alkalines neutral). In the beginning, Karlsberg2+ uses the same structure for all PACs and calculates the protonation pattern. After completion and successful modeling, they are used in a titration algorithm (based on equation (53)). The titration itself is carried out in steps of 0.5 pH units using equation (52) and the pKa values are calculated.

## 2.7 Docking

### 2.7.1 Theory

Molecular docking refers to a computer-based method used to predict a potential interaction of two or more molecules. Accompanied by the creation of the associated model. Nowadays, docking plays a central role in many drug discovery applications, such

as interaction between small molecules and macromolecules (e.g., protein-ligand docking). More recently, this method is also used for binding predictions between large molecules (protein-protein interactions) and is thus no longer limited to smaller molecules. The current basis of most docking programs is molecular mechanics. An improved agreement between theoretical models and experimental data can be achieved by more accurate parameters such as charge, torsion and angular geometries[326].

From the early 1980s to the present, molecular- as well as protein-modeling has been achieved by using force fields and has been improved and exploited by the addition of molecular processes, such as protein-ligand binding. The latter is divided into two main methods: the rigid body approach (related to Emil-Fischer's model) and flexible docking.

In the rigid body approach, both the receptor and the ligand are considered as independent bodies. Their mutual affinity is based on shape and volume recognition. While in flexible docking, the mutual affinity is based on the respective conformational recognition[327]. Although docking has evolved uniquely to this day, the negative side effect of the increased use of this technique is that also inaccurate applications in CADD (Computer-Aided-Drug-Design) have been developed[328].

Therefore, as with all other techniques, limitations must be considered. It should be noticed that the algorithm applied must be adjusted individually for each system in order to ensure the highest possible accuracy of the poses obtained[329, 330].

Regardless of the choice of docking program, the process can basically be divided into two major steps: sampling[331, 332] and scoring/ranking[333]. Whereby in principle a "preliminary step" before sampling and scoring, the ligand and protein preparation, should not be neglected.

### **2.7.2 Ligand protein preparation**

**Ligand and protein preparation** is the preliminary step before the actual docking procedure. First, obtaining a protein structure including the bound ligand is mandatory. The need for a three-dimensional structure is ensured by the protein crystal structure, which can be obtained from the Protein Data Bank (PDB). As described in section 2.7.1, docking requires the modification of several parameters. The information of these within

the PDB is mostly not sufficient, whereby most diverse programs (for example Maestro from the Schroedinger Group or UCSF Chimera), are used for this purpose. The parameterization depends on the software used. Consequently, consistency in ligand as well as protein preparation is mandatory[334, 335].

Basically, the receptor or protein and the ligand are prepared separately from each other. Whereby both are based on similar considerations. Thus, the first step in ligand preparation is usually to extract it from the protein. If a ligand is not available as a structure, a sketch is usually prepared as a MOL file and structure optimized (minimization, charge assignment, etc.) and then processed as a separate file. The protein/receptor is prepared in very similar ways (geometry optimizations, minimizations, etc.) and despite assumptions that this may be impractical, showed that complex optimizations of both molecules can significantly increase the results and interactions of both[336-340].

An additional and essential step in protein preparation is the selection and delineation of the binding site. This is done manually or automatically based on the coordinates of the ligand by means of the program used. If no binding site is known, many programs now allow the calculation of possible cavities or potential binding sites, making even "blind" docking possible as Marazo[341] and Espinoza[342] successfully demonstrated. A common software for mapping the binding site is GRID. Essentially, a grid box with predefined dimensions, which is divided into small individual boxes, is placed over the potential binding site. The boundary atoms then reflect potential interaction contours. The size as well as the resolution of the GRID is thus essential[343]. In the recent past, Feinstein and Brylinski[344, 345] were able to show that the size has an influence on the hit identification and the binding site screening of ligands. This calculation is done either simultaneously "on the fly" or separately for each ligand. Autodock and Vina are prime examples of this situation: while the latter performs the calculations "on the fly", Autodock requires the precalculated GRID of each ligand[344, 345].

### 2.7.3 Sampling methods

The embedding and the conformation of a ligand depends on the type of docking algorithm. The search for the correct conformation can be done in two different ways: systematically or stochastically/randomly.

**Systematic sampling** for ligand conformations requires comprehensive sampling of individual conformational possibilities in conjunction with structural parameters. Due to the degrees of freedom of rotation as well as translation, a very large number of possibilities of binding positions and corresponding binding conformations results which would end up in possible combinatorial explosion. Various sampling algorithms have been developed to address this problem.

One of the more common sampling methods is the matching algorithm (MA)[346-348]. In this sampling approach, ligand formation is based on shape characteristics and chemical information of the ligand. The receptor as well as the ligand form a pharmacophore. The respective conformations are determined by calculations using a distance matrix between ligand and protein. Chemical-based properties such as hydrogen donors and acceptors are also taken into account. One advantage of these algorithms is their speed. Programs such as DOCK or LibDock use this sampling method.

A much better known and nowadays very common method to counteract the problem of combinatorial explosion is the fragmentation of individual ligand components. In the fragmentation and construction methods (incremental construction IC), the ligands are docked piece by piece to the desired active site in the receptor[349-351]. Fragmentation of the ligand is ensured by breaking the rotatable bonds. A "start fragment" is selected, which is first docked to the active site and serves as an anchor for the subsequent fragments. These are then added incrementally using different orientations and or conformations, which at the same time emphasizes ligand flexibility. This method, which is the most common today to balance computer costs and time, is used in programs such as DOCK, FlexX or SLIDE[349, 352, 353].

**Stochastic/random sampling** is essentially done using two common methods: Monte Carlo (MC) or genetic algorithm (GA). In the application of MC methods[354, 355],

different poses of a ligand are determined by rotation of the bonds, rigid body translation and rotation of the ligand itself. In this process, each achieved pose is evaluated based on an energy criterion implemented in the algorithm. If the criterion is passed, the pose is stored as a possible scenario and further modified. The evaluation of the pose is repeated until a predefined number of conformations is reached. The MC methodology has the advantage that the conformational changes can be relatively large, overcoming possible energy barriers of the ligand on the potential energy surface. MC methods find application for example in AutoDock[356], QXP[357] or Affinity[358].

#### 2.7.4 Scoring functions

The search for a correct conformation, which is referred to the sampling process, can result in an extraordinarily large number of structures. Of these, only a certain percentage are biologically relevant[333]. This means that the purpose of scoring is to distinguish the right from the wrong poses. This distinction is based on broad properties including intermolecular interactions, (de)solvation, and electrostatic and entropic effects. Scoring can be essentially divided into three types: force-field-based, empirical, and knowledge-based.

**Force-field-based** scoring functions in a classical manner, obtain the binding energy by calculating the sum of the non-binding (electrostatic and van der Waals forces) interactions. The electrostatic part is calculated by the Coulomb equation[359-361]. Extensions of the classical treatment of force-field-based calculations, additionally include hydrogen bonding, solvation effects, and entropy effects. Whereas typical scoring functions imply especially the latter not completely.

A very widely used method today to solve this problem of solvation is the combination of Molecular Mechanics (MM) and Poisson-Boltzmann Surface Area/Generalized Born Surface Area (MM/PB(GB)SA)[362-366].

This method is the combination of classical force field and a continuous solvation model for the calculation of binding free energy. Through this combination it is possible to calculate the binding energies in an efficient way. Basically, the following equations are used to calculate a receptor (R) and a ligand (L) as a complex structure (RL) and its binding free energy, respectively[367]:

$$\Delta G_{bind} = \Delta G_{RL} + \Delta G_R + \Delta G_L \quad (54)$$

and

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S \quad (55)$$

with

$$\Delta E_{MM} = \Delta E_{bonded} + \Delta E_{nonbonded} \quad (56)$$

$$\Delta E_{bonded} = \Delta E_{angles} + \Delta E_{dihedrals} \quad (57)$$

$$\Delta E_{nonbonded} = \Delta E_{electrostatics} + \Delta E_{vdW} \quad (58)$$

The free energy of binding is divided into different terms given by equation (54) and (55). Apart from entropy and temperature, the total molecular mechanics energy  $\Delta E_{MM}$  and the solvation energy  $\Delta G_{solv}$  play an important role. The  $\Delta E_{MM}$  in equation (56) is divided into binding and non-binding energy terms. The bonded energy terms see equation (57) are divided into angles and dihedrals, while the nonbonded energy terms are divided into electrostatic interactions and van der Waals forces.

The free energy of solvation  $\Delta G_{solv}$  is also divided into different terms. The subdivision is made into a polar and non-polar term. The latter one represents the non-polar part and the solvent accessible surface area. In the polar part, the difference is found in the applied method.

$$\Delta G_{solv} = \Delta G_{GB/PB} + \Delta G_{surf} \quad (59)$$

Thus, equation (59) is instrumental in determining how the solvation is calculated, which is either Poisson-Boltzmann (PB) or the Generalized Born (GB) method[367].

**Empirical scoring** divides the binding energy into different components. In addition to that, reproducing experimental data is also involved. In this respect, they will be related to the QSAR models: linear regressions of descriptors to model protein-ligand interactions[368-374]. The scoring function and the parameters used (as for example in

PLANTS[375]) are therefore simpler. However, the QSAR counterparts argued that such energy parameters were not robust enough[333, 376]. This assumption does not exist without reason since most empirical scores are derived from heterogeneous training data and are thus individual and limited. Nevertheless, one advantage of this method is that the calculation can be performed faster than the methods that are force field based[374, 377].

**Knowledge-based scoring** was developed to represent structures rather than energies, in contrast to its other representatives. The former are constructed using pairwise potentials emanating from known ligand-receptor complexes. The methodology is based on the inverse Boltzmann relation (vide infra), starting from statistical mechanics. Mean force potentials are calculated instead of real ones[333, 378, 379]. Due to the balance between power and time of calculation, an advantage of this method cannot be denied. In addition, other rather unusual interactions (for example, sulfur-aromatic interactions)[376] can be taken into account. However, on the other hand, a limitation is given by the inverse Boltzmann relation. For this, a reference state must first be defined (the potentials are set to 0). However, this is anything but trivial and influences the result significantly[380]. To improve the precision of the prediction, a form of "hybrid" approaches have been developed to date. A well-known example is Small Molecule Growth (SMoG2016), which combines empirical data with knowledge-based data[381].

### 2.7.5 Docking methods

The docking procedure requires some preparation as described in the previous sections. Apart from that, it also requires the choice of the type of docking.

**Rigid ligand and rigid receptor docking[382-385]** is the choice when both binding partners are treated as rigid bodies. The search space is very limited by the rigidity of both components by limitation to three translational and three rotational degrees of freedom. However, addressing the flexibility, the solution could be through precomputed ligand conformation sets. On the other hand, this can be accomplished by allowing for a degree an atom-atom overlap between the receptor and the ligand. Examples of earlier



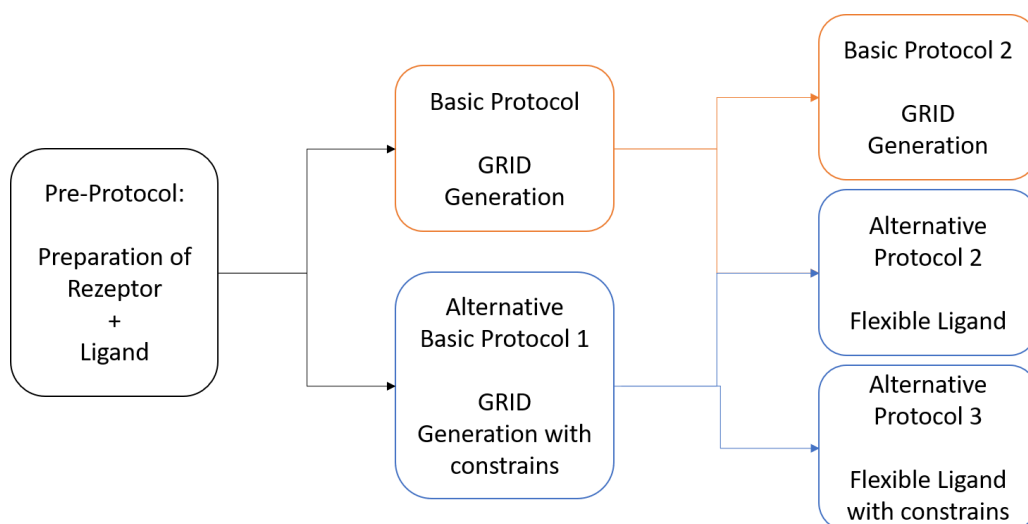
versions of docking programs that use the rigid receptor and rigid ligand method are DOCK, FLOG, and protein-protein docking programs such as FTDock[351, 352, 386-389].

**Flexible ligand and rigid receptor docking[382-385]** is used in systems that follow the behavior of the induced fit paradigm. For these systems, it is important to consider ligand and protein and their respective flexibility to address the conformational change of both. This is important because the conformational changes involve the formation of a minimum-energy perfect-fit complex[390, 391]. Due to the high computational time cost of also having a flexible receptor, the standard approach is a compromise to keep the ligand flexible and the receptor rigid. Almost all docking programs such as AutoDock and FlexX have already implemented this type of docking[349, 392].

**Flexible ligand and flexible receptor docking[382-385]** is rarely used because of computational cost. As Teague has demonstrated[393], there is a tightly linked correlation between intrinsic mobility of proteins and ligand binding. To date, incorporating receptor flexibilities in docking has been a major challenge. Ideally, MD simulations could help determine the degrees of freedom of ligand-receptor complexes. The problem of the extremely large computational effort involved has not yet been solved.

### 2.7.6 Examples of docking applications

**Glide[394, 395] (non-covalent docking)** is a variation in which a flexible ligand docks onto a rigid receptor. In this case, a degree-of-freedom-dependent scanning of the ligand takes place based on conformational, orientally and positional degrees of freedom. Basically, there are three different modes of proceeding which differ by sampling variation of the ligand degrees of freedom and the scoring function itself.



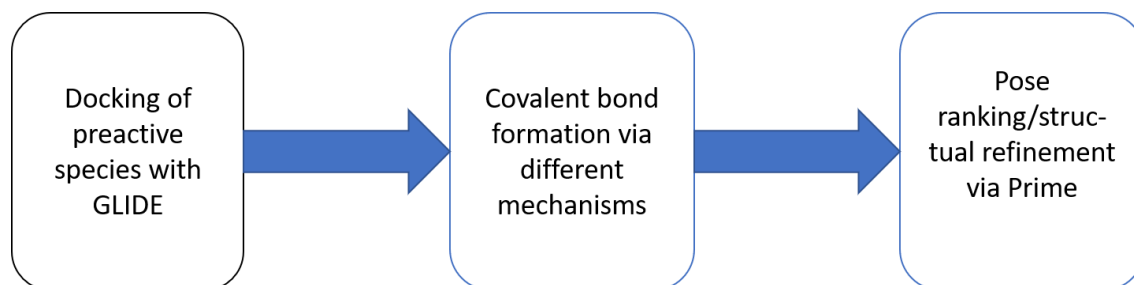
*Figure 24: Representation of different processing options of the docking program GLIDE. The black box represents the preparation of the ligand as well as the receptor or protein. The orange boxes and associated arrows represent the standard/base protocol. The blue boxes and arrows represent possible alternative protocols.*

The single precision (SP) Glide Score function places the connections using SP or high-throughput virtual screening (HTVS). The third variant uses "extra precision" (XP) scoring. XP basically starts with SP docking and is refined using an "anchor-and-grow" algorithm. This results in a more precise degree-of-freedom sampling[373, 394-396]. Before the GRID protocols of docking are performed, the receptor and ligand still need to be prepared according to pre-protocol (Figure 24). Then the main steps of GLIDE docking follow.

In the basic protocol (Figure 24), a lattice generation is performed, which generates the corresponding lattice potentials for the protein/receptor. Next, in basic protocol 2, docking is performed using a flexible ligand with GLIDE (Figure 24).

In alternative basic protocol 1, lattice generation is done first based on certain constraints preselected by the user and followed then by basic protocol 2 or two alternate protocols (Figure 24). In alternative protocol 2, the docking process is based on a flexible ligand docking experiment. Thus, the calculated docking is similar to an experimental data set. In contrast, in alternative protocol 3 (Figure 24), docking is performed by applying the constraints specified beforehand and is not based on experimental data[373, 394-396].

**Covalent docking (CovDock)**[397, 398] approach is a powerful combination of two distinct programs, which is implemented, for example, in Schrödinger's Maestro suite to perform docking for covalently bound ligands. Basically, the procedure is divided into three steps which can be seen in Figure 25.



*Figure 25: Representation of the proceeding order of CovDock program.*

GLIDE (part of the program combination) is first used to dock a “precursor”, a non-covalently bound form of the ligand, to the receptor site. This is followed by local sampling using the Prime program (after the covalent bond between receptor and ligand has been formed), the second part of the program combination. Parameters of covalent bonds are taken from the OPLS3e force field. In addition, CovDock allows encoding of different chemistries to establish the covalent bond. An extension of these bond variants can also be implemented. To remove redundant solutions, CovDock has additional filtering and clustering steps built into the program flow. Thus, ligand positions are displayed at the end, which are optimized and were found either by "Virtual-Screening" (VS) mode or the "Lead Optimization Mode". The speed of calculation is dependent on the size of the whole system, as well as the reaction chemistry (necessary for the formation of the covalent bond).

**Virtual screening (VS)** is still a very popular approach of finding additional molecules that were not previously considered as potential ligands. Moreover, this method allows a very good comparison between potential and already identified ligands. In principle, the molecules found should be relatively small, at least for drug discovery, since this is easier to modify experimentally. Nevertheless, the VS remains a good method to find new ligands even beyond a size of 500 g/mol as described before. With VS it is also possible

to gain theoretically structures which were not available in the PDB but could be modelled now[399].

The applications of VS are manifold. Among others, VS is used in ligand- and structure-based methods. More than 100 methods of this kind exist. Either a 2D structure (topology fingerprints/descriptors, described by Hert *et al.*) or 3D representatives[400-404] are used for this purpose. Whether ligand-based or structure-based methods have priority in terms of reliability has not yet been clarified uniformly. Both methods have delivered comparable results. Clarification of this problem is countered by the suggestion that combining the two methods or even post-pharmacor VS filtering leads to even better results[405-413].

### 3 Motivation

For living organisms, fatty acids are of great importance and play a fundamental role. In addition to providing energy, fatty acids are also fundamental building blocks that are used to form a variety of biological compounds. The latter are necessary for supporting processes at the cellular level. Fatty acids serve as substrates for the formation of membranes additionally to their function in signal transduction pathways[414].

Moreover, fatty acids can also act as precursors of hormones or bile acids. However, the fundamental role of fatty acids in cells is underlined by the identification of genetic diseases in relation to protein defects. This is achieved by the involvement of the defective proteins in fatty acid metabolism. Due to the essential roles and functions of fatty acids in conjunction with their physicochemical properties, the importance of gaining an understanding of the mechanisms that control and ensure the uptake, transport and metabolism is significant. The complex as well as chiral molecules of lipids differ physically as well as chemically in aqueous solution[414].

The most important pathways of lipid metabolism are already known, but there are still processes that have not been elucidated in detail. These include the structural properties of the enzymes that metabolize the respective lipids and lipid classes. An essential process in fatty acid metabolism, which has not yet been elucidated in detail regarding the enzymes and mechanisms involved, is the  $\beta$ -oxidation. This process serves as a source of energy and is thus an essential component of fatty acid degradation.

As described by Lazarow and coworkers[415, 416] the  $\beta$ -oxidation occurs in the mitochondria and peroxisomes[417]. It is an enzyme-mediated reaction. While the mitochondrial  $\beta$ -oxidation is coupled with energy metabolism the peroxisomal  $\beta$ -oxidation is responsible for shortening fatty acids.

There are three types of multifunctional enzymes in mammalian cells that participate in  $\beta$ -oxidation: the trifunctional enzyme complex (TFE)[91], type-2 of the multifunctional enzyme system (MFE-2)[418-421] and the multifunctional enzyme type 1 (MFE-1)[121]. MFE-2 uses a different stereochemistry[421, 422] than its counterpart MFE-1, which complements the  $\Delta^3$ - $\Delta^2$ -enoyl-CoA isomerase activity that the TFE enzyme lacks.

Therefore, the MFE-1 has 2-enoyl-CoA hydratase-1,  $\Delta^3$ - $\Delta^2$ -enoyl-CoA isomerase, and (3S)-hydroxyacyl-CoA dehydrogenase activities in a polypeptide chain[112]. Enzymes with common ancestors are structurally very similar, although there can be significant

divergences concomitantly, which is why different chemical reactions can be catalyzed, as in the case of the TFE, MFE-2 and MFE-1. Independently, these enzymes are able to transport the fatty acid derivatives and intermediates by passing the substrate between their active sites. Thus, these enzyme systems establish a highly complex and essential structure-function relation that has not yet been elucidated in detail.

The scope of this dissertation was to determine possible relationships between structure and function at the atomistic level. Using classical and accelerated molecular dynamics simulations, different MFE-1 constructs were investigated and compared to explore possible structural-functional relationships. Essentially, the active site I was studied to gain detailed insight into the functional nature of the active site I and to demonstrate a fundamental understanding of the interactions and catalytic functionalities as proposed by Wierenga *et al.* [121, 122, 233]. Furthermore, dynamic cross correlation studies were used to investigate the fundamental dynamics of the entire MFE-1 system to draw conclusions about domain movement and the likely resulting substrate channeling mechanism.

To complete the basic understanding of MFE-1 at the mechanistic and structural level, a previously unavailable crystal structure with a substrate in the second active site of MFE-1 was modeled using the docking method. This should help to understand the second active site on a structural level in more detail and additionally to identify other potential ligands MFE-1 is capable to metabolize. Overall, this dissertation is intended to sharpen the fundamental understanding of MFE-1 at the dynamics and on the structural level while providing a basis for future structure-function relations from a molecular dynamics perspective.

## 4 Results – Active site I of MFE-1

### 4.1 Introduction

The atomistic description of the multifunctional enzyme type 1 (MFE-1) is one objective of this thesis. Since it is involved in several metabolizing pathways, MFE-1 contains two active sites, as described in more detail in section 1. The following section deals with the so-called crotonase part (domain A), which harbors the first active site (Figure 26) that catalyzes two of the three reactions of the fatty acid metabolism (see section 1).

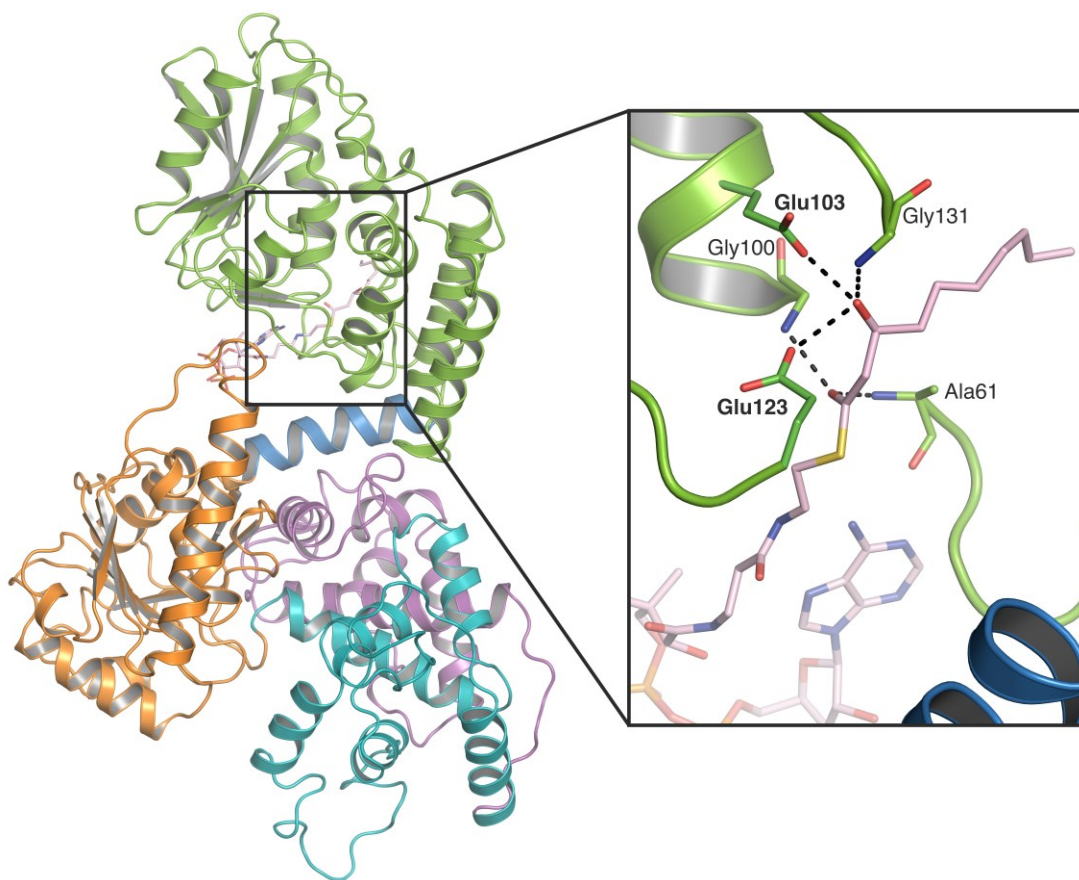


Figure 26: **Left:** Overall fold of the MFE-1 in cartoon representation. Domain A is green, domain B is blue, domain C is orange, and domains D and E are cyan and magenta, respectively. **Right:** First active site (hydratase active site) with highlighted glutamates 103 and 123 as stick models. The amino acids glycine 100 and alanine 61 as part of the oxyanion hole are also shown. 3S-hydroxydecanol-CoA represents the ligand and is shown in light-pink.

Crystallographic studies[423] have demonstrated that substrates like 2E-enoyl-CoA are hydrated in the first active site, while the fatty acid tail is buried inside domain A. This hydrogenation reaction is assumed to require a water molecule as a trigger. In the first active site, a water molecule was found to be in hydrogen bonding distance to glutamate 103 and 123 and is thus stabilized[122]. In that case, Kasaragod *et al.*[122] suggested that one glutamate plays a crucial role for the hydrogenation reaction. Concerning the MFE-1, it might be either glutamate 103 or glutamate 123. Kasaragod and coworkers suggested further that Glu123 triggers hydrogenation by reprotonating the substrate. Therefore, it is also assumed that on the one hand, glutamate 123 is protonated, and on the other hand, glutamate 103 is deprotonated[122]. Nevertheless, the correct protonation state and the catalytic mechanism at the first active site are still unknown. To determine the most probable protonation state of catalytic glutamates,  $pK_a$  calculations were done using the program Karlsberg2+[320]. This information allows to investigate how the protonation states influence the stability of the first active site.

## 4.2 Model setup and preparation

The initial model structure of MFE-1 was constructed from the Protein Data Bank entry 3ZW8[424]. The cartesian coordinates of the fatty acid complex 3S-hydroxydecanoyl-CoA were gained out of the PDB 3ZWC and inserted after superimposing into the 3ZW8. The cofactor NAD<sup>+</sup> was extracted out of PDB 3ZW9, superimposed and inserted in the structure of 3ZW8.

Therefore, a total of three MFE-1 systems were prepared: system 1 = apoprotein; system 2 = MFE-1 complexed with 3S-hydroxydecanoyl-CoA present in the first active site; system 3 = holoprotein (including NAD<sup>+</sup> present in the second active site).

Each system consists of 722 amino acids, and all hydrogens were added according to pH7 and built via psf-gen tool VMD (vers. 1.9.1). In addition, the  $pK_a$  calculations were performed using Karlsberg2+[425, 426] to accurately predict the protonation states of glutamates E103 and E123.

Afterwards all three MFE-1 systems were solvated in a 125x110x120 Å<sup>3</sup> water box using an explicit TIP3P model[427] with a 10 Å periodic boundary conditions cut-off. The resulting models contained more than 150000 atoms. The systems were neutralized by



adding 10 mM NaCl ions. These complexes finally were minimized (20000 steps), heated (30000 steps and 300K) and equilibrated (2fs time step, 30000 steps).

To check the structure stabilities of each MFE-1 system, RMSD were computed from a 20 ns short classical MD performed for each system. Afterwards, classical and accelerated molecular dynamics simulations (200ns) were performed for each system on the HLRN (The North German Supercomputing Alliance) supercluster with NAMD (vers. 2.9) using the CHARMM27 force field. All calculations were also carried out with an NPT ensemble at 300K with atmosphere pressure using the Langevin method[258].

### **4.3 Protonation state of Glu103 and Glu123**

#### **4.3.1 Structure stability**

To analyze the structural stability of each MFE-1 model (apoprotein, complexed with 3S-hydroxydecanoyl-CoA in active site I and holoprotein), the root mean square deviation (RMSD) of the backbone C $\alpha$  atoms were calculated (Figure 27). First, a classical molecular dynamics simulation with a length of 200ns was performed for the apoprotein to check if the structure is stable. The result shows that the MFE-1 structure with an RMSD of 2.5 Å (black curve in Figure 27) is stable, despite structural fluctuations observed around 70 and 140 ns. The fluctuation of the apoprotein are due to loop movement of domain A or due to the motion of domain A relative to the other domains caused by the flexible hinge region[121].

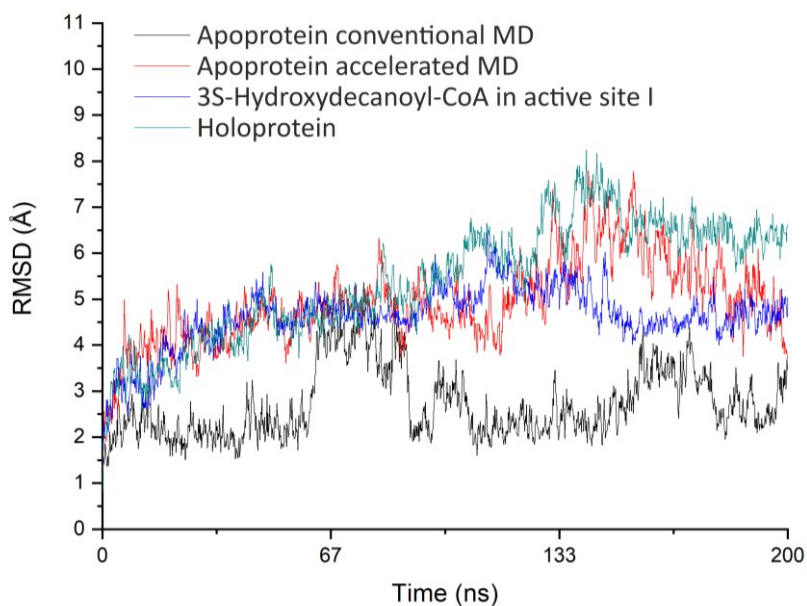


Figure 27: Root mean square deviations (RMSD) of Ca atoms for different MFE-1 systems relative to the crystal structure (3ZW8). The black curve represents the RMSD of the apoprotein with conventional 200ns MD; the red curve represents the RMSD of the apoprotein with accelerated 200ns MD; the blue curve represents the RMSD of the accelerated 200ns MD with 3S-hydroxydecanoyl-CoA molecule present in the first active site; cyan curve represents the RMSD of the accelerated 200ns MD holoprotein (including NAD<sup>+</sup> present in the second active site).

To overcome possible trapped energy states and thereby increase conformational sampling (described in chapter 2.2), 200ns accelerated MD simulations were performed. The corresponding RMSD values are plotted in Figure 27. In contrast to the conventional MD, all models show higher flexibility as expected for aMD, but also very good stability as reflected by the flatter of RMSD evolution plots. The higher RMSD values are due to enhanced exploration of the conformational space as shown in Battocchio *et al.*[428]. The RMSD for the apoprotein (red curve) increases to 4.8 Å, while the RMSD converges to 4.6 Å for the model with the 3S-hydroxydecanoyl-CoA in the first active site (blue curve). For the third model with both active sites occupied, the RMSD is slightly higher at around 6.8 Å but still achieves very good stability throughout its structure. Higher RMSD values and thus flexibilities and motions were observed for domain A and the corresponding hinge region (data not shown). Since the RMSD values of MFE-1 obtained

from accelerated MD are overall comparable to the obtained RMSD values from conventional MD, the following results and analyses are based on the aMD method.

### 4.3.2 pK<sub>a</sub> calculations

As described in the introduction in chapter 4.1, the two glutamates 103 and 123 are of particular importance since they are involved in the catalytic mechanism[122]. The same has also been shown by structural studies on MFE-1 and the reaction mechanism of its first active site in domain A[122]. However, the exact protonation state of both glutamates remains elusive[121, 122, 423]. Therefore, to obtain an initial model the protonation pattern of E103 and E123 had to be calculated using the crystal structure (PDB: 3ZW8). These protonation states were used for all the following, models, and further investigations of the first active site. To compute the protonation pattern of the crystal structure Karlsberg2+[320] was used. Calculations were performed using default parameters excluding water molecules. The gained values are given in Table 1.

*Table 1: pK<sub>a</sub> Values for the glutamate103 and glutamate123 calculated for the crystal structure and apoprotein of MFE-1 (PDB: 3ZW8).*

pK <sub>a</sub> glutamate 103	pK <sub>a</sub> glutamate 123
+6.89	-3.79

These calculations predict a high pK<sub>a</sub> value for glutamate 103. Under physiological conditions and suggesting that this side chain is most likely protonated, whereas the pK<sub>a</sub> of glutamate 123 is much lower, whereby it is most likely deprotonated. Subsequently, these predicted protonation states were used to analyze the pK<sub>a</sub> values of both glutamates throughout the whole MD trajectory. The result of this calculation is given in Figure 28.

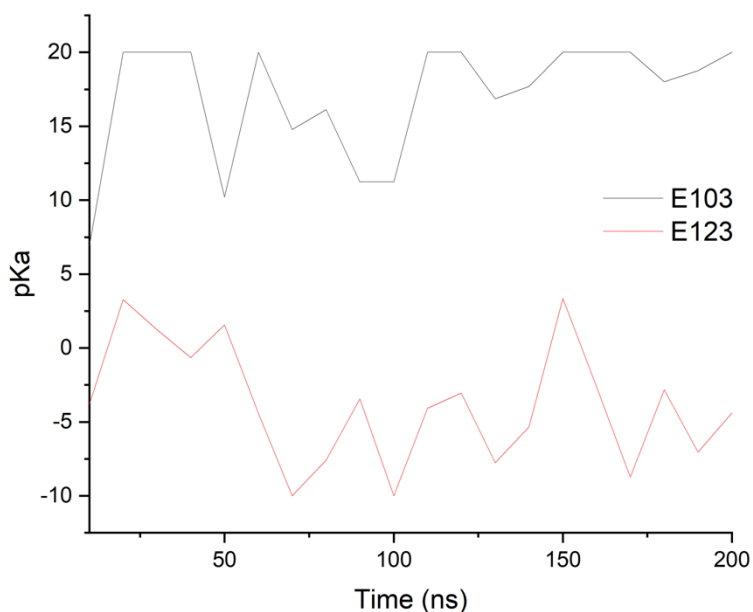
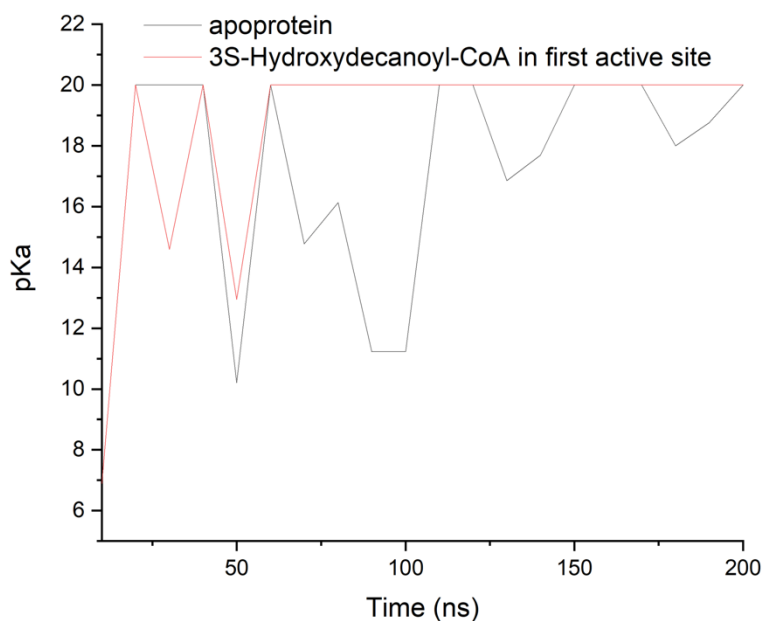


Figure 28: Computed  $pK_a$  values over a 200 ns accelerated MD run of the apoprotein. The black curve represents the  $pK_a$  value of the glutamate 103, and the red curve the  $pK_a$  value of the glutamate 123.

The diagram shows the evolution of the  $pK_a$  value of glutamate 123 and 103 during a 200 ns MD. The black curve represents the  $pK_a$  value of glutamate 103, while the red curve represents the evolution of the  $pK_a$  value of glutamate123. Both values show a steady evolution. While the  $pK_a$  value of glutamate103 is always above seven, at the same time, the  $pK_a$  value of glutamate 123 is less than 7 and even reaches a negative  $pK_a$  value of -5 or less. The  $pK_a$  value of glutamate 103, on the other hand, fluctuates about 10 units starting with the  $pK_a$  value of 7 and ranges up to 20. Takiden *et al.*[429] could show for phytochromes containing propionate side chains a stable protonation state using  $pK_a$  value calculations depending on a MD simulation. Therefore, based on the similar calculations shown here, it is assumed that glutamate 103 is protonated and glutamate 123 is deprotonated.

Subsequently, the evolution of the protonation state in the presence of the 3S-hydroxydecanoyl-CoA substrate in the first active site and its possible impact on the protonation state was investigated. The following diagram shows the  $pK_a$  values during the MD trajectory of both glutamates 123 and 103 (Figure 29).



*Figure 29: Computed  $pK_a$  values over a 200 ns accelerated MD run with the apoprotein and the 3S-hydroxydecanoyl-CoA substrate present in the first active site. The black curve represents the  $pK_a$  value of the glutamate 103 for the apoprotein. The red curve represents the  $pK_a$  value of the glutamate 103 with the 3S-hydroxydecanoyl-CoA present in active site I.*

Figure 29 shows the evolution of the  $pK_a$  values of glutamate 103 of MFE-1 with 3S-hydroxydecanoyl-CoA in the first active site compared to the apoprotein. Interestingly, both models show very similar behavior with a steady  $pK_a$  value above 7. However, the  $pK_a$  value for the apoprotein fluctuates between 10 and 20, while the  $pK_a$  value significantly increases due to the presence of the substrate and remains stable at 20. The latter behavior reinforces and underlines the assumption of the protonation state for the following analyses and MD simulation results, where glutamate 103 is protonated while glutamate 123 is deprotonated ( $pK_a$  values of glutamate 123 not shown but are similar to the apoprotein in Figure 28).

## 4.4 Structural changes of glutamates E103 and E123

### 4.4.1 MFE-1 apoprotein

Proposed by Kasaragod *et al.*[122] the two glutamates E103 and E123 seem to interact with a water molecule that could be part of the hydrogenation mechanism. Thus, the ability to form a hydrogen bond interaction between the two glutamates and the water molecule depends on the distance between them[122]. Accordingly, the distance between the two oxygen atoms Oε1 of E103 and E123 was investigated over the whole trajectory for each MD simulation (apoprotein, MFE-1 complexed with 3S-hydroxydecanoyl-CoA in the first active site and holoprotein). As the first step, this distance was investigated for the apoprotein, and the result is illustrated in Figure 30.

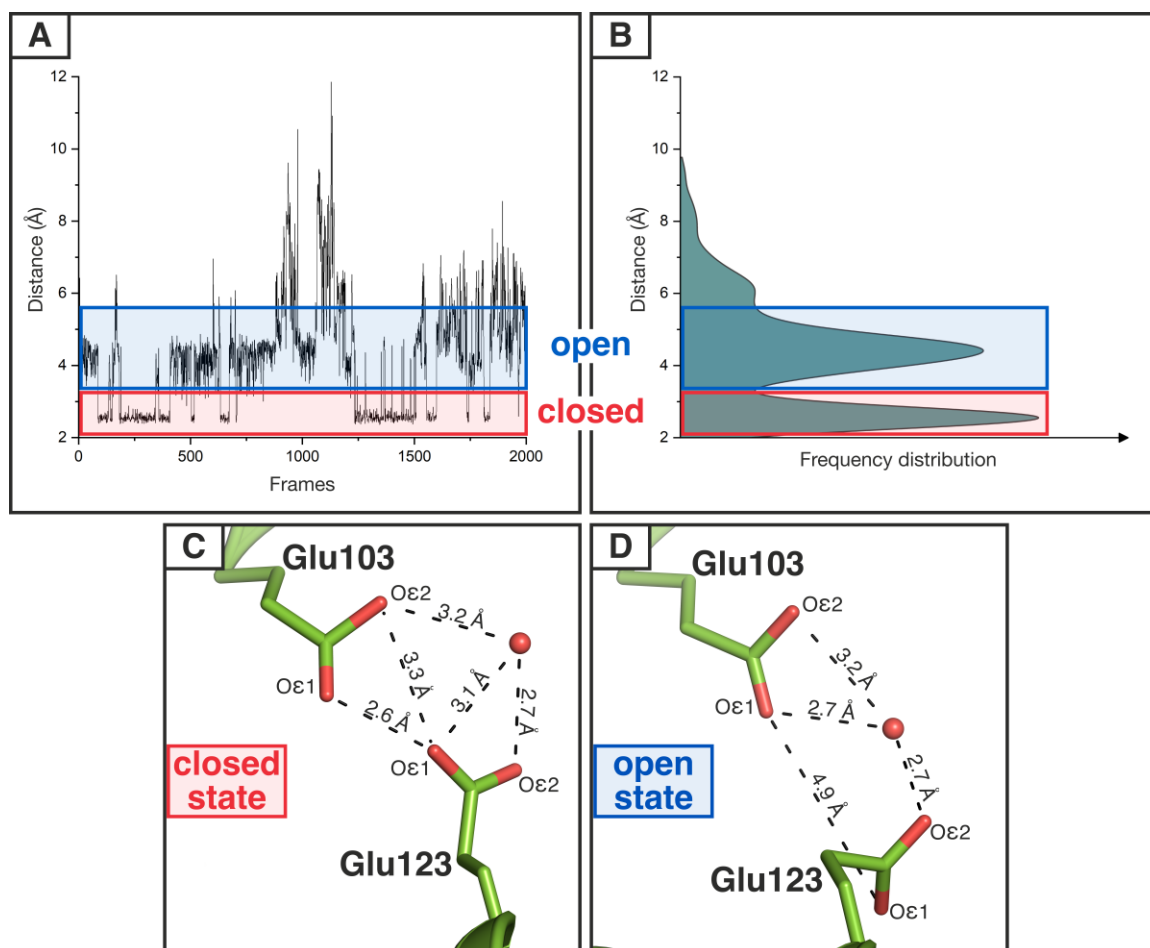


Figure 30: **A** Calculated distance plot for the apoprotein between the oxygen atoms Oε1 of the protonated glutamate 103 and deprotonated glutamate 123. The distance is measured over the whole 200 ns accelerated MD trajectory for the apoprotein (PDB: 3ZW8). The blue box represents the open state, and the red box represents the closed

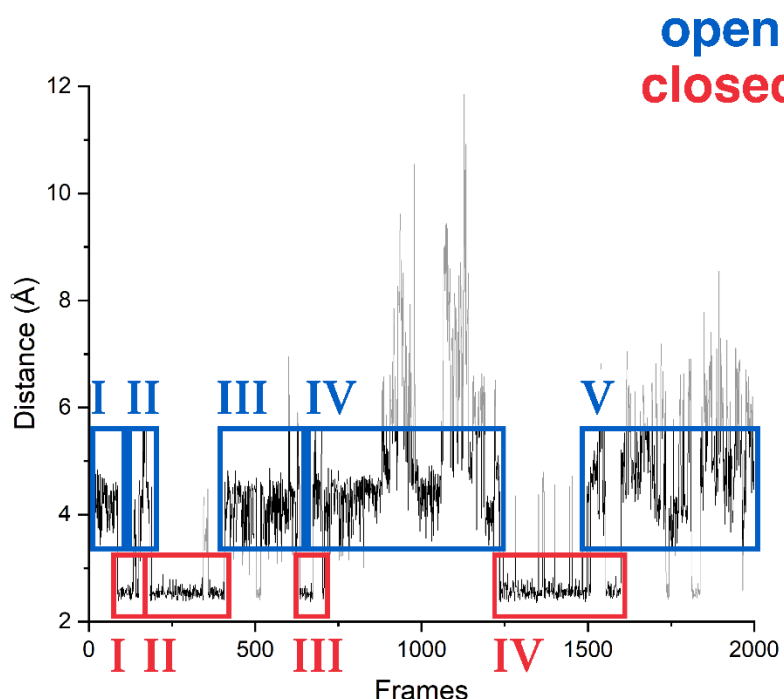
state, referring to the distance between the two oxygen atoms of both glutamates. **B** Frequency distribution plot of the distances between the O $\epsilon$ 1 oxygen atoms of E103 and E123. Two main peaks are visible, referring to the open (blue box) and closed (red box) state. **C** Structural representation of the closed state based on the crystal structure of MFE-1 (PDB: 3ZW8). The distance between the two O $\epsilon$ 1 oxygen atoms of both glutamates 103 and 123 is 2.6 Å as indicated by the dashed line. **D** Structural representation of the open state based on the crystal structure of MFE-1 (PDB: 3ZW8). The glutamates undergo conformational changes compared to the closed state in panel C. The distance between the O $\epsilon$ 1 oxygen atoms of both glutamates is 4.9 Å representing the open state.

As a starting point for the trajectory the crystal structure of the MFE-1 (PDB: 3ZW8) with a distance of 5.2 Å between the glutamates E103 and E123 (measured between O $\epsilon$ 1 of Glu103 and Glu123) was used. The distance plot over the whole 200 ns accelerated MD trajectory for the apoprotein of MFE-1 is shown in Figure 30 (panel A). In combination with the frequency distribution plot (panel B), which counts the quantity of the distance (between O $\epsilon$ 1 of E103 and E123), the distance plot allows to identify different states. Here, two major peaks reoccurring for the whole trajectory with a similar high frequency are visible. One peak at 2.6 Å is narrow (red box) and the second peak around 4.9 Å is slightly broader (blue box). A structural relation between the distances and the corresponding states of both glutamates is observed. According to the distances, the glutamates adopt two different conformations (Figure 30 panel C and D), which will be denoted as: closed (panel C) and open state (panel D).

Furthermore, a water molecule can be observed in the surrounding of Glu103 and Glu123. It occupies different positions in the open and closed state (Figure 30 panel C and D). In the open state the water molecule is closer to both glutamates (2.7 Å, distance measured between the oxygen atom of the water molecule and O $\epsilon$ 1 of E103 and E123). The observed water molecule in the open state could correspond to the one described by Kasaragod *et al.*[122] and thus might be important for the interaction and hydrogenation reaction of the active site I with water.

For a further look into the evolution of the different states referring to the distances assigned to the open and closed state, the 2000 frames of the distance plot were

grouped into subsections (I – V for the open state and I – IV for the closed state) to describe and compare conformational representatives for both states (Figure 31).



*Figure 31: Calculated distance plot for the apoprotein between the O $\epsilon$ 1 oxygen atom of the protonated glutamate 103 and deprotonated glutamate 123. The distance is measured over the whole 200 ns accelerated MD trajectory for the apoprotein (PDB: 3ZW8). The blue boxes represent the open states subdivided into the sections I-V, referring to the distance between the two O $\epsilon$ 1 oxygens of both glutamates. The red box represents the closed states subdivided into the sections I-IV, referring to the distance between the two O $\epsilon$ 1 oxygen atoms of both glutamates.*

The subdivision into the different sections, as shown in Figure 31, facilitates the analysis of the two previously identified states. Each section is examined in more detail with respect to possible conformational changes of the glutamates E103 as well as E123. Considering the very high number of possible structural representations (2000 frames in total), two representative structures from each section were selected to illustrate the open and closed states. These represent the structural changes over the whole trajectory, also (Figure 32).



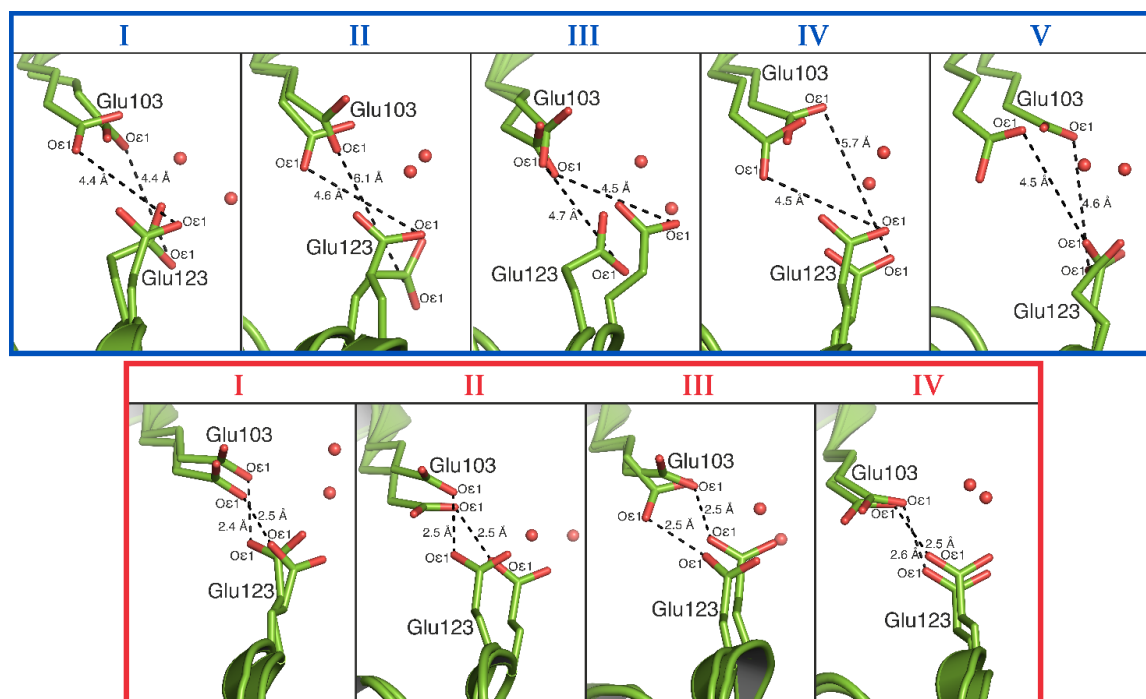


Figure 32: Representation of the glutamates E103 and E123 as stick models and their conformation during the aMD trajectory of the apoprotein. Each section is illustrated by two representatives describing minor deviations but a general agreement for the respective state. The snapshots in the blue box represent each section (I – V) of the open state, as indicated by the distance plot in Figure 31. Each snapshot in the red box illustrates the sections I – IV referring to the closed state of both glutamates as indicated by the distance plot in Figure 31. The red sphere in each picture represents a water molecule. Distances were measured between the oxygen atoms Oε1 of E103 and E123 and are highlighted as dashed lines.

Figure 32 shows on the one hand the conformational differences within respective state (open: blue boxes I – V and closed: red boxes I – IV) and on the other hand the conformational differences between the open and closed state of both glutamates E103 and E123. All snapshots of the open state show a structural recurring distance of 4.5 Å between E103 and E123 and a similar conformational appearance except for minor positional deviations caused by the aMD trajectory. In addition, all the snapshots of the open state show a water molecule between both glutamates, that seems to be in a trapped position. Although during the entire trajectory an exchange between the trapped and the surrounding water molecules takes place, the two glutamates and the water molecule adopt the open conformation repeatedly.

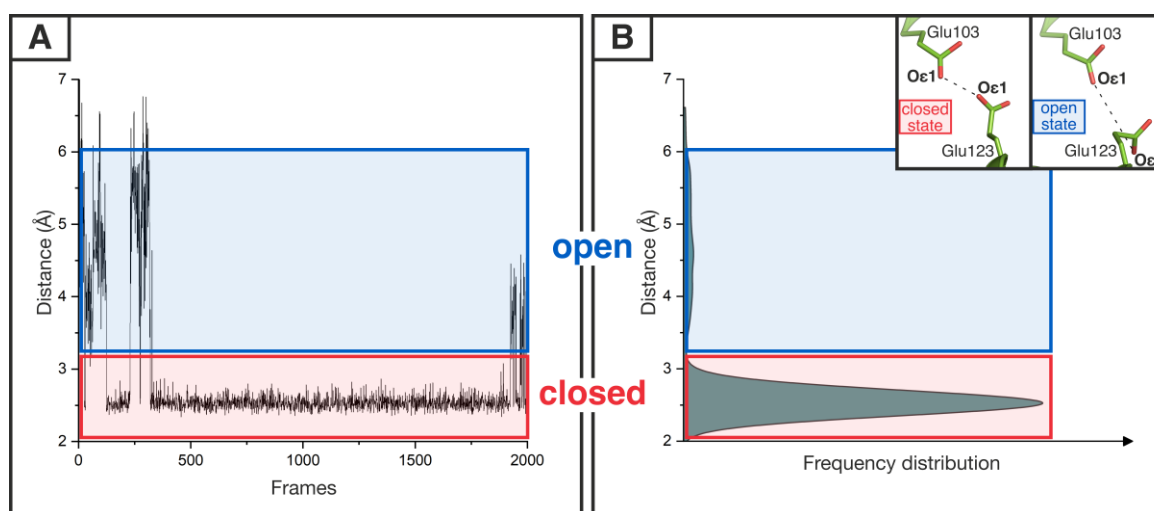
In the closed state, the conformation of the glutamates E103 and E123, which move closer to each other, can be observed repeatedly in the same way as in the open state (shown in Figure 32). Accordingly, the measured distance of both oxygen atoms Oε1 of E103 and E123 is at 2.5 Å. Positional deviations due to the aMD are also present, but a general agreement can also be seen.

An important result is the position of the water molecule. It is also in the closed state in hydrogen bond distance to both glutamates but is no longer located between the two amino acids, instead it is transient located outside the glutamate plane. Thus, the water molecule is not in a trapped position anymore, but this position could be understood as a temporary position before the glutamates move apart from each other and the water molecule can "enter" for the hydrogenation reaction.

Both states (open and closed) shown here, support the importance of both glutamates described by Kasaragod *et al.*[122] and the associated interaction with catalytic water molecule which seems to be necessary. Furthermore, the results indicate that the glutamates E103 and E123 seem to not only have one fixed position but alternating between the open and closed conformation, which are defined in this thesis. This behavior seems to be fundamentally important to ensure reactivity in the first active site.

#### **4.4.2 MFE-1 complexed with 3S-hydroxydecanoyl-CoA**

The 3S-hydroxydecanoyl-CoA substrate was embedded into the first active site, and the behavior of both glutamates was analyzed in an analogous manner. For this, the distance of the two oxygen atoms (Oε1) of Glu103 and Glu123 was measured for the whole aMD trajectory and plotted in Figure 33.



**Figure 33: A** Calculated distance plot between the Oε1 oxygen atom of the protonated glutamate 103 and deprotonated glutamate 123 for the MFE-1 with an embedded 3S-hydroxydecanoyl-CoA substrate in the first active site (domain A). The distance is measured over the whole 200 ns accelerated MD trajectory. The blue box represents the open state, referring to the distance between the two Oε1 oxygen atoms of both glutamates. The red box represents the closed state, referring to the distance between the two Oε1 oxygen atoms of both glutamates. **B** Frequency distribution plot of the different distances between the Oε1 oxygen atoms of both glutamates. The distribution depends on the counts of each distance. Two main peaks are visible, referring to the open (blue box) and closed (red box) state. The two small pictures in the top right corner in panel B represent a reminder of the structural representation of the closed and open state based on the crystal structure of MFE-1 (PDB: 3ZW8).

Analog to the apoprotein, the crystal structure of the MFE-1 was used as reference (distance between Oε1 of the glutamates E103 and E123 is 5.2 Å) with the 3S-hydroxydecanoyl-CoA substrate embedded into the first active site within domain A. The distance plot (Figure 30 panel A) over the whole aMD trajectory for the MFE-1 complex was investigated likewise the apoprotein. Here, also in combination with the frequency distribution plot (panel B) different states, the open and closed state, can be observed. As described in chapter 4.3.3, this plot visualizes the quantity of the distance (between Oε1 of glutamates 103 and 123) distributed over the entire trajectory. Similar to the apoprotein in chapter 4.3.3, a narrow peak around 2.6 Å is present, representing the closed state (red box). In contrast to that, the open state (blue box) with an average

distance of 4.9 Å is much less pronounced and very broad compared to the apoprotein. Therefore, the open state seems to be much less present in the MFE-1 complex with the embedded 3S-hydroxydecanoyl-CoA substrate. Nevertheless, two conformational states of the two glutamates are still distinguishable. These conformations are similar to the ones observed for the apoprotein.

Furthermore, surrounding water molecules occupy different positions in the open and closed state as it was described for the apoprotein before in section 4.3.3. Thus, the open state could also correspond to the state described by Kasaragod *et al.*[122], that seems to be important for the interaction with water at the active site I and for the hydrogenation reaction.

For a more detailed investigation of the MFE-1 complex with 3S-hydroxydecanoyl-CoA substrate, the two states were grouped into subsections (I – III for the open state and I – II for the closed state) to describe and compare conformational representatives for both states (Figure 34).

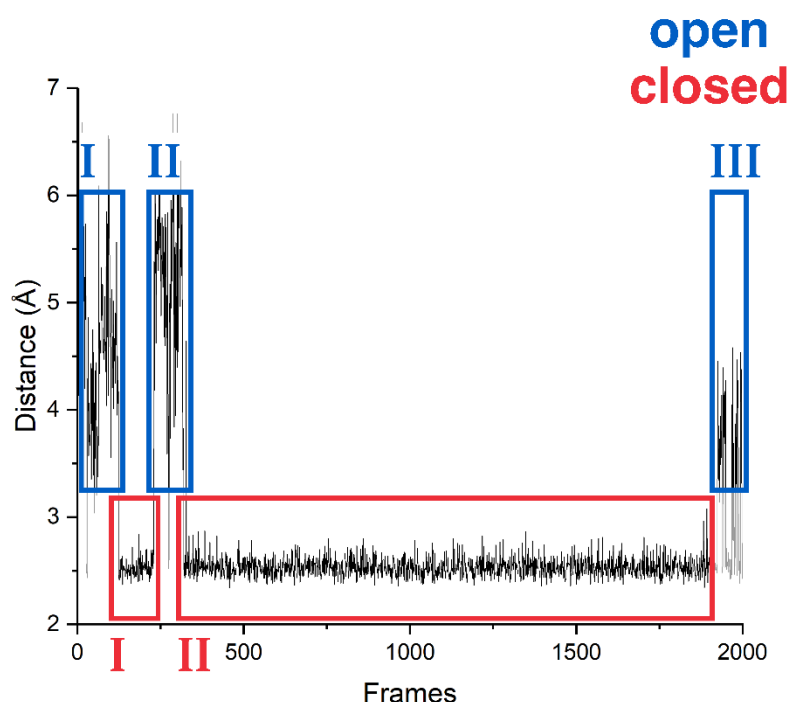


Figure 34: Calculated distance plot for the MFE-1 with 3S-hydroxydecanoyl-CoA substrate present in the first active site in domain A measured between the O $\epsilon$ 1 oxygen atoms of the protonated glutamate 103 deprotonated glutamate 123. The distance is measured over the whole 200 ns accelerated MD trajectory for the MFE-1 complex. The blue boxes represent the open states subdivided into sections I – III, referring to the distance

between the two O $\epsilon$ 1 oxygen atoms of both glutamates. The red box represents the closed states divided into sections I – II, referring to the distance between the two O $\epsilon$ 1 oxygen atoms of both glutamates.

The subdivision into the different sections facilitates the analysis of the two previously identified states but shows less subsections for the open and closed state with the 3S-hydroxydecanoyl-CoA substate present in the first active site compared to the apoprotein. While for the apoprotein, the open state could be divided into five sections, in this complex, there are three subunits in the open state. In the closed state, the number of sections is two, while in the apoprotein four subdivisions can be distinguished. This suggests the presence of a substrate in the first active site stabilizes the MFE-1 compared to the apoprotein in this region.

Each section is examined in more detail with respect to possible conformational changes of the glutamates E103 as well as E123 and represented by two representative structures from each section illustrating the open and closed states (Figure 35).

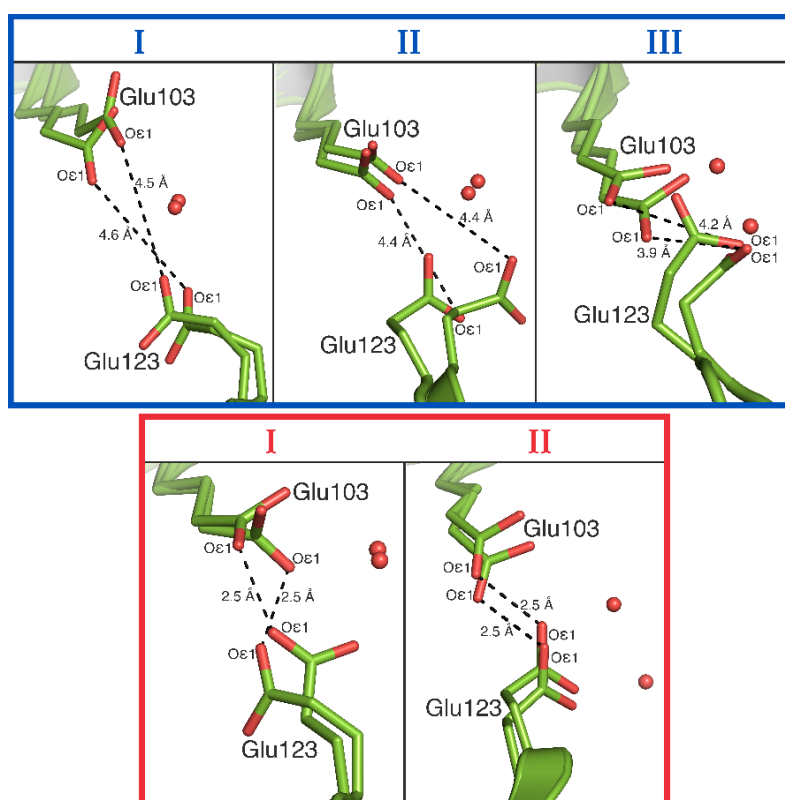


Figure 35: Representation of the glutamates E103 and E123 as stick models and their conformation during the aMD trajectory of the MFE-1 with 3S-hydroxydecanoylCoA

*substrate embedded in the first active site. Each section is illustrated by two representatives describing minor deviations but a general agreement for the respective state. The snapshots in the blue box represent each section (I – III) of the open state, as indicated by the distance plot in Figure 34. The snapshots in the red box illustrate sections I and II referring to the closed state of both glutamates, as indicated by the distance plot in Figure 34. The red spheres in each picture represent a water molecule. Distances between the oxygen atoms Oε1 of E103 and E123 were measured and are highlighted as dashed lines.*

Figure 35 shows a similar behavior of the glutamates E103 and E123 in MFE-1 with 3S-hydroxydecanoyl-CoA substrate in the first active site compared to the apoprotein. Likewise the apoprotein, where the occurrence of the open and closed state describe an alternating process, in the MFE-1 complex with the substrate in the first active site this alternating conformational pattern is also observable (open: blue box, sections I – III and closed state: red box, sections I and II). In addition, the distances between the Oε1 oxygen atoms of E103 and E123 with 4.0 – 4.5 Å in the open and 2.5 Å in the closed state are in well agreement with the distances in the apoprotein, despite minor positional deviations.

Concerning water molecules surrounding both glutamates, a similar behavior as in the apoprotein can be observed. In the open state, the water molecule enters the plane in between the E103 and E123 and thus comparable to the behavior of the apoprotein the water molecule is in a trapped position. In the closed state, the water molecule is outside this plane and is found close to the glutamates. From this temporary position the water molecule can enter the plane after the conformational change of both glutamates towards the open state. Because of that, the hydrogenation reaction is facilitated.

Overall, these results support likewise the apoprotein the state described by Kasaragod *et al.*[122]. Moreover, the presence of the 3S-hydroxydecanoyl-CoA substrate does not seem to prevent the alternation of the open and closed state but rather influences the equilibrium to favor the closed state.

#### 4.4.3 MFE-1 holoprotein

In this section, the MFE-1 is additionally complexed with NAD<sup>+</sup> in the second active site in domain C. As described by Kasaragod *et al.*[121], this cofactor is necessary for the dehydrogenation reaction in the second active site. Moreover, the insertion of NAD<sup>+</sup> serves to analyze whether a similar behavior of both glutamates (E103 and E123) can be observed for the holoprotein as well. For this, the distance between O $\epsilon$ 1 of Glu103 and Glu123 during the entire aMD trajectory was plotted in Figure 36.

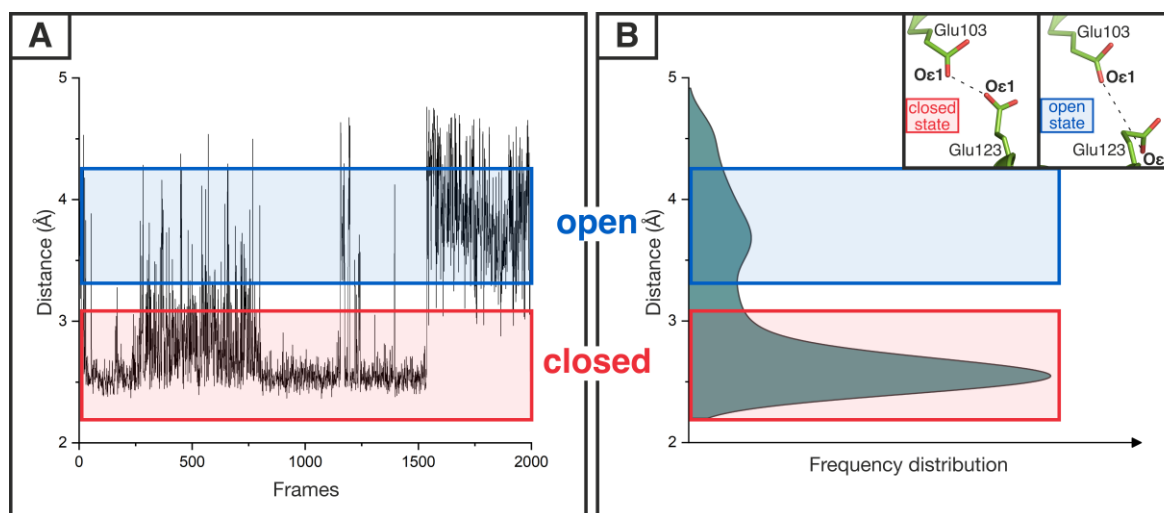


Figure 36: **A** Calculated distance plot between the O $\epsilon$ 1 oxygen atoms of the protonated glutamate 103 and the deprotonated glutamate 123 for the MFE-1 with an embedded 3S-hydroxydecanoyl-CoA substrate in the first active site (domain A) and NAD<sup>+</sup> present in the second active site (domain C). The distance is measured over the whole 200 ns accelerated MD trajectory. The blue box represents the open state and the red box represents the closed state. **B** Frequency distribution plot of the different distances between the O $\epsilon$ 1 oxygen atoms of both glutamates. The distribution depends on the counts of each distance. Two main peaks are visible, referring to the open (blue box) and closed (red box) state. The two small pictures in the top right corner in panel B represent a reminder of the structural representation of the closed and open state based on the crystal structure of MFE-1 (PDB: 3ZW8).

Analog to sections 4.3.3 and 4.3.4, the crystal structure was used as a reference (5.2 Å between O $\epsilon$ 1 of E103 and E123). In addition to the presence of the 3S-hydroxydecanoyl-

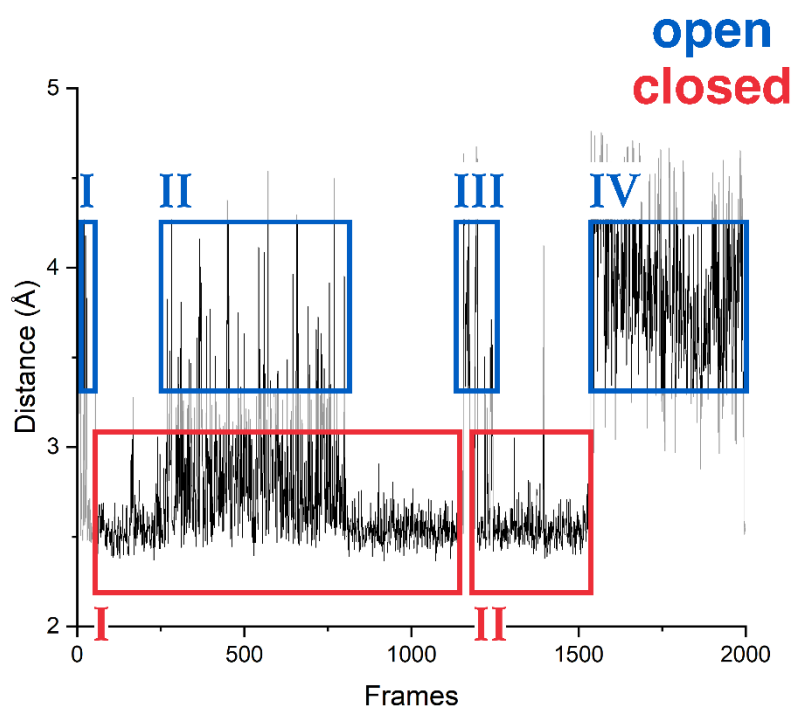
CoA substrate in the first active site, the second active site is complexed with NAD<sup>+</sup>. In this MFE-1 holoprotein, the distance was also measured over the entire aMD trajectory and plotted (Figure 36, panel A). Like the apoprotein and the MFE-1 complex with 3S-hydroxydecanoyl-CoA substrate, two states (open and closed) can be identified depending on the distance between both Oε1 oxygen atoms of Glu103 and Glu123 for the holoprotein. The frequency distribution of both states (panel B) visualizes and quantify the probability of the open and closed state.

Consistent with the apoprotein in chapter 4.3.3 and the MFE-1 complex with 3S-hydroxydecanoyl-CoA in chapter 4.3.4, two peaks can be identified. Compared to the previous MFE-1 systems, a shift in the distribution of each state can also be observed. The holoprotein reveals a higher quantity and a narrower peak of the open state as for the MFE-1 complex with 3S-hydroxydecanoyl-CoA, but still shows less counts and a broader distribution than the apoprotein. Despite the shift in the frequency distribution, both states (open: blue box and closed state: red box) occur in an alternating pattern. Here, the distance of the closed state is 2.6 Å, whereas, in the open state the distance is around 3.7 Å.

Furthermore, as described in sections 4.3.3 and 4.3.4, depending on the respective state a change in the position of the surrounding water molecule can also be observed in for the holoprotein. As observed before also in this MFE-1 complex with NAD<sup>+</sup> present in the second active site, the open state supports the state described by Kasaragod *et al.*[122].

For a more detailed investigation the distance plot was divided into subsections (I – IV for the open and I – II for the closed state).





*Figure 37: Calculated distance plot for the MFE-1 complexed with 3S-hydroxydecanoyl-CoA in the first active site in domain A and NAD<sup>+</sup> present in the second active site in domain C. Distance is measured between the O $\epsilon$ 1 oxygen atoms of the protonated glutamate 103 and the deprotonated glutamate 123. The distance is measured over the whole 200 ns accelerated MD trajectory for the MFE-1 complex. The blue boxes represent the open states subdivided into the sections I – IV and the red box represents the closed state subdivided into sections I – II, referring to the distance between the two O $\epsilon$ 1 oxygen atoms of both glutamates.*

Based on this subdivision, it can be observed that the alternation between the open and closed state is different compared to the apoprotein and the MFE-1 complex with 3S-hydroxydecanoyl-CoA in the first active site. Here, for the holoprotein four subsections can be determined for the open state as compared to five subsections for the apoprotein and three for the MFE-1 complex with 3S-hydroxydecanoyl-CoA. The closed state reveals two subsections for the holoprotein as for the MFE-1 complex with 3S-hydroxydecanoyl-CoA, in contrast to four sections for the apoprotein. These findings suggest, as described in section 4.3.4, that the presence of a substrate in the first active site seems to stabilize domain A. According to the frequency distribution the holoprotein favors the closed state likewise the MFE-1 complex with 3S-hydroxydecanoyl-CoA. In addition, the

presence of NAD<sup>+</sup> in the second active site seems to cause a more frequent alternation between the closed and open state compared to the MFE-1 complex (section 4.3.4), where NAD<sup>+</sup> is missing.

In order to investigate structural changes of Glu103 and Glu123 in more detail as described in sections 4.3.3 and 4.3.4, two representative snapshots from each subsection for the open and closed state are illustrated in Figure 38.

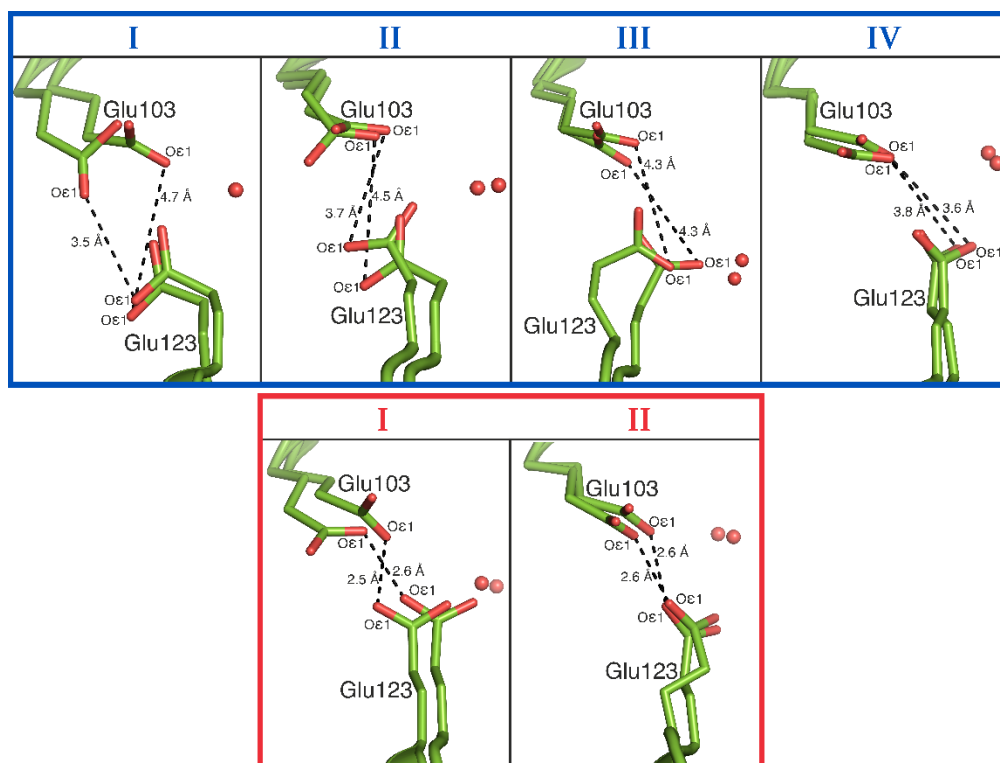


Figure 38: Representation of the glutamates E103 and E123 as stick models and their structural adaption during the aMD trajectory of the MFE-1 with 3S-hydroxydecanoyl-CoA substrate embedded in the first active site and NAD<sup>+</sup> present in the second active site. The snapshots in the blue box represent the section (I – IV) of the open state, as indicated by the distance plot in Figure 37 and the snapshots in the red box illustrates sections I and II referring to the closed state. Each section is illustrated by two representatives describing minor deviations but a general agreement for the respective state. The red spheres in each picture represent a water molecule. Distances between the oxygen atoms Oε1 of E103 and E123 were measured and are highlighted as dashed lines.

Figure 38 shows the conformational differences of E103 and E123 between the open and closed state. This illustrates that the structural changes of both glutamates also occur in

the presence of NAD<sup>+</sup> in the second active site. Thus, the open and closed state also underwent an alternating change in the MFE-1 holoprotein. The open state represented by sections I – IV (blue box) reveal comparable distances between the oxygen atoms Oε1 of E103 and E123 around 3.8 – 4.5 Å to the apoprotein and MFE-1 complexed with 3S-hydroxydecanoyl-CoA. The same applies for the closed state, sections I and II (red box), with a distance between the oxygen atoms Oε1 of E103 and E123 of 2.5 Å.

The structural conformational changes of the two investigated glutamates E103 and E123 are very similar for all three MFE-1 constructs.

The surrounding water molecule again shows a very similar behavior as in the previously studied MFE-1 constructs. In the closed state, the water is close to the glutamates but not within the glutamate plane. The open state, however, reveals the water molecule position in the plane between the two glutamates, as observed before for the apoprotein and the MFE-1 construct with 3S-hydroxydecanoyl-CoA in the first active site. In accordance with the state described by Kasaragod *et al.*[122], the glutamates are interacting with a water molecule that seems to be important for the hydrogenation reaction in the open state. The alternation between the open and closed states of glutamate E103 and E123 is also enabled even with NAD<sup>+</sup> present in the second active site. It can be observed that the frequency distribution is different compared to the two other MFE-1 constructs, with a shifted equilibrium towards the closed state that is more pronounced than the apoprotein but less shifted as the MFE-1 construct with 3S-hydroxydecanoyl-CoA in first active site.

## **4.5 Principal component analysis**

### **4.5.1 MFE-1 apoprotein**

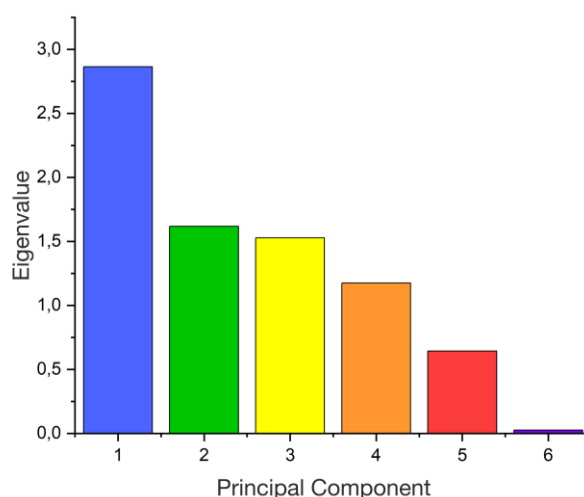
The results of sections 4.3.3 to 4.3.5 show a distance-dependent conformational change of glutamates E103 and E123. These changes lead to the open and closed states as defined in this thesis. However, the trigger for this alternating process is not known so far and will be further investigated by principal component analysis (PCA) in this section. To identify a possible relation[430], parameters such as distances and dihedrals between specific atoms and molecules were selected and measured over the entire trajectory for each MFE-1 system (data table not shown). Initially, this process was performed for the apoprotein, resulting in the usage of a total of six parameters: dihedral change of E103

and E123, distance between O $\epsilon$ 1-O $\epsilon$ 1 as well as O $\epsilon$ 2-O $\epsilon$ 2 oxygen atoms of E123 and E103, distance of water molecules to glutamate E103 and E123. The resulted eigenvalues are given in Table 2.

*Table 2 Calculated eigenvalues of the correlation matrix of the apoprotein of MFE-1.*

Principal Component	Eigenvalue	Variance in %
1	2.86335	36.5
2	1.61636	20.6
3	1.52627	19.4
4	1.17378	15
5	0.64237	8.2
6	0.02554	0.3

Each eigenvalue describes the amount of variance in the principal components. Thus, these are arranged accordingly in descending order, which leads to the following graph (Figure 39).



*Figure 39 Calculated eigenvalues are displayed as a histogram. The arrangement is from left to right, starting with the largest eigenvalue and the smallest eigenvalue on the right.*

From the arrangement of the eigenvalues according to the PCA, the order of the eigenvalues is assigned to principal components (PC) 1 – 6 based on their variances. The

eigenvectors associated with the eigenvalues and thus the PCs are represented in the following table (Table 3).

*Table 3: Loading matrix for all six principal components of the MFE-1 apoprotein; Dihed: dihedral change of E103 or E123; Oε1-Oε1: distance between E103 and E123 measured between Oε1 oxygen atoms; Water: distance of water molecules to Oε1 of glutamate E103 and E123; Oε2-Oε2: distance between E103 and E123 measured between Oε2 oxygen atoms.*

Variables	PC1	PC2	PC3	PC4	PC5	PC6
Dihed103	<b>-0.54390</b>	-0.16652	-0.04818	0.00296	<b>-0.82056</b>	0.02778
Dihed123	-0.00262	0.17439	-0.10948	<b>0.97723</b>	-0.02213	0.04590
Oε1-Oε1	<b>-0.58567</b>	-0.03288	0.06723	0.05381	<b>0.36686</b>	<b>-0.71686</b>
Water198	0.11819	<b>-0.93018</b>	<b>-0.29277</b>	<b>0.13667</b>	0.12792	-0.00563
Water84	-0.09071	<b>0.27141</b>	<b>-0.94401</b>	<b>-0.15272</b>	0.05966	-0.00780
Oε2-Oε2	<b>-0.58218</b>	-0.04326	0.06552	-0.00976	<b>0.41432</b>	<b>0.69507</b>

The loading matrix of the MFE-1 apoprotein shows the relation between the selected variables. The correlation values lie between -1 and +1. Negative values indicate an anti-correlation and positive values a correlation. Values around 0 represent a non-correlated motion. Significant values ( $> 0.1$ ) are broadly distributed over all principal components and highlighted in bold and light-blue shaded in Table 3.

To start with, PC1 and PC5 are similar, both reveal an anti-correlated (PC1) and correlated (PC5) motion of the dihedral change of Glu103 and distance change between Oε1 and Oε2 oxygen atoms of Glu103 and Glu123. PC4 describes a correlation of the Glu123 dihedral to Water198 and an anti-correlation to Wat84. Therefore, the described correlations of these three principal components indicate that the dihedral change of E103 is independent from its distance to the water molecules. Additionally, the dihedral changes of E103 and E123 seem not to be structural dependent from each other.

Next, PC2 and PC3 describe an anti-correlation between the water molecules among themselves. Moreover, PC6 illustrates the anti-correlated relation of the changing distances between Oε1- Oε1 and Oε2- Oε2 of the two glutamates.

Based on all six principal components, water molecules seem to affect structural changes of Glu123 slightly more than Glu103. Furthermore, both glutamates seem to undergo structural changes independently from each other.

Overall, the PCA results of the MFE-1 apoprotein seem to indicate a high structural flexibility of Glu103 and Glu123. This flexibility seems to be accompanied by an interaction to a water molecule. Considering the state of the first active site described by Kasaragod *et al.*[122] where a water molecule is proposed to be necessary, this interaction with the water network could be a trigger for the structural changes.

#### 4.5.2 MFE-1 complexed with 3S-hydroxydecanoyl-CoA

As the principal component analysis showed in section 4.4, water molecule (as suggested by Kasaragod *et al.*[122]) seems to be essential as a trigger for the conformational change of glutamate E103 and E123. To investigate this in more detail, PCA was also performed in the presence of the substrate 3S-hydroxydecanoyl-CoA in the first active site of MFE-1. The same parameters as described in section 4.5.1 were used, and the following eigenvalues and principal components were calculated first (Table 4 and Figure 40).

*Table 4: Calculated eigenvalues of the correlation matrix for MFE-1 with embedded 3S-hydroxydecanoylCoA.*

Principal Component	Eigenvalue	Variance in %
1	5.45952	37.4
2	3.85865	26.5
3	2.31398	15.9
4	1.16418	8.0
5	0.98187	6.7
6	0.80429	5.5

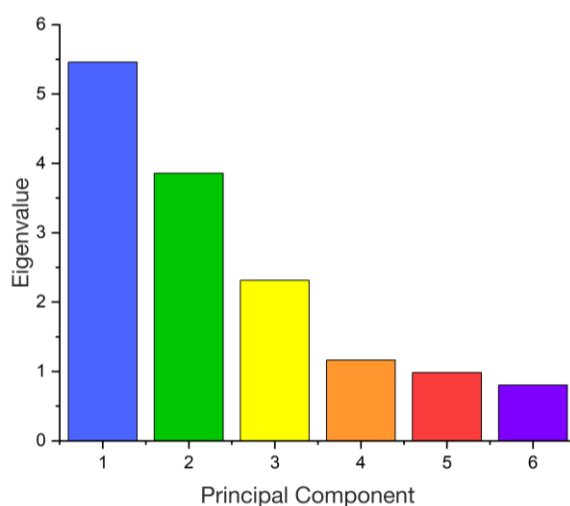


Figure 40 Calculated eigenvalues are displayed as a histogram. The arrangement is from left to right, starting with the largest eigenvalue and the smallest eigenvalue on the right.

Analogous to the apoprotein, for the MFE-1 complex with 3S-hydroxydecanoyl-CoA, the eigenvalues are assigned to principal components and their associated eigenvectors are shown in Table 5.

Table 5: Loading matrix for all six principal components of MFE-1 complexed with 3S-hydroxydecanoyl-CoA; Dihed: dihedral change of E103 or E123; Oε1-Oε1: distance between E103 and E123 measured between Oε1 oxygen atoms; Water: distance of water molecules to Oε1 of glutamate E103 and E123; Oε2-Oε2: distance between E103 and E123 measured between Oε2 oxygen atoms. Significant values are highlighted in bold and light-blue shaded.

Variables	PC1	PC2	PC3	PC4	PC5	PC6
Dihed103	-0.08696	<b>-0.26006</b>	<b>-0.94630</b>	-0.03294	0.02492	0.16615
Dihed123	<b>-0.27997</b>	<b>-0.78336</b>	<b>0.31742</b>	0.08346	0.04035	<b>0.44565</b>
Oε1-Oε1	<b>-0.61011</b>	-0.22512	-0.00704	-0.17703	-0.10285	<b>-0.73151</b>
Water124	0.33195	-0.22505	0.03795	<b>-0.54252</b>	<b>0.71551</b>	-0.17727
Water35	0.31074	<b>-0.17566</b>	0.03214	<b>-0.64171</b>	<b>-0.67654</b>	0.04500
Oε2-Oε2	<b>-0.57884</b>	<b>0.43188</b>	0.03509	<b>-0.50444</b>	0.13227	<b>0.45301</b>

Table 5 shows correlation for MFE-1 complexed with 3S-hydroxydecanoyl-CoA. The contribution of the dihedral of Glu123 to the eigenvectors has significantly increased.

This is illustrated by PC1 and PC6, which suggest a higher dependence of the Glu123 conformation on the distance between the oxygen atoms O $\epsilon$ 1-O $\epsilon$ 1 as well as O $\epsilon$ 2-O $\epsilon$ 2.

In accordance, PC2 and PC3 illustrate structural correlations of conformational changes of Glu103 and Glu123 in dependence on the distances between the oxygen atoms O $\epsilon$ 1-O $\epsilon$ 1 and O $\epsilon$ 2-O $\epsilon$ 2. Moreover, these two principal components show a mutual coupled structural change of glutamate 103 and 123 (significant values for E103 and E123 dihedral changes). In addition, especially PC2 describes a slightly higher correlation of the O $\epsilon$ 1-O $\epsilon$ 1 and O $\epsilon$ 2-O $\epsilon$ 2 distances to water molecules.

The correlation of the water molecules among each other but also to the O $\epsilon$ 1 and significantly O $\epsilon$ 2 oxygen atom distances are described by PC4 and PC5.

Taken together, in presence of 3S-hydroxydecanoyl-CoA substrate in the active site I of MFE-1 the mutual dependence of the conformational changes of E103 and E123 seem to increase. Thus, a structural stability within the first active site appears to be favored due to the presence of the substrate.

Again, the importance of the water interaction with the glutamates E103 and 123 is reflected by these results and is also described by Kasaragod *et al.*[122].

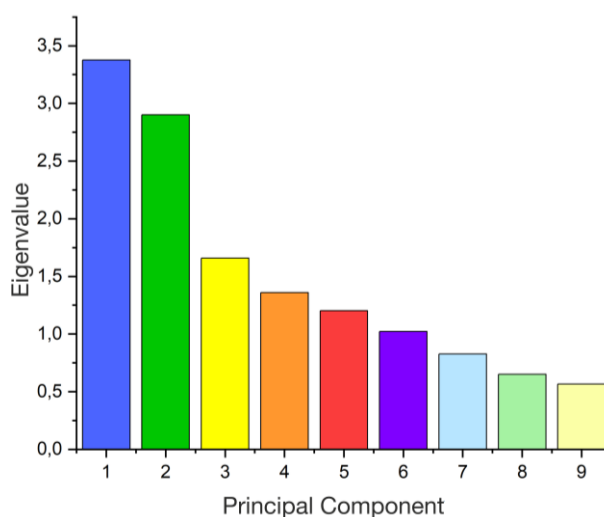
#### **4.5.3 MFE-1 holoprotein**

Section 4.4 (apoprotein) and 4.5 (MFE-1 complex with 3S-hydroxydecanoyl-CoA in the first active site) reveal a water molecule as described by Kasaragod *et al.*[122], that seem to be the trigger of the conformational changes of glutamate E103 and E123. In the following, for the MFE-1 holoprotein, that is additionally complexed with NAD<sup>+</sup> in the second active site, it was investigated whether the presence of NAD<sup>+</sup> has an influence on the structural behavior of MFE-1 concerning the two glutamates and the water interactions or not. The same parameters were used as in sections 4.5.1 and 4.5.2, but three more water molecules were added to check their influence in more detail, and the following eigenvalues and principal components were calculated (Table 6 and Figure 41).



*Table 6: Calculated eigenvalues of the correlation matrix for MFE-1 with embedded 3S-hydroxydecanoylCoA in the first active site and NAD<sup>+</sup> present in the second active site.*

Principal Component	Eigenvalue	Variance in %
1	3.37591	24.9
2	2.90059	21.4
3	1.65860	12.2
4	1.35911	10.0
5	1.20153	8.9
6	1.02147	7.5
7	0.82732	6.1
8	0.65097	4.8
9	0.56538	4.2



*Figure 41: Calculated eigenvalues are displayed as a histogram. The arrangement is from left to right, starting with the largest eigenvalue and the smallest eigenvalue on the right.*

For the holoprotein and analogous to the other MFE-1 models, the eigenvalues are assigned to principal components and their associated eigenvectors (Table 7).

*Table 7: Loading matrix for all nine principal components of MFE-1 holoprotein; Dihed: dihedral change of E103 or E123; Oε1-Oε1: distance between E103 and E123 measured between Oε1 oxygen atoms; Water: distance of water molecules to Oε1 of glutamate E103 and E123; Oε2-Oε2: distance between E103 and E123 measured between Oε2 oxygen atoms. Significant values are highlighted in bold and light-blue shaded.*

Variables	PC1	PC2	PC3	PC4	PC5
Dihed103	<b>-0.39133</b>	-0.16451	-0.02572	<b>0.68438</b>	<b>-0.42433</b>
Dihed123	-0.14584	-0.01438	-0.20532	<b>0.49133</b>	<b>0.53618</b>
Oε1-Oε1	<b>-0.58141</b>	0.03499	-0.00140	<b>-0.48294</b>	0.08266
Water562	-0.00499	-0.01608	<b>-0.91462</b>	-0.13769	-0.26332
Water878	0.13086	<b>-0.51579</b>	-0.08084	0.02552	<b>0.50864</b>
Water879	0.14576	<b>-0.57450</b>	-0.14828	-0.12451	0.03909
Water990	<b>0.03394</b>	<b>-0.51249</b>	<b>0.29899</b>	-0.07633	<b>-0.36093</b>
Water1557	<b>0.38664</b>	<b>0.33514</b>	0.01141	0.11224	0.08185
Oε2-Oε2	<b>-0.54634</b>	0.01309	0.05082	-0.05857	0.24300

Variables	PC6	PC7	PC8	PC9
Dihed103	<b>0.38244</b>	0.05619	0.12497	-0.07519
Dihed123	<b>-0.54333</b>	<b>0.19435</b>	0.24866	0.10497
Oε1-Oε1	0.15021	<b>0.21507</b>	<b>0.59312</b>	-0.00168
Water562	<b>-0.04608</b>	-0.11984	-0.02571	0.24027
Water878	0.46339	<b>-0.44659</b>	0.15331	0.11503
Water879	<b>0.06404</b>	<b>0.69357</b>	-0.19893	-0.29147
Water990	<b>-0.35846</b>	-0.03211	0.15648	<b>0.59723</b>
Water1557	0.41692	<b>0.45730</b>	0.16940	<b>0.55416</b>
Oε2-Oε2	0.11254	0.07870	<b>-0.67363</b>	<b>0.40450</b>

The dependence of the dihedral change to the Oε1-Oε1 and Oε2-Oε2 distances to Glu103 and Glu123 is described by PC1 and PC7, respectively. Moreover, PC1 and PC7 illustrate the correlation of the dihedral change to the distances to water molecules. However, these two principal components do not indicate a coupling of the structural changes of E103 and E123 among each other.

PC8 and PC9 describe the correlation of the Oε1 and Oε2 oxygen atom distances among each other but also to the water molecules.

When considering PC4, PC5 and PC6 a significant correlation between the dihedral changes of E103 and E123 are observable. In addition, these three principal components show a dependence of the conformational changes of the two glutamates on the distances to water molecules.

As observed before in section 4.4, PC2 and PC3 illustrate the correlation of the water molecules among each other.

In summary, the MFE-1 holoprotein shows a similar behavior as the apoprotein and MFE-1 complexed with 3S-hydroxydecanoyl-CoA. Based on the results for the holoprotein which shows a higher degree of coupling between the conformational changes of the two glutamates E103 and E123, but also to the water distances, a significant interdependence can be suggested, that is more pronounced than for the other two MFE-1 constructs. Thus, the presence of NAD<sup>+</sup> seems to enhance the structural mutual dependence of glutamate conformational changes, that is in close interplay with water molecules. As mentioned before, the interaction with water and its important role as a trigger is also suggested by Kasaragod *et al.*[122].

## 4.6 Discussion

The MFE-1 is an important enzyme for the fatty acid metabolism catalyzing an isomerization, hydrogenation and following dehydrogenation reaction[121, 122]. The hydratase reaction takes place within the first active site. Comparing MFE-1 with monofunctional hydratases reveal two glutamates to play a major role at the active site[121, 122]. Kasaragod *et al.*[122] identified by structural studies and comparison of the crotonase part of MFE-1 that amino acids E103 and E123 seem to be the two essential glutamates for the hydrogenation reaction. However, the respective protonation state of both glutamates or possible structural changes are not known so far. The results of section 4 for the investigated MFE-1 constructs reveals the most probable protonation state and structural changes and thus contribute to the elucidation of the MFE-1 mechanism.

### Protonation state of E103 and E123

In the first active site of MFE-1 are two titratable residues, E103 and E123. Based on the suggestion by Kasaragod *et al.*[121, 122], these two residues are interacting with a water molecule and with the substrate. Therefore, both glutamates seem to be responsible for the catalytic hydrogenation of the substrate in the first active site of MFE-1. Since the knowledge about the correct protonation state of both glutamates remains elusive[121, 122, 423] it would be beneficial to determine the protonation state.

Calculations of the pKa values of these glutamates according to the crystal structure of the MFE-1 apoprotein indicate that E103 carries the proton and E123 is deprotonated.

The calculations over the entire 200 ns trajectory for the apoprotein confirmed the protonation states that were initially valid for the crystal structure of MFE-1 (PDB: 3ZW8). The pKa values for E103 and E123 remain positive and negative throughout the aMD trajectory, respectively. Thus, during the whole trajectory E103 seems to be protonated and E123 deprotonated.

In addition, this behavior was investigated with the 3S-hydroxydecanoyl-CoA substrate present in the first active site. Although a different chemical environment is prevailing in the first active site due to the presence of the substrate, the results of these pKa value calculations confirm the trend observed so far. Moreover, the presence of this substrate seems to stabilize the protonation state of E103 represented by relatively consistent higher pKa values compared to the apoprotein. Based on these results, a protonated E103 and deprotonated E123 are assumed in this dissertation. These protonation states confirm the starting point with the educt and the final state with the hydrated product of the proposed mechanism for the first active site described by Kasaragod *et al.*[122].

### **Distance depending conformational changes of E103 and E123**

All three examined MFE-1 constructs (apoprotein; MFE-1 with 3S-hydroxydecanoyl-CoA in the first active site; holoprotein) provide a deeper insight into the behavior of the first active site in domain A. All constructs have a protonated E103 and the deprotonated E123. As described by Kasaragod *et al.*[122] these two glutamates and their interaction with water molecules could be of great importance. This assumption can be supported by the obtained results shown in section 4. The distance dependence of both glutamates to each other was investigated for the apoprotein, the MFE-1 construct with 3S-hydroxydecanoyl-CoA, and the holoprotein (with NAD<sup>+</sup> present in active site II). For all three MFE-1 constructs, two different forms of conformational states could be defined, which alternate in their frequency of recurrence. The two states are referred to as open and closed in this dissertation, while the open state is defined by a distance around 4.0 Å between the Oε1 oxygen atoms of E103 and E123 and the closed state by a distance of 2.5 Å. Apart from the different distances which indicate a structural change, the conformations of both amino acids differ significantly from each other. Rotational

changes of the side chains results in the open state and could be essential for the hydrogenation mechanism.

First, for the apoprotein the distribution between these two states is relatively similar (section 4.3.3, Figure 30). In contrast, the MFE-1 construct with the substrate (section 4.3.4, Figure 33) and the holoprotein (section 4.3.5, Figure 36) each tends to favor the closed state, with the open state occurring but with a lower probability. As described by Wierenga *et al.*[121] when crystallizing MFE-1 in its apo form, it is very flexible. The substrate 3S-hydroxydecanoyl-CoA particularly affects domain A. Thus, its presence in the first active site stabilizes domain A and hence, the flexibility of this domain is decreased. Consequently, the frequency of the alternation between the two states is reduced. When NAD<sup>+</sup> is present in the second active site, the open state is more likely than for the MFE-1 construct with the substrate but less likely than for the apoprotein.

This finding suggests that the behavior of both E103 and E123 amino acids, and the associated behavior in the first active site, appears to be intrinsic and may represent a possible key feature of the reaction mechanism for the hydrogenation reaction. Additionally, the presence of a substrate is essential to stabilize the respective state. NAD<sup>+</sup> seems to have a minor influence on the frequency distribution and thus contributes to the stability of the conformational glutamate state. This relation could be based on allosteric communication within the MFE-1, but this question is beyond the scope of this dissertation. To investigate this, further atomistic as well as dynamic investigations and analyses are necessary.

Another important result reveals an additional significant difference between the open and closed state. The results show that the water molecule described by Kasaragod *et al.*[122] seems to play a central role. Accordingly, a water molecule occupies a specific position to interact with the glutamates and thus seems to enable the hydrogenation reaction. First, a water molecule is in a position outside the plane spanned by E103 and E123 in the closed state. Once the conformational change of the two amino acids towards the open state occurs, the water molecule moves between both glutamates and thus can interact with the amino acids (Figure 42).

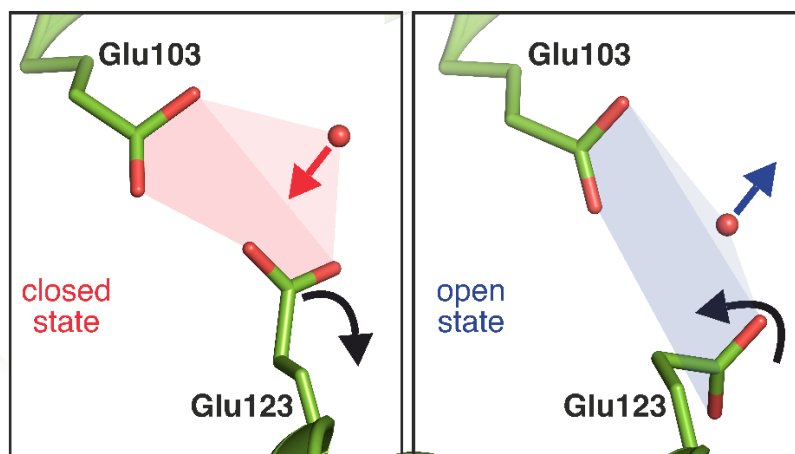


Figure 42: Comparison between open and closed state. **Left:** Structural representation of the closed state of E103 and E123. The red sphere represents a water molecule. The movement of water due to structural rearrangement out of an imaginary plane (illustrated by the red area) of E103 and E123 is indicated by the red arrow. **Right:** Structural representation of the open state. The movement of the water molecule (red sphere) after structural rearrangement of both amino acids into the imaginary plane (illustrated by the blue area) of E103 and E123 is indicated by the blue arrow. The arrows shown in black suggest the structural change of E103.

As seen in Figure 42 and suggested by the colored arrows, the water molecule seems to move in and out of the glutamate plane during the conformational change between the closed and open state. This occurs in all three investigated MFE-1 constructs. The only difference between all constructs concerning the water molecule, is that the water molecule in the apoprotein occupies a position otherwise assigned to the substrate.

Since neither the conformation change of glutamates E103 and E123 nor the knowledge of the movement of water molecules were previously known, the principal component analysis was used to investigate possible relations. PCA confirm the trend for all three MFE-1 constructs that the conformational changes of the glutamates and the water movement seem to be dependent on each other. However, minor deviations are observable for the three constructs. While the apoprotein shows the most variable correlations, the mutual dependency increases due to the presence of a substrate in the first active site. This coupled structural dependence reveals the highest extend for holoprotein (3S-hydroxydecanoyl-CoA in the first active site and NAD<sup>+</sup> in the second active site), which also indicate an allosteric communication within the MFE-1.

Consequently, and based on these results, it can be assumed that the movement of water molecules seems to be the trigger for the structural changes of E103 and E123 or *vice versa*.

To allow proton transfer to the substrate and hydrogenation to occur, the two glutamates apparently need to be a certain distance apart to capture the water molecule. At this position, the water molecule dwells during the reaction before the glutamates release it changing to the closed state and it leaves the first active site. Subsequently, this sequence repeats with a different water molecule. In connection with the alternation between open and closed states and the corresponding glutamate distance, a pattern of a water flow occurs. This behavior seems to be in agreement with the mechanism predicted and activated state described by Kasaragod *et al.*[122] and may play an important role in the overall dynamic behavior of the first active site and thus affecting the overall dynamics of MFE-1.

## 4.7 Summary

The crotonase part of the multifunctional enzyme MFE-1 and the associated first active site catalyze, among others, the hydrogenation of enoyl-CoA molecules during the  $\beta$ -oxidation of fatty acids in fatty acid metabolism. The structural investigations of various MFE-1 constructs described in this chapter focus in particular on the glutamates E103 and E123, which are assumed to be responsible for the hydrogenation reaction. Based on accelerated molecular dynamic simulations (aMD), a distance and conformational change dependence between the two residues was found when comparing MFE-1 constructs. In this context, the so far unknown protonation state of E103 and E123 was investigated and elucidated in more detail using pKa value calculations. Furthermore, this apparent mutual dependence could furthermore be attributed to the water network present in the first active site. The interplay between distance of the two glutamates to each other, the glutamate conformation, and water molecules thus provides better insight to understand the functionality of MFE-1. Moreover, the results support the state suggested by Kasaragod *et al.*[122] that appears to be necessary for the catalytic hydrogenation reaction.





## 5 Results – Docking

### 5.1 Introduction

The multifunctional enzyme type 1 has two active sites. Both sites play a central role in the  $\beta$ -oxidation of fatty acid metabolism[121, 233]. While the isomerization and hydrogenation reaction occur in the first active site, the second active site (Figure 43) is responsible for the dehydrogenation reaction. Since the product from the first active site is converted in the latter reaction, the structural investigation of the binding pocket in terms of functionality and binding modes is of great importance.

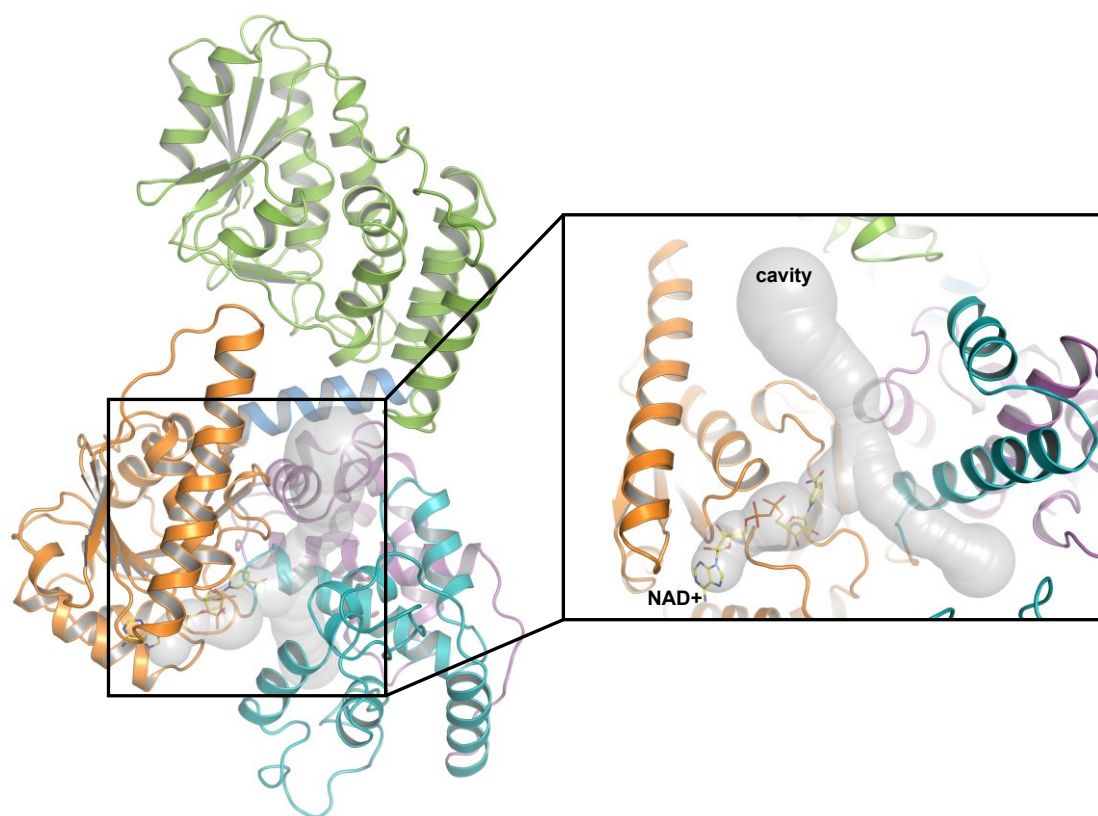


Figure 43: **Left:** Overall fold of the MFE-1 in cartoon representation. Domain A is green, domain B is blue, domain C is orange, and domains D and E are cyan and magenta, respectively. **Right:** Second active site (dehydrogenase active site) with NAD+ highlighted in yellow shown as stick models. The cavities (shown as transparent spheres) of the second active site were calculated with CAVER3.0[431, 432].

In the absence of availability of crystal structures with a substrate physiologically relevant to MFE-1 bound in the second active site, this chapter addresses the issue of modeling a ligand in the second active site. So far, only the research group of Wierenga[121, 122, 233] succeeded in describing corresponding crystal structures of MFE-1 with 3S-hydroxydecanoyl-CoA present in the first active site.

Therefore, this chapter describes how to embed the substrate, that is obtained from the first active site, in the second active site and modeling a 3D structure. Moreover, other potential ligands aside the 3S-hydroxydecanoyl-CoA substrate, which could also serve as future model structures, were investigated.

To model a structure of MFE-1 containing a substrate in the second active site, Virtual Screening (VS) together with docking was used. Molecular docking is to date an established method used especially in the pharmaceutical industry to identify potential ligands, complexes, or interactions of those with specific protein regions. The information derived from this can be used for future studies to elucidate drugs or previously unknown structures. Therefore, in this chapter, the docking method is used to predict a 3D model with 3S-hydroxydecanoyl-CoA present in the second active site. The obtained conformers are ranked, and the most probable 3D model is determined in connection with the binding energy.

## 5.2 Modell setup and preparation

Docking requires a 3D structure as a basis. Since no structure for the MFE-1 complex with a substrate embedded in the second active site was available, homology modeling[433] was used to construct a starting model from a comparable structure. Therefore, as comparable structure for the second active site of MFE-1, the L-3-hydroxyacyl-CoA dehydrogenase extracted from PDB 1FOY was used.

In this dissertation, the homology model (PDB: 1FOY) was used for the second active site due to its very good structural agreement (RMSD of 1.8 Å) with the dehydrogenase active side of MFE-1. For the MFE-1, the PDB 3ZW8 (apoprotein structure of the MFE-1) and the PDB 3ZW9 for the coordinates of the NAD<sup>+</sup> in the second active site of the MFE-1 were used. All these PDBs were imported into the program Chimera[434] and superimposed. This was saved as a new structure file for the following preparation for docking.

The docking procedure required minor preparations of the newly created model. Therefore, it is necessary to describe the region to which the ligand binds and prepare the ligand itself. The workflow for the protein (Protein Preparation Wizard) and ligand preparation (LigPrep) are automated processes and were performed completely with the Schrodinger software Maestro[435].

After successful preparation, docking the ligand in the protein within the second active site could be performed. The substrate 3S-hydroxydecanoyl as the fatty acid (extracted out of PDB 3ZWC), bound to the CoA, was used as the ligand.

Finally, a modified version of CovDock[436] and CovDock-VS[398], which is described in detail in chapter 2.7, referred as PLOP[398, 437-444] (modified by the group of Prof. Matt Jacobson, Department of Pharmaceutical Chemistry, University of California San Francisco, USA) was used for docking the ligand and rank the poses by calculating the relative binding energy (Prime Energy  $\Delta G_{bind}$ ). The CovDock-VS was done additionally to find similar molecules which possibly also serve as ligands.

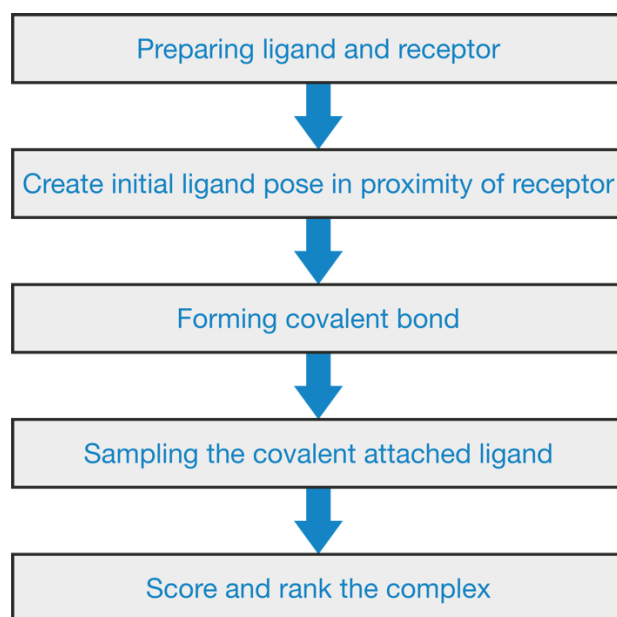


Figure 44: CovDock workflow during the docking process.

### 5.3 Virtual screening

After ligand and protein preparation with the Maestro software, virtual screening (described in detail in chapter 2.7) was performed. Since no crystal structures with possible ligands are available for the second active site, as described in section 5.1, the

so-called receptor-based screening was used to obtain possible ligands. The KEGG (Kyoto Encyclopedia of Genes and Genomes) library[445-447] was used. This database provides the perfect resource for genomic, chemical, and systemic functional information. Outstanding is the additional information provided by gene catalogs, which are linked to other ecosystems with superordinate cell functions. In total, the virtual screening for the second active site of MFE-1 identified 311 compounds with the highest potential. These were ranked by binding energy  $\Delta G_{bind}$ . All components from rank 176 onward have positive energy and were therefore not considered further. The others can be divided into seven main classes, as shown in Table 8.

*Table 8: Number of components received from virtual screening using KEGG library. R = variable rest.*

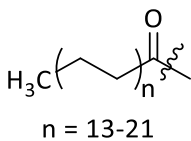
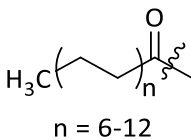
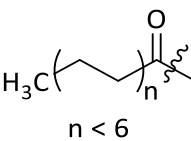
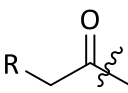
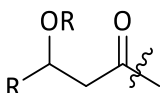
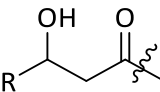
Compound family	Structure pattern	Total number of found compounds
Long-chain fatty acyls	 $n = 13-21$	27
Mid-range fatty acyls	 $n = 6-12$	6
Short-chain fatty acyls	 $n < 6$	10
Acyl-CoAs		44
Oxo-acyl-CoAs		15
3-Hydroxy-acyl-CoAs		11
Miscellaneous	no special motive	63

Table 8 shows that the last four entries in total dominate among the potential ligands found. But according to chapter 1, they seem to play a subordinate role due to the substrate specificity[233]. If the ranking from the result of virtual screening is included, five of these apparently subordinate compounds are nevertheless far ahead in the ranking (Table 9).

*Table 9: Top-ranked molecules of virtual screening of less relevant molecules of MFE-1.*

Overall rank (virtual screening)	Compound family	Compound	Binding energy in $\frac{kcal}{mol}$
1	Acyl-CoA	2-Methylacetoacetyl-CoA	-224.355
2	Miscellaneous	Sinapoyl-CoA	-223.713
3	3-Oxo-Acyl-CoAs	3-Oxo-OPC8-CoA	-216.819
4	3-Oxo-Acyl-CoAs	3-Hydroxy-OPC6-CoA	-211.207
18	3-Hydroxy-Acyl-CoAs	2,6-Dimethyl-5-methylene-3-oxo-heptanoyl-CoA	-202.334

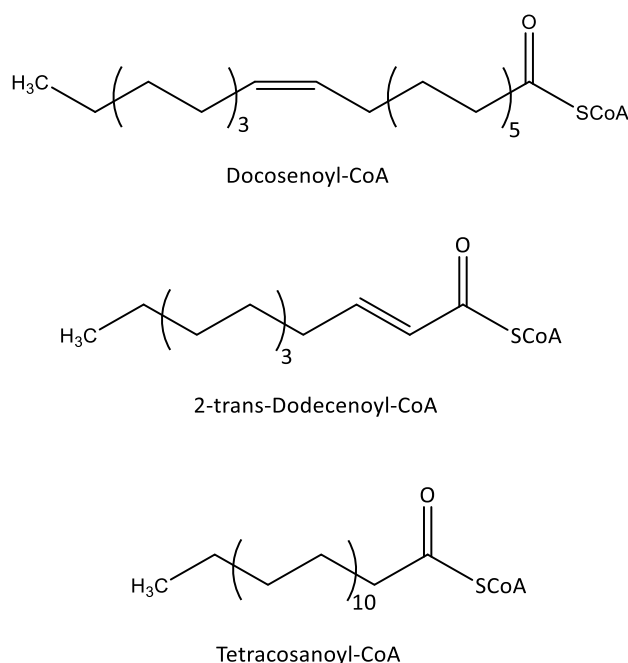
Table 9 shows the top-ranked molecules in the overall ranking determined by the virtual screening. These five molecules are physiologically irrelevant ligands because these substrates do not correspond to the substrate specificity of MFE-1[121]. Looking at all ranked positions 1 – 50, the percentage of fatty acid derivatives, which are physiologically relevant, dominates. As shown in Table 8, the remaining derivatives can be divided into long-chain, mid-range, and short-chain fatty acids that will be describes in more detail in the following.

Table 10: Possible compounds of long-chain fatty acyls delivered by virtual screening.

Overall rank (virtual screening)	Compound	Binding energy in $\frac{kcal}{mol}$
5	Docosenoyl-CoA	-211.137
6	Tetracosaeptaenoyl-CoA	-210.365
7	Tetracosahexaenoyl-CoA	-210.340
8	Docosanoyl-CoA	-210.171
9	Tetracosatetraenoyl-CoA	-209.949
10	4,8,12-Trimethyltridecanoyl-CoA	-207.389
11	(Z)-Icos-5-enoyl-CoA	-206.382
12	(2E)-Hexadecenoyl-CoA	-205.703
13	18-Hydroxyoleoyl-CoA	-205.350
14	(4Z,7Z,10Z,13Z,16Z)-Docosapentaenoyl-CoA	-204.393
15	gamma-Linolenoyl-CoA	-204.308
16	alpha-Linolenoyl-CoA	-203.713
17	(5Z,8Z,11Z,14Z,17Z)-Icosapentaenoyl-CoA	-202.991
19	(11Z)-Hexadec-11-enoyl-CoA	-202.064
24	2-trans-Dodecenoyl-CoA	-201.036
27	Decanoyl-CoA	-199.448
29	9Z-Hexadecenoyl-CoA	-198.962
31	cis,cis-3,6-Dodecadienoyl-CoA	-198.309
34	3,5,7-Trioxododecanoyl-CoA	-197.831
49	(13Z,16Z)-Docosadienoyl-CoA	-193.387
72	(S)-3-Hydroxytetradecanoyl-CoA	-185.636
82	Tetracosanoyl-CoA	-181.143
115	(7Z,10Z,13Z,16Z)-Docosatetraenoyl-CoA	-164.901
142	(3R,6Z,9Z,12Z,15Z,18Z,21Z)-3-Hydroxytetracosahexaenoyl-CoA	-122.446
162, 164	(Very-long-chain)trans-2,3-dehydroacyl-CoA derivatives	-79.399
175	(2E,6Z,9Z,12Z,15Z,18Z)-Tetracosahexaenoyl-CoA	-23.411

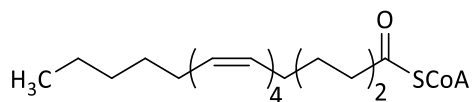
The trend, which is also observable with the molecules predicted to be less relevant (small distances between the binding energies, Table 9), continues with long-chain fatty acids. The first entry of long-chain fatty acids (Docosenoyl-CoA) on the overall rank five has a binding energy of  $-211.137 \frac{\text{kcal}}{\text{mol}}$  (Figure 45), and the 2-trans-Dodecenoyl-CoA (rank 24) still has an energy of  $-201.036 \frac{\text{kcal}}{\text{mol}}$  (Figure 45), which is only a difference of  $10.101 \frac{\text{kcal}}{\text{mol}}$  within 15 steps.

Additionally, an energy difference of  $29.994 \frac{\text{kcal}}{\text{mol}}$  between the first entry of the long-chain fatty acids (Docosenoyl-CoA) and the Tetracosanoyl-CoA (rank 82) entry (Figure 45) can be observed.

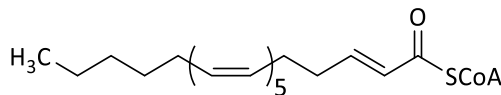


*Figure 45: Structures of Docosenoyl-CoA, 2-trans-Dodecenoyl-CoA and Tetracosanoyl-CoA. All three belong to the long-chain fatty acyls.*

Between the (7Z,10Z,13Z,16Z)-Docosatetraenoyl-CoA with rank 115 (Figure 46, Table 10) with the energy of  $-164.90 \frac{\text{kcal}}{\text{mol}}$ , and the last molecule ((2E,6Z,9Z,12Z,15Z,18Z)-Tetracosahexaenoyl-CoA, Figure 46,  $-23.411 \frac{\text{kcal}}{\text{mol}}$ ), the difference of the individual binding energies is much more significant (with a total number of  $141.490 \frac{\text{kcal}}{\text{mol}}$ ). This suggests a decrease in the relevance of these ligands for the MFE-1 in the second active site.



(7Z,10Z,13Z,16Z)-Docosatetraenoyl-CoA



(2E,6Z,9Z,12Z,15Z,18Z)-Tetracosahexaenoyl-CoA

Figure 46: Structures of (7Z,10Z,13Z,16Z)-Docosatetraenoyl-CoA and (2E,6Z,9Z,12Z,15Z,18Z)-Tetracosahexaenoyl-CoA. Both belong to the long-chain fatty acyls.

A similar trend with small differences in binding energy between the individual ranked positions is also observed for the mid-range fatty acids. The results from the virtual screening are shown in Table 11, arranged according to the overall ranking.

Table 11: virtual screening compounds for mid-range fatty acyls. In bold is the known ligand of MFE-1 present in active site I of the crystal structure PDB 3ZWC.

Overall rank (virtual screening)	Compound	Binding energy in $\frac{kcal}{mol}$
33	(E,E)-Piperoyl-CoA	-198.203
37	(2E)-Octenoyl-CoA	-197.395
39	(2E)-Hexenoyl-CoA	-194.870
47	Octanoyl-CoA	-194.185
<b>58</b>	<b>3S-Hydroxydecanoyl-CoA</b>	<b>-189.707</b>
108	3-cis-Dodecenoyl-CoA	-168.591

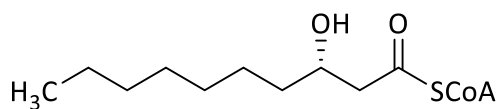
The first entry (E,E)-Piperoyl-CoA has a binding energy of  $-198.203 \frac{kcal}{mol}$  and is listed at position 33 in the overall ranking. In comparison to rank 1 in Table 9 (acyl-CoA) with  $-224.335 \frac{kcal}{mol}$ , there is only a difference of  $26.152 \frac{kcal}{mol}$ . Moreover, both molecules differ significantly in rank with a total difference of 32 places. This confirms the previously observed trend of small differences in the binding energy between the



individual fatty acid molecules. It also shows that enormous differences in structural nature might be decisive for the binding energy and thus for the ranked position.

The most interesting result is marked in bold. This entry the ligand 3S-hydroxydecanoyl-CoA (Figure 47) is the only crystallized ligand of MFE-1 in the first active site (PDB: 3ZWC). So far, as described in section 5.1, the MFE-1 could not be crystallized with the ligand 3S-Hydroxydecanoyl-CoA embedded in the second active site. According to the virtual screening, it has a binding energy of  $-189.707 \frac{\text{kcal}}{\text{mol}}$  and is ranked at 58 of the overall ranking. The relatively high rank of 3S-hydroxydecanoyl-CoA suggests a lower significance compared to the molecules ranked at the top.

However, since the difference in the binding energy compared to the overall rank 1 with  $26.152 \frac{\text{kcal}}{\text{mol}}$  is small, this ligand can be considered as important. Moreover, this ligand has already been crystallized with the MFE-1 structure in the first active site and is therefore relevant.



(S)-3-Hydroxydecanoyl-CoA

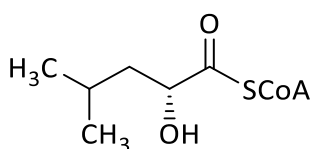
*Figure 47: Structure of (S)-3-Hydroxydecanoyl-CoA. This compound belongs to the group of mid-chain fatty acyls.*

Likewise, the fatty acid derivatives before, the short-chain fatty acids show the same trends (only small energy differences) as the previous molecules.

Table 12: Virtual screening compounds for short-fatty acyls.

Overall ranking (virtual screening)	Compound	Binding energy in $\frac{kcal}{mol}$
30	(R)-2-Hydroxy-4-methylpentanoyl-CoA	-198.584
53	Butanoyl-CoA	-192.228
57	Pentanoyl-CoA	-191.016
68	Hexanoyl-CoA	-187.207
85	3-Methyl-vinylacetyl-CoA	-179.705
86	4-Methylpentanoyl-CoA	-178.845
116	2-Methylpropanoyl-CoA	-162.901
163, 165	Short-chain trans-2,3-dehydroacyl-CoA derivates	-79.399

The short-chain fatty acid acyls show a similar behavior, as shown in Table 12. The first five entries differ by only  $18.879 \frac{kcal}{mol}$ . The top-ranked (R)-2-Hydroxy-4-methylpentanoyl-CoA (overall rank 30) is shown in Figure 48.



(R)-2-Hydroxy-4-methylpentanoyl-CoA

Figure 48: Structure of (R)-2-Hydroxy-4-methylpentanoyl-CoA. This compound is the highest-ranked compound by virtual screening of the short-chain fatty acyls.

When comparing 2-Methylpropanoyl-CoA with the energy of  $-162.901 \frac{kcal}{mol}$  at rank 116 and the overall top 1 ranked AcylCoA with  $-224.355 \frac{kcal}{mol}$ , a slightly larger difference of  $61.454 \frac{kcal}{mol}$  can be observed than with the other molecules (long-chain and mid-range fatty acids). However, generally no major outliers can be observed.

Within the short-chain fatty acid group is only a difference of  $18.879 \frac{\text{kcal}}{\text{mol}}$  between entries 1-5 (rank 30 and 85). Only the last entry with  $-79.399 \frac{\text{kcal}}{\text{mol}}$  is more distinct. This is also valid for mid-range fatty acids.

Overall, the results of the virtual screening show clear trends and the dependence on the structural composition of individual ligands. It was found that in the overall ranking from 1-50, fatty acids and their derivatives dominate, and only in the positions from 60 upwards do essentially other ligands show a higher presence.

Due to the fact that the ligand 3S-hydroxydecanoyl-CoA could already be crystallized in the structure of MFE-1 in the first active site[122] and that this ligand was also determined to have a good binding energy for docking into the second active site ( $-189.707 \frac{\text{kcal}}{\text{mol}}$ ) during virtual screening, the next part of the results focuses on docking this ligand into the second active site to gain a reliable structure.

## 5.4 Ligand conformers

The (S)-3-Hydroxydecanoyl-CoA is used to find the best pose for a reliable structure. To accomplish this, each ligand atom is matched with the active site II grid (gained from the homology model PDB 1FOY) with the lowest energy of the conformer. In total, 50 conformers of the (S)-3-Hydroxydecanoyl-CoA were computed. The following table lists the conformer models with the lowest total energy.

*Table 13: 3S-hydroxydecanoyl-CoA conformers with the lowest total energy for the second active site. The five compounds are ordered upon their total energy.*

Model	Total energy in $\frac{\text{kcal}}{\text{mol}}$
45	-31336.076
44	-31334.495
<b>50</b>	<b>-31331.570</b>
49	-31330.661
37	-31329.337

Table 13 shows that all five models have similar total energies. The distance between model 49 and 37 is  $-6.739 \frac{\text{kcal}}{\text{mol}}$ , which seems small compared to the total energy of the respective conformer.

Highlighted in bold is the conformer with the best binding pose in the second active site. Although, this model has not the lowest total energy, this result and the computed small discrepancies between the respective conformers allow conclusions about additional dependencies of the decisive poses. To understand these additional dependencies, the various conformers were docked and visualized. The following figure shows the different conformers in the second active site (Figure 49).

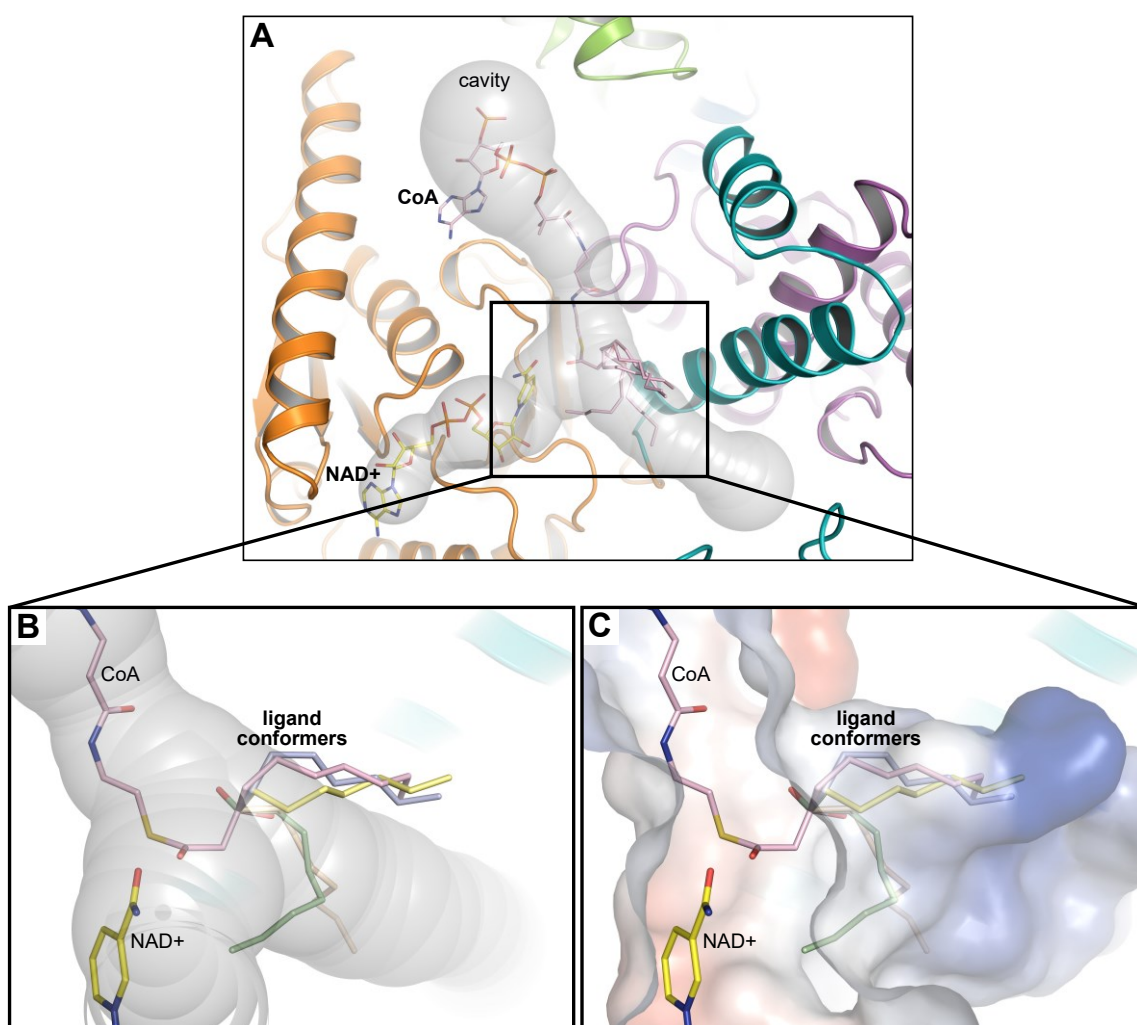


Figure 49: Overview of the five different conformers of the (S)-3-Hydroxydecanoyl-CoA compound. **A** Overview of the second active site with domains C (orange), D (cyan), and E (magenta) in cartoon representation. CoA and NAD+ are presented as stick models within the cavities (grey spheres) calculated with CAVER 3.0[431, 432]. **B** Zoom into the second

*active site II. Shown are the different binding poses within the calculated cavity (grey spheres) of the five conformers as stick models (model 37 is light-orange, model 44 is light-yellow, model 45 is light-blue, model 49 is light-green, and model 50 is light-pink). NAD<sup>+</sup> and CoA are also shown as stick models. C Different binding poses of the five conformers are shown as stick models within the second active site. The cavity is now represented in electrostatic potential solid surface calculated with APBS with PyMOL[448]. The conformers and the cofactors NAD<sup>+</sup> and CoA are shown as stick models.*

Figure 49 (**Panel B**) shows conformational differences of the respective conformers, although as described above, the total energy difference is only  $6.739 \frac{\text{kcal}}{\text{mol}}$ . It is noticeable that model 37 (orange) and model 49 (light-green) lie in the calculated cavity, with the latter one showing a bend. The other three conformers (model 44 in light-yellow, 45 in light-blue, and 50 in light-pink) lie slightly displaced outside the cavity, pointing to another direction. The reason for this deviation seems to be shown in Figure 49 (**Panel C**). The image of the electrostatic surface shows a kind of pocket next to the cavity. In this pocket, the models 44, 45, and 50 seem to be embedded very well. As the docking calculation has chosen model 50 to be the best conformer, it can be assumed that on the one hand, the total potential energy plays an important role in the choice of the binding pose, but the environment and especially electrostatic pockets should not be neglected and considered as crucial as well.

## 5.5 Discussion

### Virtual Screening

The HAD part of MFE-1 and the associated second active site is responsible for the dehydrogenation reaction in the beta-oxidation pathway of fatty acid metabolism. The studies of MFE-1 conducted in this chapter focused specifically on establishing a reliable structure of MFE-1 with 3S-hydroxydecanoyl-CoA substrate bound in the second active site. Due to the lack of existing crystal structures of this complex, potential ligands were first searched for based on the Virtual-Screening method, and a previously known ligand was attempted to dock accordingly. Due to virtual screening, a large number of potential ligands could be identified and ranked according to their importance and specificity. In

principle, substrate specificity for 3Z/3E-enoyl-CoA molecules has already been described by Wierenga *et al.*[121] and Kasaragod *et al.*[233]. Thus, in this dissertation, the set of potential ligands could be limited to 176 and finally classified into seven main categories. Of which, as well as consistent with physiological substrate specificity, no fatty acids are found at the top-ranked positions. Molecules without specific structural motifs are found right at the front, and the more likely potential ligands for the MFE-1 are represented with a total of about 24% of all molecules. This percentage seems low, but when considering the ranked positions of these fatty acid classes, it is again a good result. Short-chain, mid-range, and long-chain fatty acids represent the majority.

The MFE-1 ligand complex crystallized so far with 3S-hydroxydecanoyl-CoA in the first active site belongs to the representatives of the mid-chain fatty acids and is thus represented within the 24%. It is nevertheless striking that rather small molecules are ranked at the top. This could explain why comparably larger fatty acids and molecules theoretically fit suboptimal into the second active site. Furthermore, these molecules have the most negative binding energy, which can be seen as a consequence of their molecular size[449].

Besides the molecular size, which can be a major steric factor in specific binding pockets, an additional observation was also described in the results obtained here. The differences in the respective binding energies are at such a low level and therefore are considered negligible, although they are important for the ranking of the respective ligands. Thus, although the binding energy gradations represent a trend observed through all classes of molecules, they are basically only significantly important for ranking. Therefore, as also identified in this chapter, the virtual screening is important in conjunction with the ranking of potential ligands. However, the physiological aspect related to substrate specificity ends up forming the basis of possible future crystallizations or experiments with different ligands of MFE-1.

### **Conformers**

The calculation of a reliable structure with 3S-hydroxydecanoyl-CoA in the second active site was done by using the homology modeling method. In the absence of an existing crystal structure, five potential candidates were identified. The marginal differences in the total energies of the corresponding MFE-1 constructs are striking. Since the 3S-hydroxydecanoyl-CoA has different conformations, changes due to rotatable bonds or

spatial orientations can be expected to have only minor effects on the total energy of the conformer or on the overall MFE-1 system. This behavior, therefore, seems to be reflected by the small energy differences (Table 13).

However, the decision which conformer is the most likely variant in the second active site seems to depend on the chemical environment, such as interaction between the binding pocket and the substrate, when considering the position of the fatty acid chain. This assumption could be supported by the fact that finally, the model was chosen, which seems to be best embedded in a side pocket revealed by cavity calculations but does not have the largest (negative) total energy of all five potential candidates. Further supporting data for calculated model is provided by the recently published structure of Wierenga *et al.*[450].

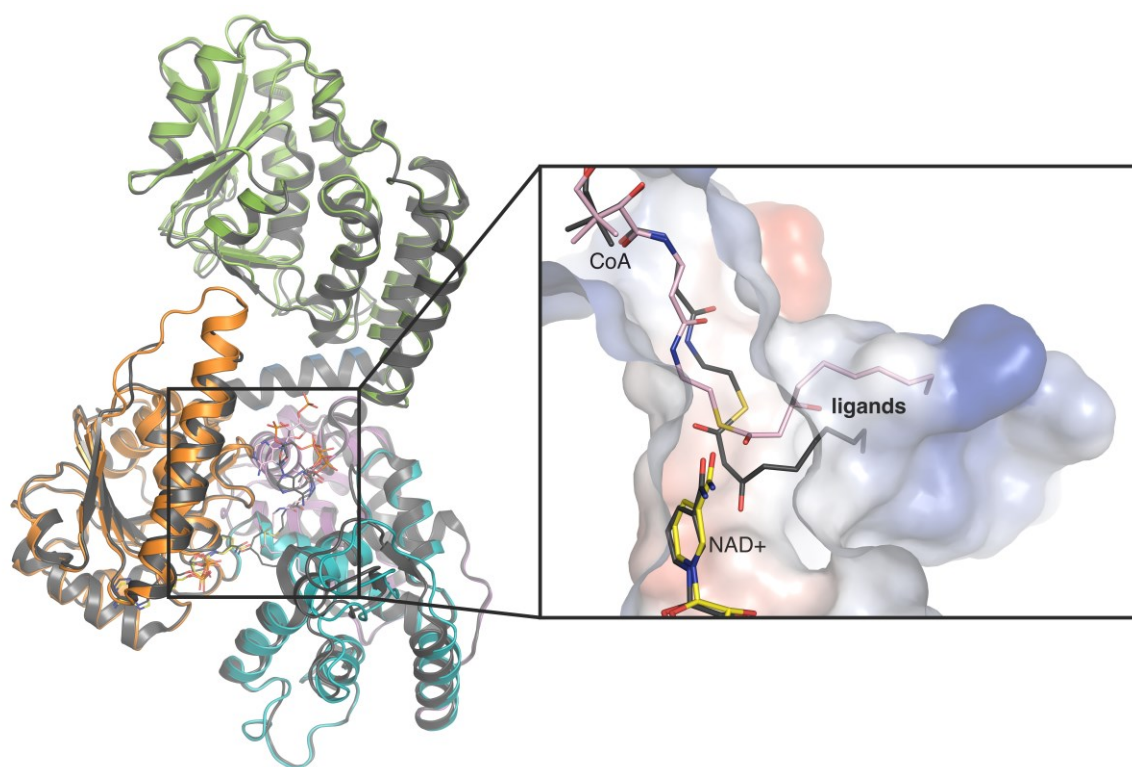


Figure 50: Overview of MFE-1 with the (S)-3-Hydroxydecanoyl-CoA compound in light-pink superimposed with the recently published MFE-1 complex (black cartoon representation) with bound 3-Ketodecanoyl-CoA compound represented in black as stick model (PDB: 6Z5F)[450].

The research group succeeded in crystallizing the MFE-1 with 3-ketodecanoyl-CoA embedded in the second active site. By superimposing this structure with the 3S-hydroxydecanoyl-CoA structure obtained by docking in the scope of this thesis (Figure 50), a very good agreement is observable. Minor differences are present as deviations in the orientation and conformation of the CoA compartment and the position of the terminal fatty acid chain. This could be due to the fact that the conformer used in this dissertation is the reactant for the second active site, while the Wierenga group[450] has crystallized the product of the dehydrogenation reaction. Since a reaction educt might have a higher affinity compared to the product small deviations might be reasonable. In addition to that, the result of the docking process describes the reactant (3S-hydroxydecanoyl-CoA), that has to be dehydrated in the second active site whereas Wierenga's structure of MFE-1 has the dehydrated product embedded. Therefore, conformational differences are plausible. Further investigations are needed to clarify these issues and to study both molecular forms in detail in MD simulations. However, the result of the docked 3S-hydroxydecanoyl-CoA conformer shows that the embedding in this form is in very good agreement with the recently available comparative model of the Wierenga group[450].

For the first time, a structure with a physiologically relevant reactant has been modeled that serves as a basis for future structural studies and possible mechanistic analyses at the second active site of MFE-1.

## 5.6 Summary

The MFE-1, because of its two active sites, essentially catalyzes three reactions, the third reaction being a dehydrogenation and occurring in the second active site. Due to the lack of crystal structures that have an embedded substrate in the second active site, it was necessary to obtain a reliable 3D structure gained by general structural modeling to embed a substrate. In order to identify potential ligands for the MFE-1 and to embed the physiologically relevant substrate 3S-hydroxydecanoyl-CoA, which has so far only been used for the first active site, virtual screening and docking were used. From a large number of candidates, it could be shown that substrate specificity plays a decisive role in the selection of ligand candidates. Hence, the selection of potential substrates can be reduced by the specificity of MFE-1 towards the ligands, although their percentage is



small compared to all investigated substrates. The calculation of the best pose for the 3S-hydroxydecanoyl-CoA conformer was also successfully performed. Although the best conformer does not have the lowest total energy, the choice of the best pose showed that the decision for a potential model structure is based on the interplay of conformation, chemical environment as well as interactions within the binding pocket. For future modeling of ligand candidates, these results provide a reliable basis and, for the first time, a basic model was gained that can be used for future structure-function investigations.



## 6 Results – Domain motions in MFE-1

### 6.1 Introduction

The knowledge of motions on an atomic level and their effects in biological processes is crucial to understand biological functions[313]. Molecular dynamics (MD) is used to obtain this kind of information. Over the past decades several MD simulations of proteins have been analyzed (*e.g.* McCammon & Harvey, 1987[314]; Brooks *et al.*, 1988[451]; Hünenberger *et al.*, 1995[452] etc.). These analyses are important for identifying groups of atoms moving concerted or possible influences of solvent on the dynamics[314, 315]. MD simulations can also be used for the elucidation of differences in the molecular dynamics between similar structures such as different coformers[315]. The analysis of the dynamics is also used to identify sidechain motions which can be a part of enzymatic reactions or have a crucial role for allosteric regulation[315].

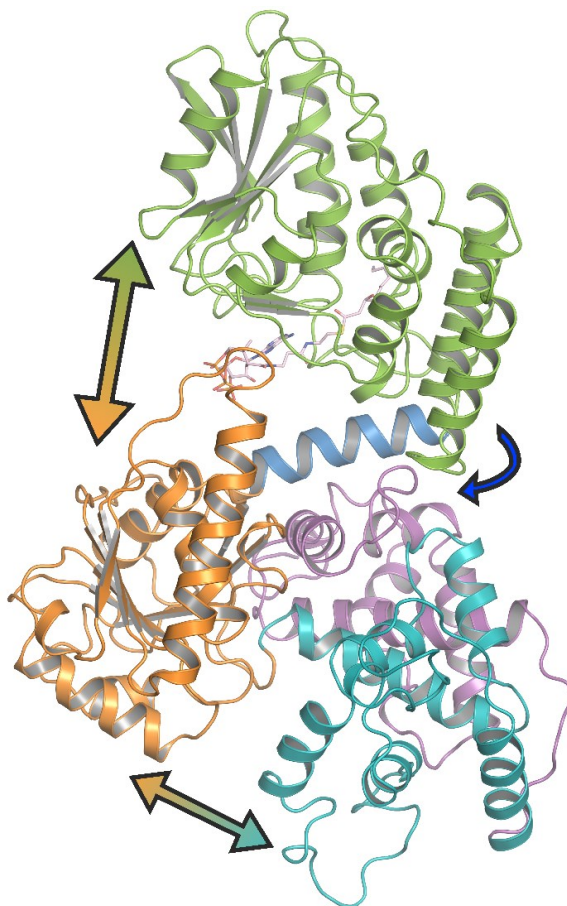


Figure 51: Overall fold of the MFE-1 in cartoon representation. Domain A is green, domain B is blue, domain C is orange and domains D and E are cyan and magenta respectively. The blue arrow on the right represents the flexible anchor connection

between the domains A and B. The two-colored arrows represent the suggested motions between domain A (green) and C (orange) and between domain C and D (cyan) necessary for the functionality of the MFE-1.

In the scope of this dissertation, it is assumed that the substrate channeling of the MFE-1 is correlated to certain domain motions (Figure 51)[233]. However, it is unknown whether there are any correlations between specific domains or side chains within MFE1 facilitating the substrate channeling. For this reason, it is necessary to investigate whether there are any correlated movements at all. To address this question modelling and analyzing techniques such as normal mode analysis, accelerated molecular dynamic simulations and dynamic cross-correlation were used and the results are described in this chapter.

## 6.2 Methods and model setup

### Normal mode analysis

Normal mode analysis was performed with the WEBnm@ v2.0 interface of the Reuter research group[312, 453]. For this, the PDB 3ZW8 was uploaded and the job was submitted. This interface uses the Elastic Network Model (ENM). The applied force field is the C $\alpha$  force field by Hinsen *et al.*[287] with each amino acid represented by a mass at the C $\alpha$  position. It is available within the Molecular Modelling ToolKit[453]. The pair potential between the C $\alpha$  is given by equation (60)[453],

$$U_{ij}(R) = k(\|R_{ij}^0\|) \cdot (\|R_{ij}\| - \|R_{ij}^0\|)^2 \quad (60)$$

with:

$$k(r) = \begin{cases} ar - b, & r < d, \\ cr^{-6}, & r \geq d. \end{cases} \quad (61)$$

The parameters in equation (60) and (61) were determined by Hinsen *et al.* They are fitted to an all-atom model[287] as follows:  $a = 8.6 \times 10^5 \text{ kJmol}^{-1}\text{nm}^{-3}$ ;  $b = 2.39 \times 10^5 \text{ kJmol}^{-1}\text{nm}^2$ ;  $c = 128 \text{ kJmol}^{-1}\text{nm}^4$ ;  $d = 0.4 \text{ nm}$ .

$R_{ij}$  represents the distance vector between two  $C\alpha$ .  $R_{ij}^0$  represents the corresponding distance vector in the input of the WEBnm@ interface configuration.  $D$  represents the distance between two neighboring  $C\alpha$  atoms. Accordingly, the total potential energy of the configuration  $R$  of the Elastic Network Model is given by:

$$U(R) = \sum_{\text{all pairs } i,j} U_{ij}(R) \quad (62)$$

The eigenvectors of the mass weighted matrix of the second partial derivatives of  $U(R)$  represent the normal modes. The six lowest frequency modes are provided by the web user interface through an applet. The modes are represented via vector field arrows showing the direction of each mode.

### Molecular dynamics simulations

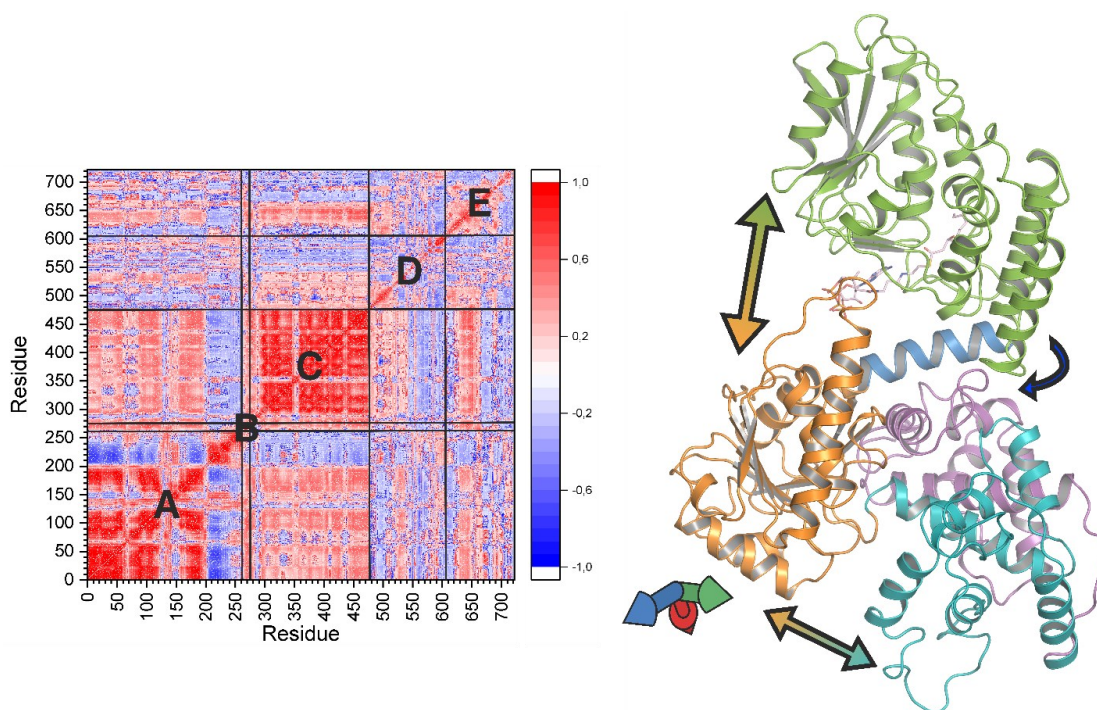
The initial model structure of MFE-1 was constructed from the Protein Data Bank entry 3ZW8[424]. The coordinates of the fatty acid complex 3S-hydroxydecanoly-CoA were gained out of the PDB 3ZWC and implemented after superimposing into the apoprotein structure (PDB: 3ZW8). The cofactor NAD<sup>+</sup> was extracted out of the PDB entry 3ZW9 and also superimposed and inserted in the apoprotein structure.

Therefore, all three MFE-1 systems were set up and prepared analogously as described in chapter 4.2.

### Dynamic cross correlation

The dynamic cross correlation matrix was calculated for each MD run. After the alignment to the domains D and E, the correlation matrices were calculated for the whole MD trajectories with Maple software[454]. Each matrix element quantifies the correlation between two coordinates of the atoms  $i$  and  $j$  (see equation (27)).

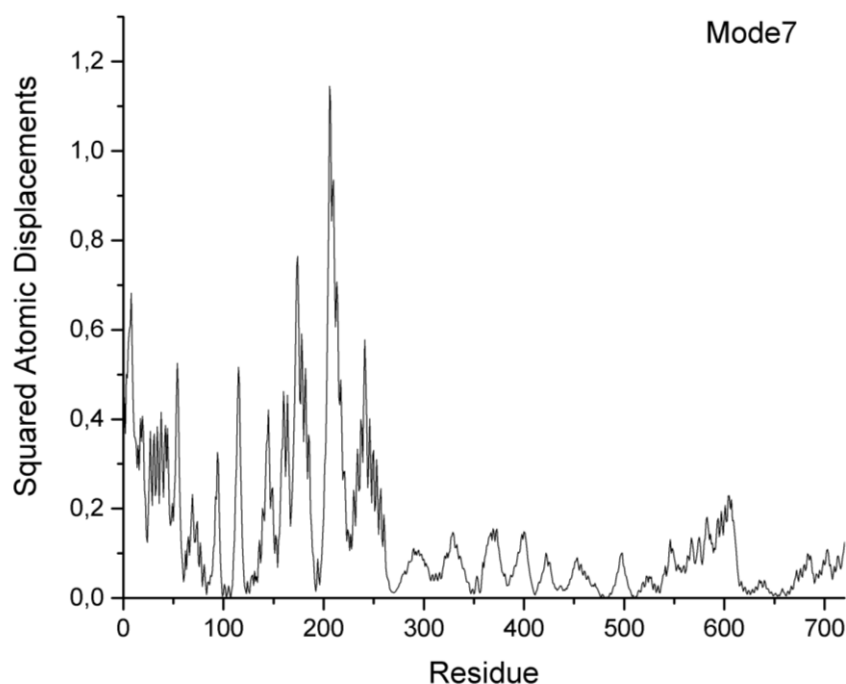
The resulting matrices show the correlations of the atoms ranging from -1 to 1. With -1 and 1 are maximal anti-correlations and correlations (Figure 52), respectively.



*Figure 52: Overall fold of the MFE-1 in cartoon representation with the associated dynamic cross correlation plot. Red areas represent positive correlation between domains, blue areas represent anti-correlation between domains. Domains A – E are highlighted in bold capital letters and illustrate the associated domain regions.*

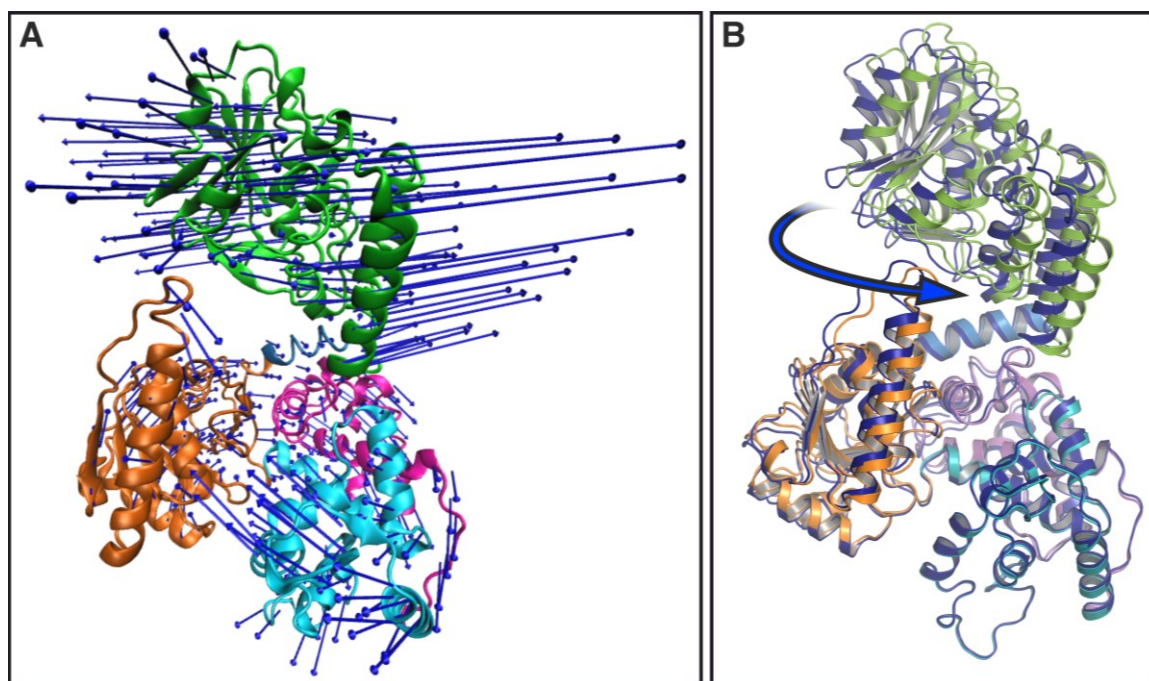
### 6.3 Normal mode analysis

To understand whether there are intrinsic motions within the MFE-1, first, the normal mode analysis was done. For this purpose, the apoprotein (PDB: 3ZW8) was used. Due to the fact that the calculated first six modes refer to global rotation and translation[455] they are not taken into account further in the analyses. Therefore, the first normal mode investigated here is mode 7. Because of this, the modes 7 to 12 are calculated and visualized. Figure 53 shows the displacement plot obtained by calculating motion 7, which is visualized with the arrow vector field in Figure 54.



*Figure 53: Calculated mode 7 of the MFE-1 represented by the atomic displacement plot in Å. The associated displacements concern mainly the residues 175 to 250 from domain A. Compared to this region, the remaining domains are relatively rigid.*

The displacement plot (Figure 53) shows that the domain A and residues 1 to 261 contributes the most to the protein motion of mode 7 in comparison to the rest of the enzyme. Especially the residues 175 to 250 (red box) have the highest displacement up to 1.2 Å. Whereas the rest of the protein shows high rigidity with an atomic displacement of 0.4 Å or less. Domain B is the most rigid one with atomic displacement below 0.02 Å.

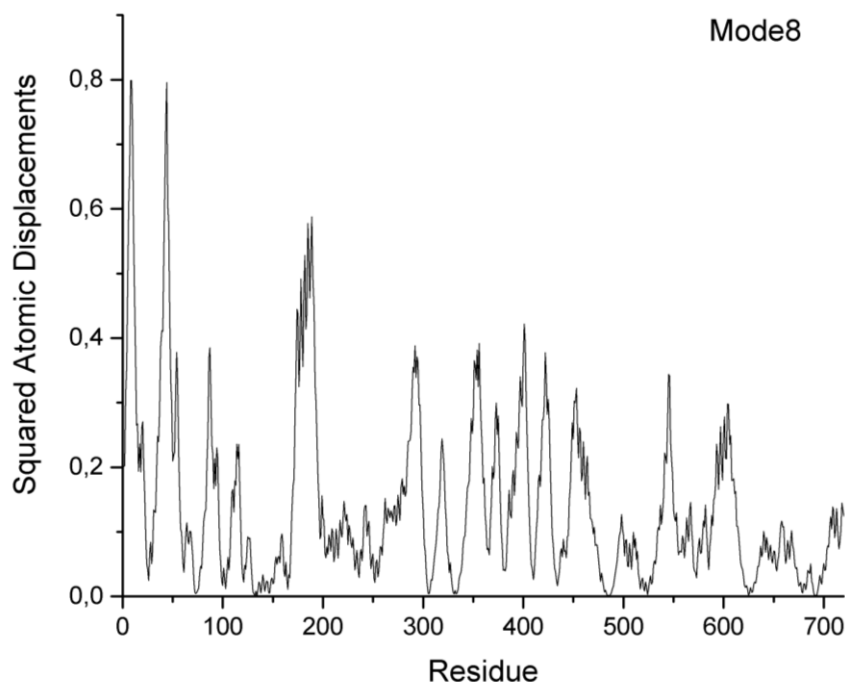


*Figure 54 Large scale collective motions from normal mode analysis of mode 7. **A** Vector field view of the normal mode 7 of the MFE-1, corresponding to twist, bend, wobble and stretching modes. **B** Overall view of the MFE-1 in cartoon representation with two different positions. Blue represents the end position of the mode 7 with an alternating motion as indicated by the blue arrow direction. The arrow indicates the main dominating motion of the corresponding domains during mode 7.*

Figure 54A shows the movement of domain A which is a turning twist with a back-and-forth motion connected to the anchor domain B. The arrow vectors of domain A are much longer than the others of the domains B to E which represents higher magnitudes of motion compared to the rest of the protein. This reveals that especially domain A is very actively moving during this mode. Barely visible is that domain A and domain C perform a movement like a dumbbell field indicated by the arrow field. Panel B (Figure 54) illustrates the dominating motion during mode 7 of the MFE-1 highlighted by the blue arrow identified during the normal mode calculation.

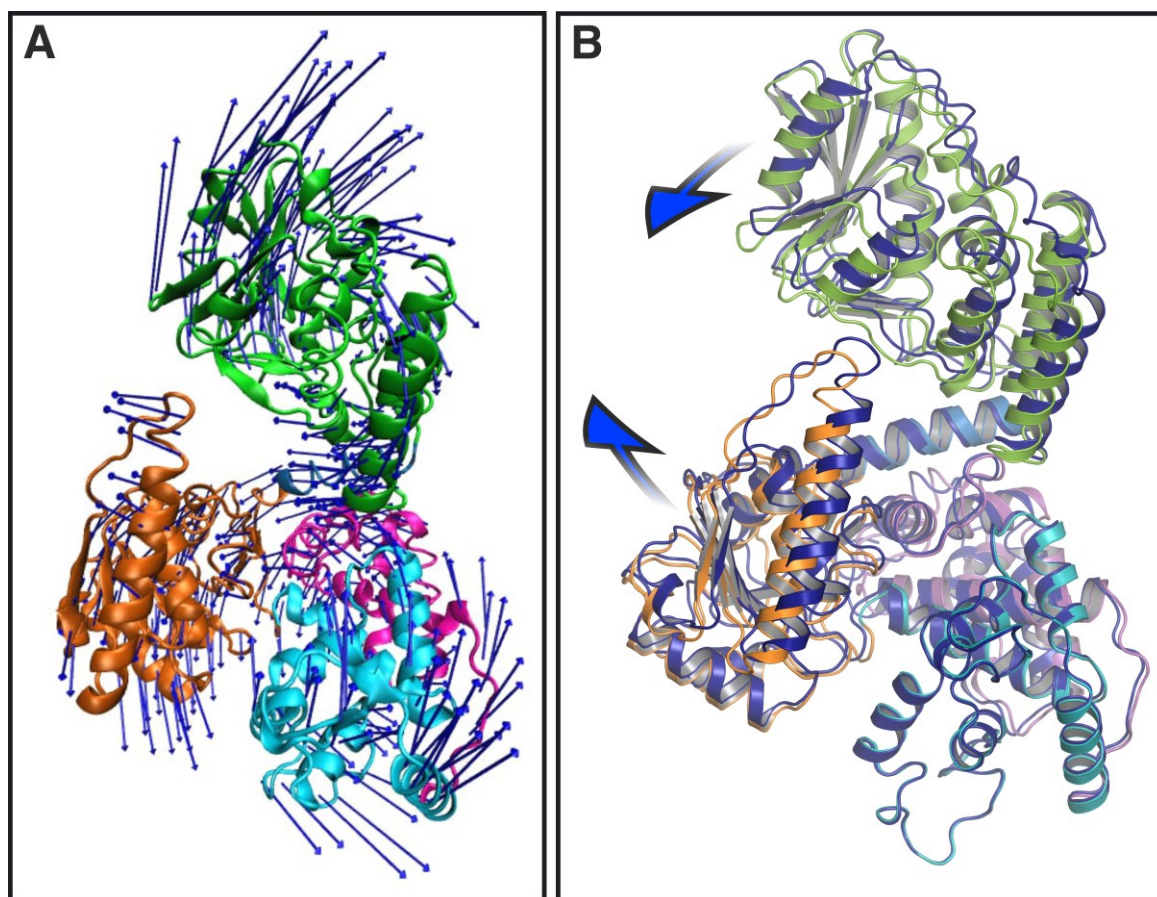
Mode 8 shows very similar results. In Figure 55 the calculated displacement and their corresponding residues are displayed.





*Figure 55: Calculated mode 8 of the MFE-1 represented by the atomic displacement plot in Å. The associated displacements concern mainly the residues 175 to 200 (0.6 – 0.8 Å). All other domains B to E of MFE-1 are rigid with a displacement maximum of 0.4 Å compared to domain A.*

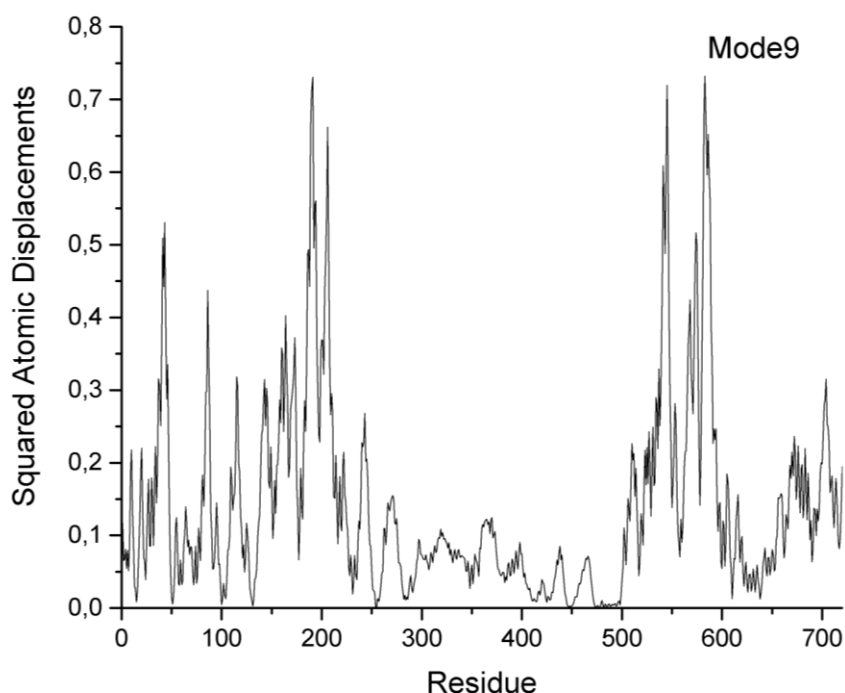
Mode 8 Figure 55 shows a very similar spectrum compared to the spectrum of mode 7. Domain A still appears to be very flexible, while the rest of the MFE-1 system is more rigid in position. The previously very rigid domain B in mode 7 is again hardly participating in atomic displacements and thus very rigid. The remaining domains of the MFE-1 show almost identical behavior and seem to be more rigid than flexible in mode 8.



*Figure 56: Large scale collective motions from normal mode analysis of mode 8. **A** Vector field view of the normal mode 8 of the MFE-1, corresponding to twist, bend, wobble and stretching modes. **B** Overall view of the MFE-1 in cartoon representation with two different positions. Blue represents the end position of the mode 8 with an alternating motion as indicated by the blue arrow direction. The blue arrows indicate the main dominating motions of the corresponding domains A and C during mode 8.*

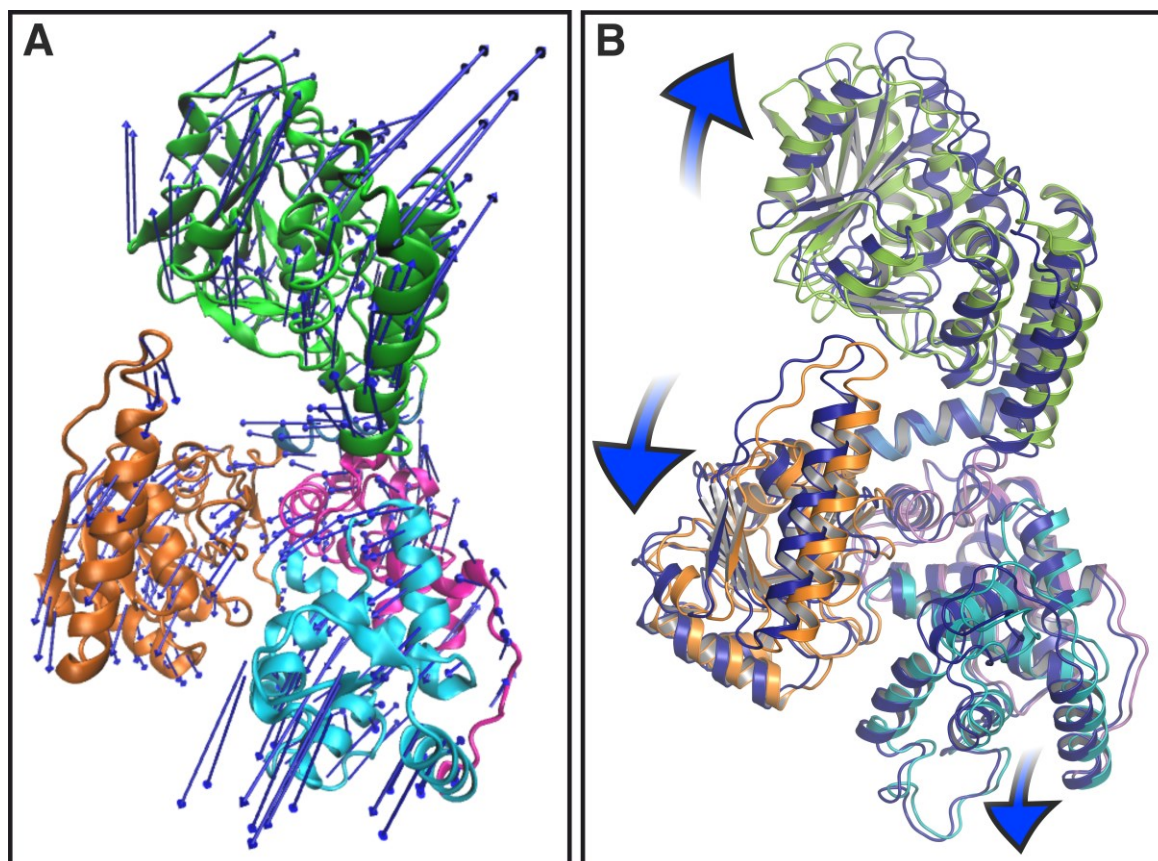
Figure 56A shows the direction of the domain A and also domain C and their residues visualized by the arrow vectors. The movement of both domains A and C are anti-correlated bending. Compared to the rest of the enzyme the magnitudes of the motion correlated to domain A and C are high as indicated by the length of the arrow vectors. This also indicates that both domains have the main contribution to the flexibility in mode 8. Only barely visible in mode 7 the dumbbell (Figure 54), is now well exhibited. These dynamic motions of especially domains A and C which harbor the active site I and II show that they are more flexible in comparison to the rest of the MFE-1. Panel B (Figure 56) illustrates the main identified motions during mode 8 of the MFE-1 highlighted by the blue arrows.

Motion 9 and the calculated displacement plot (Figure 57) and the corresponding vector field (Figure 58) shows a big difference compared to mode 7 and 8.



*Figure 57: Calculated mode 9 of the MFE-1 represented by the atomic displacement plot in Å. The associated displacements of motion 9 are dominated by the residues of domain A (0.5 – 0.7 Å) and domain D (residues 500 to 600; displacement of 0.7 Å). The domains B, C and E of MFE-1 are rigid with a displacement of 0.0 to 0.3 Å compared to domain A and D.*

Interestingly mode 9 shows the first significant change in the displacement spectrum. Basically, domain A again seems to be very flexible and domain B is again very rigid, like the rest of the entire MFE-1, except for domain C, which shows significantly higher displacements in contrast to the previous ones. At the same time, with the level of displacement up to 0.8 Å, it can be assumed that both domains show very similar flexibilities during mode 9. Overall, this displacement spectrum shows a trend continuation with domain A having the most flexible part in MFE-1, while the remaining sections always appear relatively rigid, with domain C now seeming to increase in importance in flexibility, as the calculated large-scale motions describe very well (Figure 58).

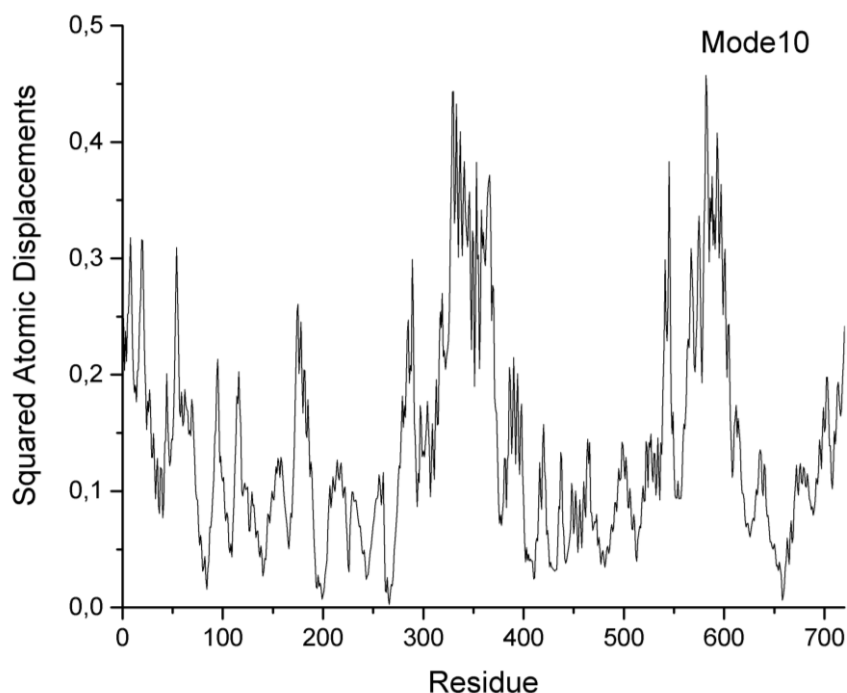


*Figure 58: Large scale collective motions from normal mode analysis of mode 9. **A** Vector field view of the normal mode 9 of the MFE-1, corresponding to twist, bend, wobble and stretching modes. **B** Overall view of the MFE-1 in cartoon representation with two different positions. Blue represents the end position of the mode 9 with an alternating motion as indicated by the blue arrow directions. The arrows indicate the main dominating motions of domains A and C during mode 9.*

The large-scale motions show the anti-correlated motion between domains A and C (illustrated by the blue arrows in Figure 58B). These motions also elucidated that domain D moves in the same direction as domain C. This result leads to the assumption that domain D is also anti-correlated to domain A in that case. The calculated vector field (Figure 58A) shows the main motion of these two domains. The magnitudes for each domain motion are similar when the vector length is compared of each domain. Likewise motion 8, the well-formed dumbbell is clearly visible in the vector field. The exhibited anti-correlated motion between the domains A and C agrees also to the suggested motion to be necessary for the substrate channeling as proposed by Wierenga and

Kasaragod[121, 122, 233]. Panel B (Figure 58) illustrates the main identified motions during mode 9 of the MFE-1 highlighted by the blue arrows.

Motion 10, the calculated displacement plot (Figure 59) and the corresponding vector field (Figure 60) shows the most significantly change in the spectra pattern compared to mode 7, 8 and 9.

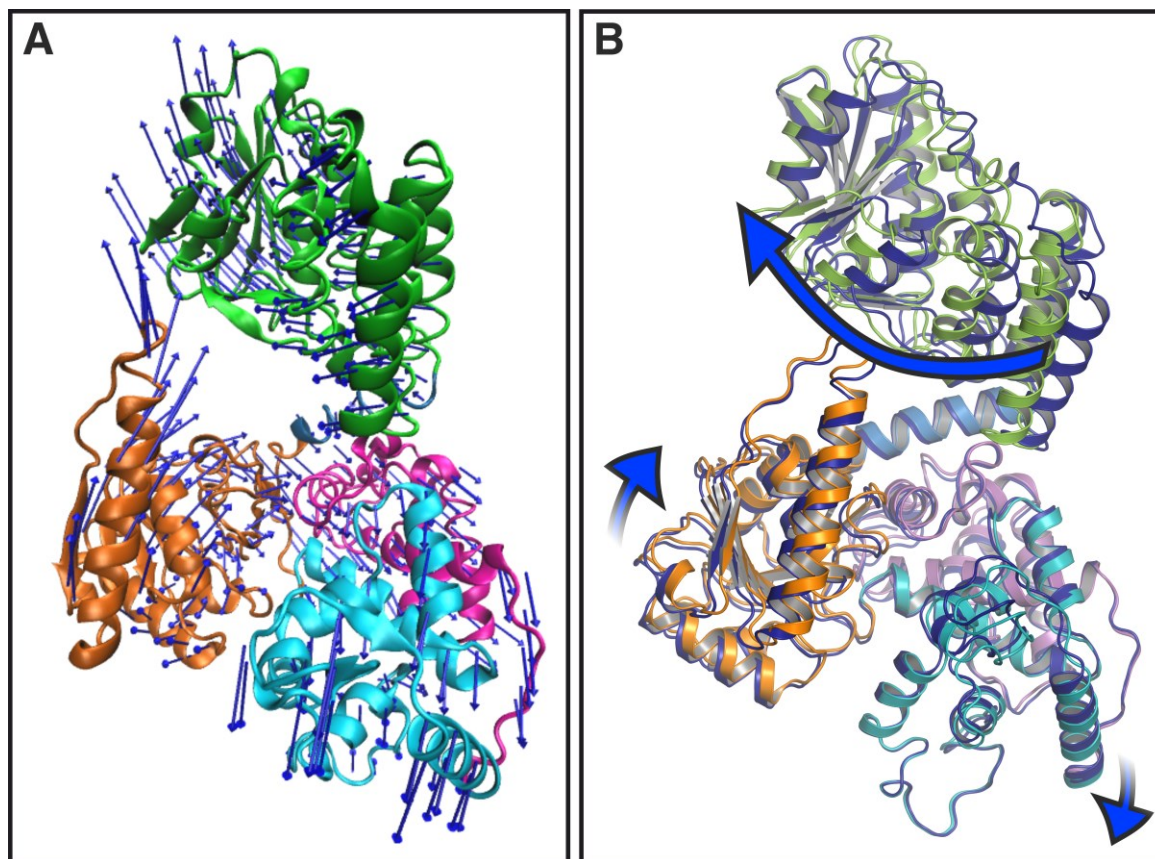


*Figure 59: Calculated mode 10 of the MFE-1 represented by the atomic displacement plot in Å. The associated displacements calculated and resulting from mode 10 are particularly dominated by domain C and domain D. Both domains have a maximum of the displacement at 0.46 Å while the remaining domains A and B as well as E experience a smaller displacement in comparison.*

For mode 10, a significant change in the spectrum occurs. While in the modes 7 – 9 mainly the domains A and C describe the most significant motion, domain D is especially active in mode 10. In this mode, domain A shows much less activity with a maximum motion of 0.3 Å compared to before when domain A had much higher displacements up to 0.8 Å. Domains C and D each stand out clearly with a maximum displacement of 0.46 Å. It is noticeable that domains B and E with displacements from 0.02 – 0.1 Å are rather rigid. So, the latter domains again show a high rigidity as before. Also in this mode



10, the trend of high flexibility of domains A and C seems to be confirmed, although the intensity has decreased. As Wierenga and coworkers assumed[121, 122, 233], domain D seems to exhibit some flexibility due to interactions with the second active site.

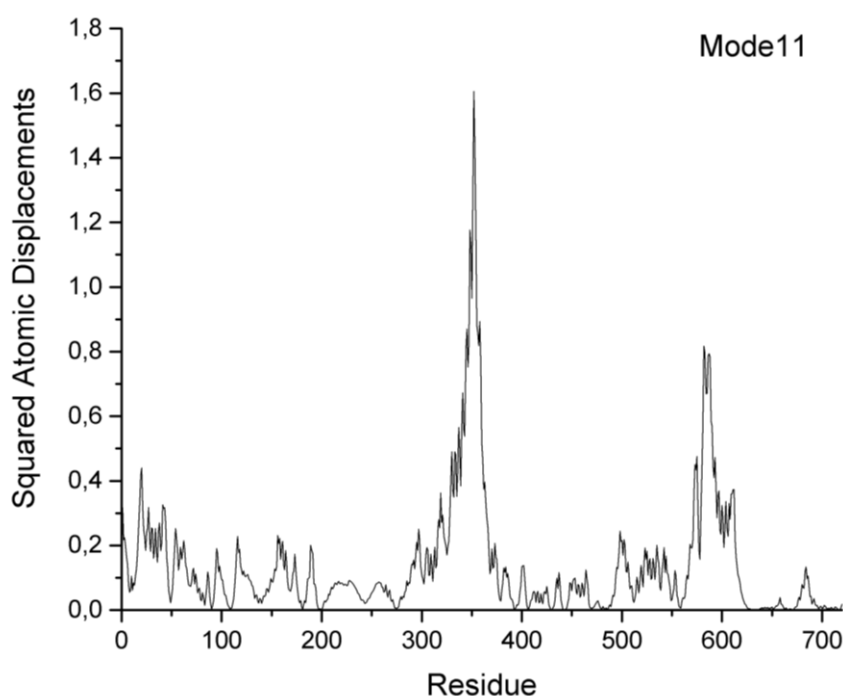


*Figure 60: Large-scale collective motions from normal mode analysis of mode 10. **A** Vector field view of the normal mode 10 of the MFE-1, corresponding to twist, bend, wobble and stretching modes. **B** Overall view of the MFE-1 in cartoon representation with two different positions. Blue represents the end position of the mode 10 with an alternating motion as indicated by the blue arrow direction. The arrows indicate the main dominating motions of the corresponding domains A, C and D during mode 10.*

The calculated large-scale motions (Figure 60) show the correlation between domains A, C and D and thus a corresponding connected motion. The vector field shows the main motion of these three domains. In contrast to the modes 7 – 9, domains A and C do not show a correlated motion. Therefore, they move in different directions independent from each other during mode 10. Domain C and D show an anti-correlated motion which is contrary to mode 9 where both are correlated. The appearance of an anti-correlated

motion between the domains C and D seems to agree to the suggestion from Wierenga *et al.*[121, 122, 233] that certain motions seem to be necessary for the substrate channeling and the functionality of the MFE-1. Panel B (Figure 60) illustrates the main identified motions during mode 10 of the MFE-1 highlighted by the blue arrows.

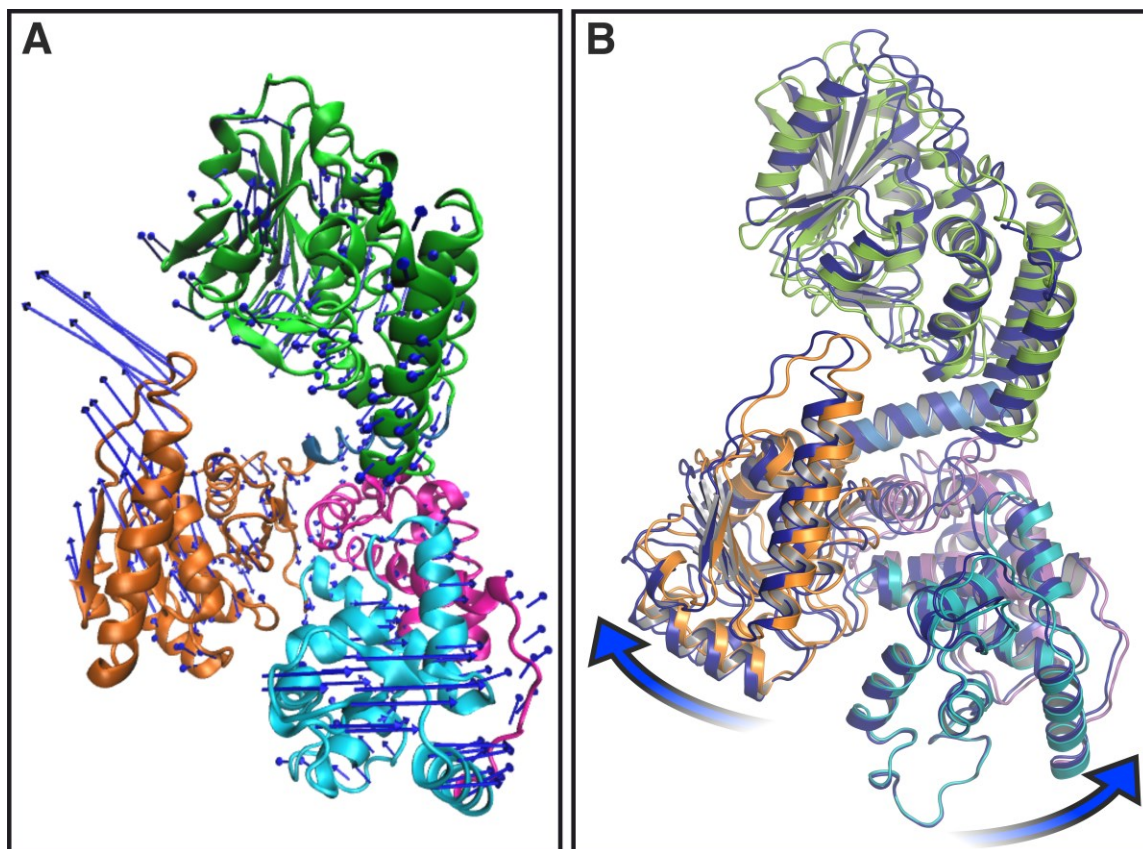
Modes 7 – 10 already clearly show the displacements of the MFE-1 essentially dependent on three domains, namely the domains A, C and D. Mode 11 provides the clearest displacement of the domains C and D in comparison to all previous modes (Figure 61).



*Figure 61: Calculated mode 11 of the MFE-1 represented by the atomic displacement plot in Å. The associated displacements resulting from mode 11 are dominated by domain C and domain D. Both domains have a maximum of the displacement at 1.6 Å (domain C) and 0.85 Å (domain D). The remaining domains A, B and E are negligible in comparison.*

In mode 11 two main domains (C and D) with the highest flexibility could be identified. Both domains have a significant peak. Remarkable is that the displacement of domain C with a maximum of 1.65 Å is by far the strongest displacement of all modes (7 – 12). Domain D with a maximum of 0.8 Å also shows a distinct peak. Both domains seem to be

significant in this mode. This result also reflects the previous trend that the essential movements of the MFE-1 are carried by mainly three domains A, C and D. Although domain A describe a smaller displacement of less 0.4 Å. The remaining domains B and E are negligible with a maximum of 0.2 Å.



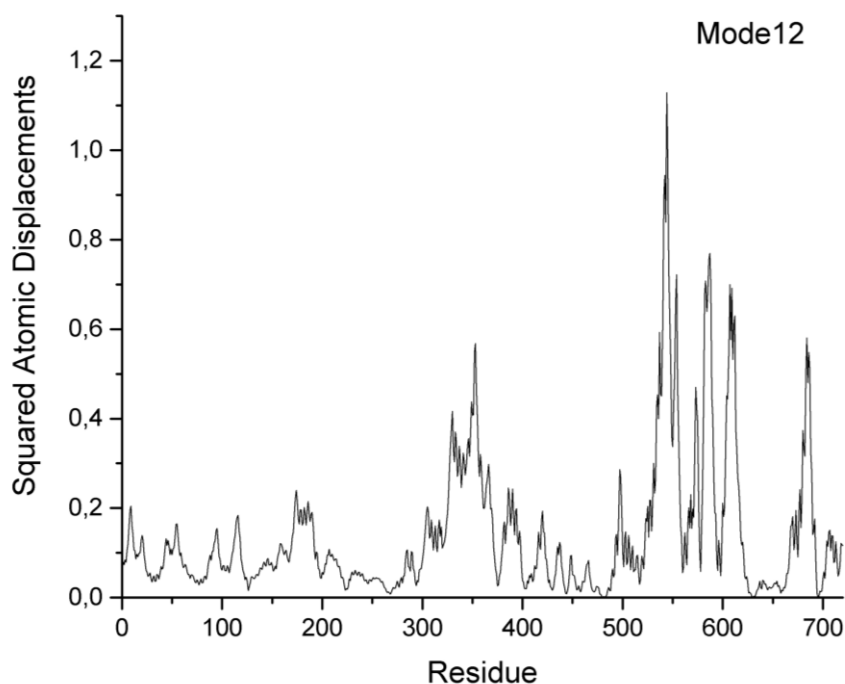
*Figure 62: Large-scale collective motions from normal mode analysis of mode 11. **A** Vector field view of the normal mode 11 of the MFE-1, corresponding to twist, bend, wobble and stretching modes. **B** Overall view of the MFE-1 in cartoon representation with two different positions. Blue represents the end position of the mode 11 with an alternating motion as indicated by the blue arrow direction. The arrows indicate the main dominating motions of the corresponding domains C and D during mode 11.*

The large-scale motions (Figure 62) shows the vector field of the dominating motions of mode 11. Mainly domains C and D are moving. This feature is also given by the higher magnitude of the vector length of the arrows of domains C and D. Both domains show clearly visible an anti-correlated motion. This interaction of both domains is also anti-correlated during mode 10 but in a different direction. Likewise in mode 10 the domain A



seems not to have a correlation (negative or positive) with both domains. The rest of the MFE-1 domains seem to be rigid represented by the low magnitude of the arrow vectors of the remaining domains B and E. This type of connection between specifically domains C and D demonstrates the discovered trend that certain patterns of motion and thereby correlation relationships of the respective domains seem to be important for the MFE-1. Figure 62B illustrates the main identified motions during mode 11 of the MFE-1 highlighted by the blue arrows.

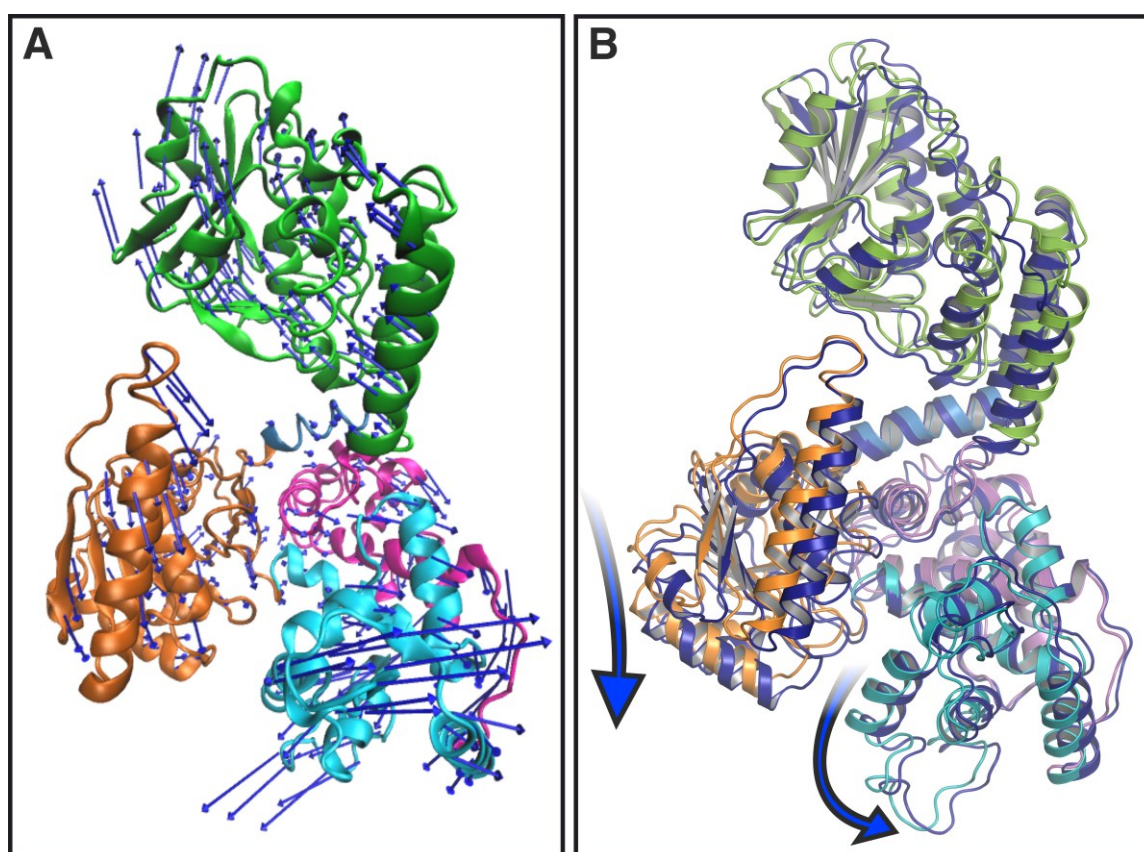
The last of the six lowest frequency modes is mode 12, which essentially addresses three of the five domains (Figure 63).



*Figure 63: Calculated mode 12 of the MFE-1 represented by the atomic displacement plot in Å. The associated displacements resulting from mode 12 are dominated by domain D, domain E and domain C. Domains C and E have a maximum of 0.6 Å each and domain D has a significantly higher maximum of 1.2 Å. The remaining two domains A and B play a minor role in this mode.*

The large-scale motions for the mode 12 also indicate that domain C and D are mainly contributing to the atomic displacements. Although domain E is not relevant in the other

modes, it plays an important role in this mode 12. Nevertheless, it seems to have a negligible influence on the actual dynamics of the MFE-1 overall. It remains striking that domains C and D as well as domain A seem to play a major role in domain flexibility in all modes, and that domain E is now at least also evident. Thus, only domain B always remains very rigid, which is also the case in this mode 12. Furthermore, this behavior shown by all modes could mean that domain B could be attributed to an anchor function, which was also described by Wierenga *et al.*[121].



*Figure 64: Mode 12 Large scale collective motions from normal mode analysis. **A** Vector field view of the normal mode 12 of the MFE-1, corresponding to twist, bend, wobble and stretching modes. **B** Overall view of the MFE-1 in cartoon representation with two different positions. Blue represents the end position of the mode 12 with an alternating motion as indicated by the blue arrow direction. The arrows indicate the main dominating motions of the corresponding domains A, C and D during mode 12.*

The large-scale motions and the corresponding vector field of mode 12 clearly shows that domain D contributes the most to the atomic displacements overall. This is also

represented by the vector field with domain D having the highest magnitude of the arrow vectors. Domain A is similar whereas the rest of the MFE-1 (domains B, C and E) seems to be very rigid during that mode. In agreement to the modes 10 and 11, here the domains C and D are also correlated. In addition to that, domain A is anti-correlated to domain C considering the vector field of the whole domains. Mode 12 also shows a certain interaction between the four domains A, C, D, and E of the MFE-1, confirming the trend that the suggested motion of the whole enzyme seems to have a key-role for the functionality itself[121, 122, 233]. Figure 64 illustrates the main identified motions during mode 12 of the MFE-1 highlighted by the blue arrows.

## 6.4 Dynamic cross correlation

### 6.4.1 MFE-1 apoprotein

To further support the assumption of a connection between substrate channeling and domain motion by Wierenga *et al.*[121, 122, 233] and based on the previous results from normal mode analysis, the MD trajectories of each MFE-1 systems were further analyzed using dynamic cross correlation (DCC) calculations.

First, the apoprotein shows the following DCC matrix (Figure 65).

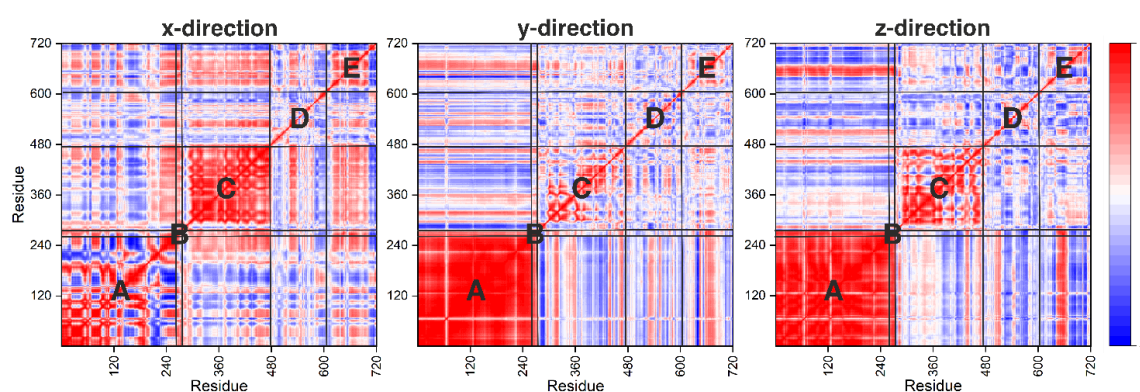
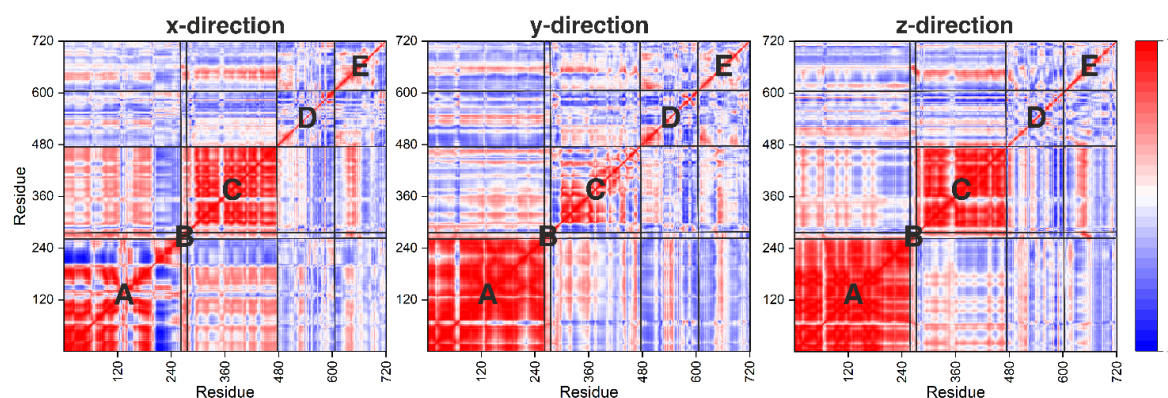


Figure 65: DCC matrix describing the correlation between domains. Red areas represent positive correlation between domains, blue areas represent anti-correlation between domains. MFE1 was aligned to domain D and E. X-, y- and z-directions refer to cartesian coordinates indicated by axis in Figure 52 in chapter 6.2. DCC maps of the apoprotein are shown with the protonated glutamate 103 in x-, y- and z-direction. The corresponding domains of the MFE-1 are illustrated in capital letters A, B, C, D and E.

The analysis of all three DCC maps in x-, y- and z-direction clearly show that especially domain A and domain C seem to be internally correlated domains represented by the respective red block in x-, y- and z-direction. In x-direction the behavior seems to be different, since domain A indicates a mixed correlation on intramolecular level (a higher anti-correlated fraction compared to y- and z-direction). Apparently, it seems to move less coordinated in this direction. While domain C in x-direction shows more distinct the internal correlation as a block. Comparing all directions, the connection between domains A and C appears to be anticorrelated. Less pronounced in the x- and z-directions, the y-direction of domain B shows an anti-correlation to both domains A and C. Domains D and E do not appear to have any other noticeable correlations with the rest of the domains. Moreover, the correlations between domains A, B, and C support the results for the domain movements shown in section 6.3.

#### 6.4.2 MFE-1 complexed with 3S-hydroxydecanoyl-CoA

Quantification of the domain motion after embedding the 3S-hydroxydecanoyl-CoA revealed first differences considering the various spatial directions (Figure 66).



*Figure 66: DCC matrix describing the correlation between domains. Red areas represent positive correlation between domains, blue areas represent anti-correlation between domains. MFE1 was aligned to domain D and E and 3S-hydroxydecanoyl-CoA is present in domain A. X-, y- and z-directions refer to cartesian coordinates indicated by axis in Figure 52 chapter 6.2. DCC maps of the protein complex are shown with the protonated glutamate 103 in x-, y- and z-direction.*

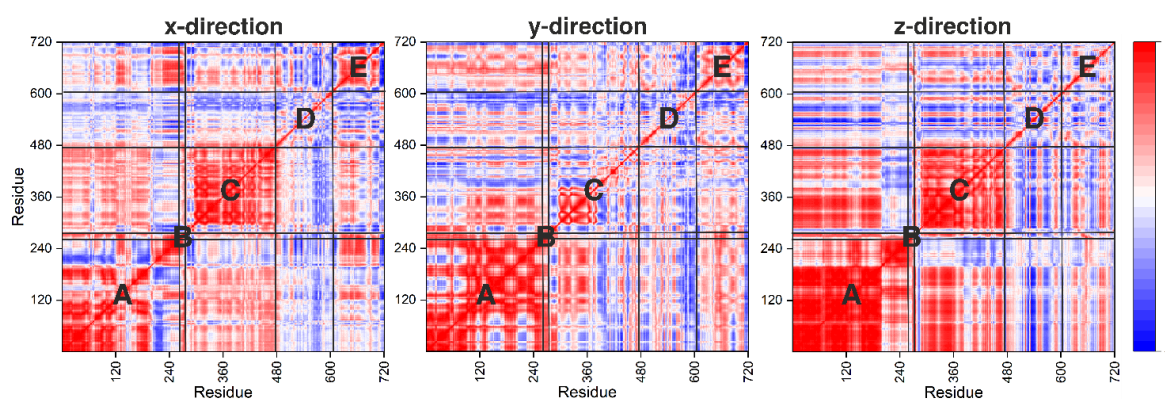
First, the analysis of the DCC maps in the spatial x-, y-, and z-directions basically revealed a similar behavior of all domains with respect to the apoprotein. Again, a clear intramolecular interaction within each of the domains A and C could be shown for this MFE-1 construct, as observed for the apoprotein. The presence of a substrate during movements in x- and z-direction seems to enhance this behavior for domains A and C in contrast to the apoprotein, where the movement is less significant. Consequently, the embedding of a molecule in the first active site could confer more stability to domain A in particular and, as a consequence, to domain C as well. Although difficult to identify in the x-, y- and z-direction, domain B shows similar behavior as already illustrated for the apoprotein. It behaves anti-correlated to domain A and C. The most apparent change occurs in domains C and D. The correlation of both domains is anti-correlated. This anti-correlation is more pronounced for the protein complex with the substrate in the active site I than for the apoprotein.

In the case a substrate is embedded, a change in correlation can be observed. While especially domains A and C seem to be strengthened in their anti-correlation, the same is true for the correlation between domains C and D. Since domain B also shows similar behavior as in the apoprotein, the general and apparently functionally important domain movement seems to be reflected overall.

#### **6.4.3 MFE-1 holoprotein**

Finally, quantification of the MFE-1 complex with embedded 3S-hydroxydecanoyl-CoA in active site I and NAD<sup>+</sup> present in active site II revealed differences when considering the spatial directions x, y and z (Figure 67).





*Figure 67: DCC matrix describing the correlation between domains of MFE-1. Red areas represent positive correlation between domains, blue areas represent anti-correlation between domains. MFE1 was aligned to domain D and E. 3S-hydroxydecanoyl-CoA is present in domain A and NAD<sup>+</sup> in domain C. X-, y- and z-directions refer to cartesian coordinates indicated by axis in Figure 52 chapter 6.2. DCC maps of the holoprotein are shown with the protonated glutamate 103 in x-, y- and z-direction.*

As with the apoprotein and the MFE-1 construct without NAD<sup>+</sup>, analysis of the DCC maps of the holoprotein in the x-, y-, and z-directions results in a comparable picture and all domains show a similar behavior. Again, domains A and C reveal a strong intramolecular correlation in the x-, y- and z-directions. The presence of NAD<sup>+</sup> does not seem to change the anti-correlation of domain A and C in the y-direction, whereas the x- and z-directions change in their behavior. In both spatial directions, the anti-correlations appear to apply only to individual residues, while the major part appear more correlated. Domain B, on the other hand, seems unaffected by the presence of NAD<sup>+</sup> and behaves anti-correlated to domains A and C in all three spatial directions. When looking at the intensities of the anti-correlation between domain C and D, there is a slight increase. In addition, the presence of NAD<sup>+</sup> also appears to affect the interaction of domain E and B. This interaction is most noticeable in this construct and was rather absent before. This fact probably indicates a positive correlation of domain A and E. Overall, the holoprotein demonstrates the interplay and predicted movement of the domains of MFE-1 that is consistent with the results obtained before by NMA and the other DCC maps. Additionally, the function of domain E is not directly apparent by these results.

## 6.5 Discussion

The first and second active site catalyze different reactions of the fatty acid metabolism. After the isomerization and hydrogenation reaction in the first active site, the intermediate product must be transported to the second active site. Crystal structure studies and corresponding investigations by the Wierenga group have suggested[233], that MFE-1 transport this by substrate channeling, for which certain domain interactions and domain movements seem to be necessary. However, due to a lack of dynamics studies, confirmation for this kind of movement is still missing. The molecular dynamics of MFE-1 investigated in this chapter could reveal those domain movements by normal mode analysis and dynamic cross correlation calculations and provide the basis for future dynamic structure investigations.

Similar results were obtained for all three examined constructs, the apoprotein, the MFE-1 3S-hydroxydecanoyl-CoA complex and the holoprotein including NAD<sup>+</sup> additionally. It was revealed by the NMA and DCC plots that four of the five domains in particular appear to play a supporting role for the functionality of MFE-1. When comparing all three structures, domains A and C in particular turn out to be very flexible. The predicted correlations could be supported by these results, whereby the consideration of the molecule movement in a certain spatial direction seems to play the decisive role. Thus, the domain flexibility predicted by Wierenga and co-workers[121] seems to take place in principle also in the recently published study[450], but is probably direction-dependent and hence difficult to interpret overall. Likewise, the predicted domain movements of domains C and D could basically be confirmed.

However, the correlation between domains A and C, also appears to be directional. In agreement with crystal structure analyses from 2010[121] and 2020[450], domain B in particular seems to assume an anchor function with very high stability at the same time. This stability could be shown for all three MFE-1 constructs and apparently confirms the function ascribed to domain B as a linker helix. Another very important finding emerges especially from the domain correlation due to the DCC investigation. Apparently, so far known from crystal structure studies, domain E plays a supporting role for domain B[121, 233]. The obtained results of the holoprotein suggest a supporting role to be possible but this has to be investigated in detail in future dynamic structure studies.

The overall results of this chapter seem to reinforce the assumptions based on crystal structure studies for the important domain correlation. According to the results shown here and to studies of Wirenga *et al.* [233], the domains A, C, and D seem to have very high flexibility. However, the results of the MFE-1 correlations have shown for the first time a possible mechanistic explanation of substrate channeling in connection with domain motions of A and C. Moreover, these findings provide a fundamental insight for further and more detailed structure-dynamics-function investigations in the future.

## 6.6 Summary

Due to the isomerization and hydrogenation reaction in the first and the dehydrogenation reaction in the second active site of MFE-1, the enzyme plays a central role in the  $\beta$ -oxidation of the fatty acid metabolism. As a representative of the enzyme family that relies on substrate channeling for the transport between their active sites, knowledge of the domain dynamics is of great importance. The molecular dynamics investigations performed in the scope of this dissertation on the respective domain motion of MFE-1, based on accelerated MD and dynamic cross correlation calculations. Various correlations between the domains for the three studied MFE-1 constructs are observable at the intra- and intermolecular level. In addition, due to the high rigidity and correlations with the four other domains, the role of domain B could be attributed as the linker helix as proposed before. Furthermore, the domain interactions showed that previously suspected movements of domains A, C, and D seem plausible.

Finally, calculated domain correlations and flexibilities contribute to the understanding of the substrate channeling as a part of the MFE-1 mechanism. Furthermore, these results also reveal MFE-1 to be a highly complex and generally very flexible enzyme. Therefore, these results provide a fundamental basis for future dynamics-function relations and further investigations with respect to the substrate channeling of MFE-1.



## 7 Conclusion and outlook

In this dissertation, the multifunctional enzyme type 1 (MFE-1) which catalyzes three reactions in the  $\beta$ -oxidation pathway of fatty acids was investigated using different constructs by molecular dynamics simulations in the form of classical and accelerated MD. Additionally, these constructs were investigated for fundamental dynamic structure-function relations using dynamic cross correlation and normal mode analysis. Moreover, a MFE-1 model with a substrate in the second active site, that had not been crystallized prior to this dissertation was generated by docking. All these investigations served to extend and furthermore deepens the fundamental knowledge at the atomistic level of the MFE-1 system. In the following, the main results and the corresponding conclusions are described.

MFE-1 is a complex enzyme with two active sites that has important functions in fatty acid metabolism. The first active site, located in the crotonase part, is responsible for the isomerization and hydrogenation of fatty acid substrates. Moreover, the second active site in the HAD part catalyzes the dehydrogenation reaction and yields in the product that is released from MFE-1. However, detailed information about the mechanism is still unknown. Residues involved in the reactions within the first active site could already be identified as well as assumptions of the mechanism are proposed in literature[**121, 122, 233**]. But knowledge about the protonation states of the key amino acids within the first active site (E103 and E123) and information of the presumed substrate channeling are missing. Furthermore, many potential ligands that can be metabolized by MFE-1 are poorly known also due to the unavailability of crystal structures with a variety of diverse ligand in the second active site.

Based on these facts, the goal of this dissertation was to investigate important mechanistic details of MFE-1. First, the basic protonation states of E103 and E123 in the first active site were calculated and the conformational behavior of these residues were studied referring to the proposed mechanism. In addition, the domain flexibility was examined for possible correlations to reveal behavioral patterns for possible substrate channeling. Finally, a 3D model with an occupied second active site was generated in ligand dependence to reveal binding relations for potential crystal structures.

The calculations performed on the apoprotein and the MFE-1 complex with 3S-hydroxydecanoyl-CoA, reveal the protonation state that was determined as protonated

for E103 and deprotonated for E123. This is consistent with the protonation states predicted by Kasaragod *et al.*[122]. Considering the complete reaction mechanism, this calculation does not indicate a possible change in protonation state at certain intermediate steps. However, this should be investigated in more detail in future studies for the mechanism to reveal potential proton transfer pathways in conjunction with the enzymatic reactions.

In addition to that, in this dissertation it has been observed, for the first active site, that the MFE-1 appears to adopt two specific states (open and closed) in conjunction with the two glutamates (E103 and E123). As also shown, these changes are associated with the water network. In this respect, the state of a water-catalyzed mechanism of the first active site predicted by Kasaragod *et al.*[122, 233] seems plausible since it agrees with the results described here. Thus, future studies with ligand variations represent another important investigation to uncover the extent to which this substrate- or substrate-water-dependent state or the water interaction changes.

Furthermore, it could be shown that defined correlations within the domains take place when considering the general domain dynamics, but that these correlations are also present on an intermolecular level. In this context, the predicted open-close domain motions described by the Wierenga group[121, 233, 450] are reflected by the obtained results. This showed that MFE-1 basically has intrinsic domain dynamics as well as domain correlations, which could be shown for all three investigated MFE-1 constructs (apoprotein, MFE-1 complex with 3S-hydroxydecanol-CoA and holoprotein) except for minor deviations.

The results show correlations between domains AC and CD, with motion correlations between domains A and C appearing to be most pronounced. Considering that active site I is located in domain A and active site II in domain C, the correlation of functional directional movement could be indicative of substrate channeling function in MFE-1. Moreover, the CD correlation could be an additional indication of product release from the second active site.

Finally, a plausible substrate was modeled into the second active site using docking. From a large number of possible substrates ranked by their binding energies, a specific proportion of potential ligands (fatty acids) could be identified based on MFE-1 substrate specificity. Based on the crystal structure with the known ligand 3S-hydroxydecanoyl-CoA,

which was previously only crystallized while embedded in the first active site, this substrate was used to create a 3D model of the second active site. The cavities present in the second binding pocket were considered to determine the ligand and the binding conformation of the substrate. By this method the basic 3D structure with 3S-ydroxydecanoyl-CoA embedded in the second active site was successfully modeled for the first time. This model is also in good agreement with the recently published study by Wierenga *et al.*[450]. Future dynamic structure studies could use this model as an initiator, in particular to serve a more detailed structural elucidation of the second active site.

These results show the high complexity of the MFE-1, but this dissertation has provided essential insights of the mechanism. The results of this doctoral thesis contribute to the general understanding of the MFE-1 and provide data and models that can be used for future investigations to uncover even more and more details and mechanistic properties of the MFE-1.



## 8 References

1. van Meer, G., *Cellular lipidomics*. *Embo j*, 2005. **24**(18): p. 3159-65.
2. Fahy, E., et al., *A comprehensive classification system for lipids*. *J Lipid Res*, 2005. **46**(5): p. 839-61.
3. Constantinides, P.P. and J.M. Steim, *Physical properties of fatty acyl-CoA. Critical micelle concentrations and micellar size and shape*. *J Biol Chem*, 1985. **260**(12): p. 7573-80.
4. Lipmann, F. and N.O. Kaplan, *A common factor in the enzymatic acetylation of sulfanilamiden and of choline*. *J Biol Chem*, 1946. **162**: p. 743-744.
5. Baddiley, J., et al., *Structure of coenzyme A*. *Nature*, 1953. **171**(4341): p. 76.
6. Kragelund, B.B., J. Knudsen, and F.M. Poulsen, *Acyl-coenzyme A binding protein (ACBP)*. *Biochim Biophys Acta*, 1999. **1441**(2-3): p. 150-61.
7. D'Ordine, R.L., P. Paneth, and V. Anderson, *<sup>13</sup>C NMR and <sup>1</sup>H-<sup>1</sup>H NOEs of Coenzyme-A: Conformation of the Pantoic Acid Moiety*. *Bioorganic Chemistry*, 1995. **23**: p. 169-181.
8. Wu, W.-J., P.J. Tonge, and D.P. Raleigh, *Stereospecific <sup>1</sup>H and <sup>13</sup>C NMR Assignments of Crotonyl CoA and Hexadienoyl CoA: Conformational Analysis and Comparison with Protein-CoA Complexes*. *Journal of the American Chemical Society*, 1998. **120**(39): p. 9988-9994.
9. Engel, C. and R. Wierenga, *The diverse world of coenzyme A binding proteins*. *Curr Opin Struct Biol*, 1996. **6**(6): p. 790-7.
10. Ofman, R., et al., *Proteomic analysis of mouse kidney peroxisomes: identification of RP2p as a peroxisomal nudix hydrolase with acyl-CoA diphosphatase activity*. *Biochem J*, 2006. **393**(Pt 2): p. 537-43.
11. Mishra, P.K. and D.G. Drueckhammer, *Coenzyme A Analogues and Derivatives: Synthesis and Applications as Mechanistic Probes of Coenzyme A Ester-Utilizing Enzymes*. *Chem Rev*, 2000. **100**(9): p. 3283-3310.
12. Kursula, P., et al., *The catalytic cycle of biosynthetic thiolase: a conformational journey of an acetyl group through four binding modes and two oxyanion holes*. *Biochemistry*, 2002. **41**(52): p. 15543-56.
13. Bhaumik, P., et al., *Structural biology of the thioester-dependent degradation and synthesis of fatty acids*. *Curr Opin Struct Biol*, 2005. **15**(6): p. 621-8.
14. White, S.W., et al., *The structural biology of type II fatty acid biosynthesis*. *Annu Rev Biochem*, 2005. **74**: p. 791-831.
15. Rawlings, M. and J.E. Cronan, Jr., *The gene encoding Escherichia coli acyl carrier protein lies within a cluster of fatty acid biosynthetic genes*. *J Biol Chem*, 1992. **267**(9): p. 5751-4.
16. Schulte, U., *Biogenesis of respiratory complex I*. *J Bioenerg Biomembr*, 2001. **33**(3): p. 205-12.
17. Roujeinikova, A., et al., *X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site*. *Structure*, 2002. **10**(6): p. 825-35.
18. Hancock, W.S., et al., *Acyl carrier protein. 28. Chemical synthesis and characterization of a protein with acyl carrier protein activity*. *J Biol Chem*, 1972. **247**(19): p. 6224-33.

19. Zhang, Y.M., et al., *Identification and analysis of the acyl carrier protein (ACP) docking site on beta-ketoacyl-ACP synthase III*. J Biol Chem, 2001. **276**(11): p. 8231-8.
20. Holak, T.A., et al., *Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry*. Eur J Biochem, 1988. **175**(1): p. 9-15.
21. Kim, Y., J.B. Ohlrogge, and J.H. Prestegard, *Motional effects on NMR structural data. Comparison of spinach and Escherichia coli acyl carrier proteins*. Biochem Pharmacol, 1990. **40**(1): p. 7-13.
22. Honeyman, G. and T. Fawcett, *Protein interactions of fatty acid synthase II*. Biochem Soc Trans, 2000. **28**(6): p. 615-6.
23. Lipmann, F., *Bacterial production of antibiotic polypeptides by thiol-linked synthesis on protein templates*. Adv Microb Physiol, 1980. **21**: p. 227-66.
24. Rock, C.O. and J.E. Cronan, *Escherichia coli as a model for the regulation of dissociable (type II) fatty acid biosynthesis*. Biochim Biophys Acta, 1996. **1302**(1): p. 1-16.
25. Xu, G.Y., et al., *Solution structure of B. subtilis acyl carrier protein*. Structure, 2001. **9**(4): p. 277-87.
26. Zhang, Y.M., et al., *Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction*. J Biol Chem, 2003. **278**(52): p. 52935-43.
27. Rock, C.O., J.E. Cronan, Jr., and I.M. Armitage, *Molecular properties of acyl carrier protein derivatives*. J Biol Chem, 1981. **256**(6): p. 2669-74.
28. Bar-Tana, J., et al., *Palmitoyl-coenzyme A synthetase. Mechanism of reaction*. Biochem J, 1973. **131**(2): p. 199-209.
29. Vanhove, G., et al., *Mitochondrial and peroxisomal beta oxidation of the branched chain fatty acid 2-methylpalmitate in rat liver*. J Biol Chem, 1991. **266**(36): p. 24670-5.
30. Steinberg, S.J., et al., *Human very-long-chain acyl-CoA synthetase: cloning, topography, and relevance to branched-chain fatty acid metabolism*. Biochem Biophys Res Commun, 1999. **257**(2): p. 615-21.
31. Van Veldhoven, P.P., et al., *Stabilisation and partial purification of Triton X-100 solubilised trihydroxycoprostanoyl-CoA synthetase from rat liver*. Biochem Mol Biol Int, 1996. **40**(3): p. 447-57.
32. Fujino, T., et al., *Molecular identification and characterization of two medium-chain acyl-CoA synthetases, MACS1 and the Sa gene product*. J Biol Chem, 2001. **276**(38): p. 35961-6.
33. Reddy, J.K. and T. Hashimoto, *Peroxisomal beta-oxidation and peroxisome proliferator-activated receptor alpha: an adaptive metabolic system*. Annu Rev Nutr, 2001. **21**: p. 193-230.
34. Aas, M., *Organ and subcellular distribution of fatty acid activating enzymes in the rat*. Biochim Biophys Acta, 1971. **231**(1): p. 32-47.
35. Hesler, C.B., C. Olymbios, and D. Halder, *Transverse-plane topography of long-chain acyl-CoA synthetase in the mitochondrial outer membrane*. J Biol Chem, 1990. **265**(12): p. 6600-5.
36. Pande, S.V., *A mitochondrial carnitine acylcarnitine translocase system*. Proc Natl Acad Sci U S A, 1975. **72**(3): p. 883-7.
37. Eaton, S., K. Bartlett, and M. Pourfarzam, *Mammalian mitochondrial beta-oxidation*. Biochem J, 1996. **320** ( Pt 2)(Pt 2): p. 345-57.

38. Vaz, F.M. and R.J.A. Wanders, *Carnitine biosynthesis in mammals*. The Biochemical journal, 2002. **361**(Pt 3): p. 417-429.
39. Kerner, J. and C. Hoppel, *Fatty acid import into mitochondria*. Biochim Biophys Acta, 2000. **1486**(1): p. 1-17.
40. Mannaerts, G.P., et al., *Evidence that peroxisomal acyl-CoA synthetase is located at the cytoplasmic side of the peroxisomal membrane*. Biochem J, 1982. **204**(1): p. 17-23.
41. Lazo, O., et al., *Cellular oxidation of lignoceric acid is regulated by the subcellular localization of lignoceroyl-CoA ligases*. J Lipid Res, 1990. **31**(4): p. 583-95.
42. Hashimoto, T., *Peroxisomal beta-oxidation: enzymology and molecular biology*. Ann N Y Acad Sci, 1996. **804**: p. 86-98.
43. Wanders, R.J., et al., *Characteristics and subcellular localization of pristanoyl-CoA synthetase in rat liver*. Biochim Biophys Acta, 1992. **1125**(3): p. 274-9.
44. Antonenkov, V.D., R.T. Sormunen, and J.K. Hiltunen, *The rat liver peroxisomal membrane forms a permeability barrier for cofactors but not for small metabolites in vitro*. J Cell Sci, 2004. **117**(Pt 23): p. 5633-42.
45. Suzuki, H., et al., *Structure and regulation of rat long-chain acyl-CoA synthetase*. J Biol Chem, 1990. **265**(15): p. 8681-5.
46. Singh, I., et al., *Transport of fatty acids into human and rat peroxisomes. Differential transport of palmitic and lignoceric acids and its implication to X-adrenoleukodystrophy*. J Biol Chem, 1992. **267**(19): p. 13306-13.
47. Mihalik, S.J., et al., *Participation of two members of the very long-chain acyl-CoA synthetase family in bile acid synthesis and recycling*. J Biol Chem, 2002. **277**(27): p. 24771-9.
48. Ockner, R.K., et al., *A binding protein for fatty acids in cytosol of intestinal mucosa, liver, myocardium, and other tissues*. Science, 1972. **177**(4043): p. 56-8.
49. Veerkamp, J.H. and R.G. Maatman, *Cytoplasmic fatty acid-binding proteins: their structure and genes*. Prog Lipid Res, 1995. **34**(1): p. 17-52.
50. Kursula, P., et al., *Crystal structure of human fatty acid binding protein 1 (FABP1)*. 2005: PDB Database.
51. Wang, G., et al., *Recent insights into the biological functions of liver fatty acid binding protein 1*. Journal of lipid research, 2015. **56**(12): p. 2238-2247.
52. *Cellular Proteins and Their Fatty Acids in Health and Disease*. 2003, Weinheim, Germany: Wiley - VCH Verlag GmbH & Co.
53. Glatz, J.F. and G.J. van der Vusse, *Cellular fatty acid-binding proteins: their function and physiological significance*. Prog Lipid Res, 1996. **35**(3): p. 243-82.
54. Bernlohr, D.A., et al., *Intracellular lipid-binding proteins and their genes*. Annu Rev Nutr, 1997. **17**: p. 277-303.
55. Coe, N.R. and D.A. Bernlohr, *Physiological properties and functions of intracellular fatty acid-binding proteins*. Biochim Biophys Acta, 1998. **1391**(3): p. 287-306.
56. Schaap, F.G., G.J. van der Vusse, and J.F. Glatz, *Fatty acid-binding proteins in the heart*. Mol Cell Biochem, 1998. **180**(1-2): p. 43-51.
57. Simpson, M.A., et al., *Biochemical and biophysical analysis of the intracellular lipid binding proteins of adipocytes*. Mol Cell Biochem, 1999. **192**(1-2): p. 33-40.
58. Bernlohr, D.A., N.R. Coe, and V.J. LiCata, *Fatty acid trafficking in the adipocyte*. Semin Cell Dev Biol, 1999. **10**(1): p. 43-9.
59. Hamilton, J.A., et al., *Fatty acid transport: the diffusion mechanism in model and biological membranes*. J Mol Neurosci, 2001. **16**(2-3): p. 99-108; discussion 151-7.

- 
60. Hamilton, J.A., *How fatty acids bind to proteins: the inside story from protein structures*. Prostaglandins Leukot Essent Fatty Acids, 2002. **67**(2-3): p. 65-72.
  61. Reese, A.J. and L.J. Banaszak, *Specificity determinants for lipids bound to beta-barrel proteins*. J Lipid Res, 2004. **45**(2): p. 232-43.
  62. LaLonde, J.M., D.A. Bernlohr, and L.J. Banaszak, *X-ray crystallographic structures of adipocyte lipid-binding protein complexed with palmitate and hexadecanesulfonic acid. Properties of cavity binding sites*. Biochemistry, 1994. **33**(16): p. 4885-95.
  63. Ory, J.J. and L.J. Banaszak, *Studies of the ligand binding reaction of adipocyte lipid binding protein using the fluorescent probe 1, 8-anilinonaphthalene-8-sulfonate*. Biophys J, 1999. **77**(2): p. 1107-16.
  64. Noy, N., *Retinoid-binding proteins: mediators of retinoid action*. Biochem J, 2000. **348 Pt 3**(Pt 3): p. 481-95.
  65. Sacchettini, J.C., et al., *Developmental and structural studies of an intracellular lipid binding protein expressed in the ileal epithelium*. J Biol Chem, 1990. **265**(31): p. 19199-207.
  66. Sacchettini, J.C., J.I. Gordon, and L.J. Banaszak, *The structure of crystalline Escherichia coli-derived rat intestinal fatty acid-binding protein at 2.5-A resolution*. J Biol Chem, 1988. **263**(12): p. 5815-9.
  67. Cowan, S.W., M.E. Newcomer, and T.A. Jones, *Crystallographic studies on a family of cellular lipophilic transport proteins. Refinement of P2 myelin protein and the structure determination and refinement of cellular retinol-binding protein in complex with all-trans-retinol*. J Mol Biol, 1993. **230**(4): p. 1225-46.
  68. Zurzolo, C., et al., *VIP21/caveolin, glycosphingolipid clusters and the sorting of glycosylphosphatidylinositol-anchored proteins in epithelial cells*. Embo j, 1994. **13**(1): p. 42-53.
  69. Szyperski, T., et al., *NMR determination of the secondary structure and the three-dimensional polypeptide backbone fold of the human sterol carrier protein 2*. FEBS Lett, 1993. **335**(1): p. 18-26.
  70. García, F.L., et al., *NMR structure of the sterol carrier protein-2: implications for the biological role*. J Mol Biol, 2000. **295**(3): p. 595-603.
  71. Choinowski, T., H. Hauser, and K. Piontek, *Structure of sterol carrier protein 2 at 1.8 Å resolution reveals a hydrophobic tunnel suitable for lipid binding*. Biochemistry, 2000. **39**(8): p. 1897-902.
  72. Seedorf, U., et al., *Defective peroxisomal catabolism of branched fatty acyl coenzyme A in mice lacking the sterol carrier protein-2/sterol carrier protein-x gene function*. Genes Dev, 1998. **12**(8): p. 1189-201.
  73. Haapalainen, A.M., et al., *Crystal structure of the liganded SCP-2-like domain of human peroxisomal multifunctional enzyme type 2 at 1.75 Å resolution*. J Mol Biol, 2001. **313**(5): p. 1127-38.
  74. Hoh, F., et al., *Structure of a liganded type 2 non-specific lipid-transfer protein from wheat and the molecular basis of lipid binding*. Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 4): p. 397-406.
  75. Chye, M.L., B.Q. Huang, and S.Y. Zee, *Isolation of a gene encoding Arabidopsis membrane-associated acyl-CoA binding protein and immunolocalization of its gene product*. Plant J, 1999. **18**(2): p. 205-14.
  76. Chye, M.L., H.Y. Li, and M.H. Yung, *Single amino acid substitutions at the acyl-CoA-binding domain interrupt 14[C]palmitoyl-CoA binding of ACP2, an Arabidopsis*
-



- acyl-CoA-binding protein with ankyrin repeats*. Plant Mol Biol, 2000. **44**(6): p. 711-21.
77. Geisbrecht, B.V., et al., *Characterization of PECL, a novel monofunctional Delta(3), Delta(2)-enoyl-CoA isomerase of mammalian peroxisomes*. J Biol Chem, 1999. **274**(31): p. 21797-803.
78. Dangi, R.S., R.K. Pal, and M. Sundd, *Crystal Structure of an Acyl-CoA Binding Protein (LmjF.17.0620) of Leishmania major*. 2017: PDB Databank.
79. Kragelund, B.B., et al., *Three-dimensional structure of the complex between acyl-coenzyme A binding protein and palmitoyl-coenzyme A*. J Mol Biol, 1993. **230**(4): p. 1260-77.
80. van Aalten, D.M., et al., *Binding site differences revealed by crystal structures of Plasmodium falciparum and bovine acyl-CoA binding protein*. J Mol Biol, 2001. **309**(1): p. 181-92.
81. Costa, E. and A. Guidotti, *Diazepam binding inhibitor (DBI): a peptide with multiple biological actions*. Life Sci, 1991. **49**(5): p. 325-44.
82. Watanabe, S., et al., *Macrocycles used as models to probe the interaction of a fatty acid derivative with its natural receptor*. Pure & Appl. Chem, 1993. **65**(3): p. 467-472.
83. Kim, J.J. and K.P. Battaile, *Burning fat: the structural basis of fatty acid beta-oxidation*. Curr Opin Struct Biol, 2002. **12**(6): p. 721-8.
84. Faergeman, N.J., et al., *Thermodynamics of ligand binding to acyl-coenzyme A binding protein studied by titration calorimetry*. Biochemistry, 1996. **35**(45): p. 14118-26.
85. Van Veldhoven, P.P., et al., *Identification and purification of a peroxisomal branched chain fatty acyl-CoA oxidase*. J Biol Chem, 1991. **266**(36): p. 24676-83.
86. Kundu, R.K., G.S. Getz, and J.H. Tonsgard, *Induction of (omega-1)-oxidation of monocarboxylic acids by acetylsalicylic acid*. J Lipid Res, 1993. **34**(7): p. 1187-99.
87. Hiltunen, J.K. and Y. Qin, *beta-oxidation - strategies for the metabolism of a wide variety of acyl-CoA esters*. Biochim Biophys Acta, 2000. **1484**(2-3): p. 117-28.
88. Reddy, J.K. and G.P. Mannaerts, *Peroxisomal lipid metabolism*. Annu Rev Nutr, 1994. **14**: p. 343-70.
89. Mannaerts, G.P., P.P. Van Veldhoven, and M. Casteels, *Peroxisomal lipid degradation via beta- and alpha-oxidation in mammals*. Cell Biochem Biophys, 2000. **32 Spring**: p. 73-87.
90. Carpenter, K., R.J. Pollitt, and B. Middleton, *Human liver long-chain 3-hydroxyacyl-coenzyme A dehydrogenase is a multifunctional membrane-bound beta-oxidation enzyme of mitochondria*. Biochem Biophys Res Commun, 1992. **183**(2): p. 443-8.
91. Uchida, Y., et al., *Novel fatty acid beta-oxidation enzymes in rat liver mitochondria. II. Purification and properties of enoyl-coenzyme A (CoA) hydratase/3-hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase trifunctional protein*. J Biol Chem, 1992. **267**(2): p. 1034-41.
92. Kamijo, T., et al., *Molecular cloning of the cDNAs for the subunits of rat mitochondrial fatty acid beta-oxidation multienzyme complex. Structural and functional relationships to other mitochondrial and peroxisomal beta-oxidation enzymes*. J Biol Chem, 1993. **268**(35): p. 26452-60.
93. Ishikawa, M., et al., *Reconstitution, morphology and crystallization of a fatty acid beta-oxidation multienzyme complex from Pseudomonas fragi*. Biochem J, 1997. **328 ( Pt 3)**(Pt 3): p. 815-20.

94. Kamijo, T., et al., *Structural analysis of cDNAs for subunits of human mitochondrial fatty acid beta-oxidation trifunctional protein*. Biochem Biophys Res Commun, 1994. **199**(2): p. 818-25.
95. Ishikawa, M., et al., *Structural basis for channelling mechanism of a fatty acid beta-oxidation multienzyme complex*. Embo j, 2004. **23**(14): p. 2745-54.
96. Yao, K.W. and H. Schulz, *Intermediate channeling on the trifunctional beta-oxidation complex from pig heart mitochondria*. J Biol Chem, 1996. **271**(30): p. 17816-20.
97. Eaton, S. and K. Bartlett, *Tissue specific differences in intramitochondrial control of beta-oxidation*. Adv Exp Med Biol, 1999. **466**: p. 161-8.
98. Liang, K., et al., *Cryo-EM structure of human mitochondrial trifunctional protein*. Proc Natl Acad Sci U S A, 2018. **115**(27): p. 7039-7044.
99. Taskinen, J.P., et al., *Structural studies of MFE-1: the 1.9 Å crystal structure of the dehydrogenase part of rat peroxisomal MFE-1*. J Mol Biol, 2006. **355**(4): p. 734-46.
100. *The Proteomics Protocols Handbook*. 2005, Totowa, NJ: Humana Press, Inc.
101. Holden, H.M., et al., *The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme A thioesters*. Acc Chem Res, 2001. **34**(2): p. 145-57.
102. Kiema, T.R., et al., *Organization of the multifunctional enzyme type 1: interaction between N- and C-terminal domains is required for the hydratase-1/isomerase activity*. Biochem J, 2002. **367**(Pt 2): p. 433-41.
103. Savolainen, K., et al., *A mouse model for alpha-methylacyl-CoA racemase deficiency: adjustment of bile acid synthesis and intolerance to dietary methyl-branched lipids*. Hum Mol Genet, 2004. **13**(9): p. 955-65.
104. Nguyen, S.D., M. Baes, and P.P. Van Veldhoven, *Degradation of very long chain dicarboxylic polyunsaturated fatty acids in mouse hepatocytes, a peroxisomal process*. Biochim Biophys Acta, 2008. **1781**(8): p. 400-5.
105. Ferdinandusse, S., et al., *Identification of the peroxisomal beta-oxidation enzymes involved in the degradation of long-chain dicarboxylic acids*. J Lipid Res, 2004. **45**(6): p. 1104-11.
106. Kurosawa, T., et al., *Conjugation reactions catalyzed by bifunctional proteins related to beta-oxidation in bile acid biosynthesis*. Steroids, 2001. **66**(2): p. 107-14.
107. Xu, R. and D.A. Cuebas, *The reactions catalyzed by the inducible bifunctional enzyme of rat liver peroxisomes cannot lead to the formation of bile acids*. Biochem Biophys Res Commun, 1996. **221**(2): p. 271-8.
108. Palosaari, P.M., et al., *Amino acid sequence similarities of the mitochondrial short chain delta 3, delta 2-enoyl-CoA isomerase and peroxisomal multifunctional delta 3, delta 2-enoyl-CoA isomerase, 2-enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase enzyme in rat liver. The proposed occurrence of isomerization and hydration in the same catalytic domain of the multifunctional enzyme*. J Biol Chem, 1991. **266**(17): p. 10750-3.
109. Zhang, D., et al., *Functional characterization of Delta3,Delta2-enoyl-CoA isomerases from rat liver*. J Biol Chem, 2002. **277**(11): p. 9127-32.
110. Filppula, S.A., et al., *Delta3,5-delta2,4-dienoyl-CoA isomerase from rat liver. Molecular characterization*. J Biol Chem, 1998. **273**(1): p. 349-55.
111. Kiema, T.R., et al., *Mutagenic and enzymological studies of the hydratase and isomerase activities of 2-enoyl-CoA hydratase-1*. Biochemistry, 1999. **38**(10): p. 2991-9.

- 
112. Palosaari, P.M. and J.K. Hiltunen, *Peroxisomal bifunctional protein from rat liver is a trifunctional enzyme possessing 2-enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and delta 3, delta 2-enoyl-CoA isomerase activities*. J Biol Chem, 1990. **265**(5): p. 2446-9.
  113. Yang, S.Y., D. Cuebas, and H. Schulz, *Channeling of 3-hydroxy-4-trans-decenoyl coenzyme A on the bifunctional beta-oxidation enzyme from rat liver peroxisomes and on the large subunit of the fatty acid oxidation complex from Escherichia coli*. J Biol Chem, 1986. **261**(33): p. 15390-5.
  114. Miles, E.W., S. Rhee, and D.R. Davies, *The Molecular Basis of Substrate Channeling* THE JOURNAL OF BIOLOGICAL CHEMISTRY, 1999. **274**(18): p. 12193-12196.
  115. Raushel, F.M., J.B. Thoden, and H.M. Holden, *Enzymes with molecular tunnels*. Acc Chem Res, 2003. **36**(7): p. 539-48.
  116. Holden, H.M., J.B. Thoden, and F.M. Raushel, *Carbamoyl phosphate synthetase: a tunnel runs through it*. Curr Opin Struct Biol, 1998. **8**(6): p. 679-85.
  117. Hyde, C.C., et al., *Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium*. J Biol Chem, 1988. **263**(33): p. 17857-71.
  118. Manjasetty, B.A., J. Powlowski, and A. Vrielink, *Crystal structure of a bifunctional aldolase-dehydrogenase: sequestering a reactive and volatile intermediate*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6992-7.
  119. Elcock, A.H., et al., *Electrostatic channeling in the bifunctional enzyme dihydrofolate reductase-thymidylate synthase*. J Mol Biol, 1996. **262**(3): p. 370-4.
  120. Atreya, C.E., et al., *Probing electrostatic channeling in protozoal bifunctional thymidylate synthase-dihydrofolate reductase using site-directed mutagenesis*. J Biol Chem, 2003. **278**(31): p. 28901-11.
  121. Kasaragod, P., et al., *Crystal structure of liganded rat peroxisomal multifunctional enzyme type 1: a flexible molecule with two interconnected active sites*. J Biol Chem, 2010. **285**(31): p. 24089-98.
  122. Kasaragod, P., et al., *The isomerase and hydratase reaction mechanism of the crotonase active site of the multifunctional enzyme (type-1), as deduced from structures of complexes with 3S-hydroxy-acyl-CoA*. Febs j, 2013. **280**(13): p. 3160-75.
  123. Hamed, R.B., et al., *Mechanisms and structures of crotonase superfamily enzymes--how nature controls enolate and oxyanion reactivity*. Cell Mol Life Sci, 2008. **65**(16): p. 2507-27.
  124. Barycki, J.J., et al., *Sequestration of the active site by interdomain shifting. Crystallographic and spectroscopic evidence for distinct conformations of L-3-hydroxyacyl-CoA dehydrogenase*. J Biol Chem, 2000. **275**(35): p. 27186-96.
  125. Barycki, J.J., et al., *Pig heart short chain L-3-hydroxyacyl-CoA dehydrogenase revisited: sequence analysis and crystal structure determination*. Protein Sci, 1999. **8**(10): p. 2010-8.
  126. Partanen, S.T., et al., *The 1.3 A crystal structure of human mitochondrial Delta3-Delta2-enoyl-CoA isomerase shows a novel mode of binding for the fatty acyl group*. J Mol Biol, 2004. **342**(4): p. 1197-208.
  127. Engel, C.K., et al., *Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: a spiral fold defines the CoA-binding pocket*. Embo j, 1996. **15**(19): p. 5135-45.
-

128. Hofstein, H.A., et al., *Role of glutamate 144 and glutamate 164 in the catalytic mechanism of enoyl-CoA hydratase*. *Biochemistry*, 1999. **38**(29): p. 9508-16.
129. Bahnson, B.J., V.E. Anderson, and G.A. Petsko, *Structural mechanism of enoyl-CoA hydratase: three atoms from a single water are added in either an E1cb stepwise or concerted fashion*. *Biochemistry*, 2002. **41**(8): p. 2621-9.
130. Mursula, A.M., J.K. Hiltunen, and R.K. Wierenga, *Structural studies on delta(3)-delta(2)-enoyl-CoA isomerase: the variable mode of assembly of the trimeric disks of the crotonase superfamily*. *FEBS Lett*, 2004. **557**(1-3): p. 81-7.
131. Spivey, H.O. and J. Ovádi, *Substrate channeling*. *Methods*, 1999. **19**(2): p. 306-321.
132. Anderson, K.S., *Fundamental mechanisms of substrate channeling*. *Methods Enzymol.*, 1999. **308**: p. 111-145.
133. Hakobyan, D. and K. Nazaryan, *Molecular dynamics simulation of interactions in glycolytic enzymes*. *Biochemistry (Moscow)*, 2006. **71**: p. 370-375.
134. Srere, P.A., *COMPLEXES OF SEQUENTIAL METABOLIC ENZYMES*. *Annual Review of Biochemistry*, 1987. **56**(1): p. 89-124.
135. Wilson, K.J., et al., *The behaviour of peptides on reverse-phase supports during high-pressure liquid chromatography*. *Biochem J.*, 1981. **199**(1): p. 31-41.
136. Westerhof, H.V. and G.R. Welch, *Enzyme organization and the direction of metabolic flow: physicochemical considerations*. *Curr Top Cell Regul.*, 1992. **33**: p. 361-390.
137. Rudolph, J. and J. Stubbe, *Investigation of the Mechanism of Phosphoribosylamine Transfer from Glutamine Phosphoribosylpyrophosphate Amidotransferase to Glycinamide Ribonucleotide Synthetase*. *Biochemistry*, 1995. **34**(7): p. 2241-2250.
138. Ushiroyama, T., et al., *Substrate channeling of NADH in mitochondrial redox processes*. *Curr Top Cell Regul.*, 1992. **33**: p. 291-307.
139. Ovádi, J., Y. Huang, and H.O. Spivey, *Binding of malate dehydrogenase and NADH channelling to complex I*. *J Mol Recognit.*, 1994. **7**(4): p. 265-272.
140. Dewar, M.J. and D.M. Storch, *Alternative view of enzyme reactions*. *Proc. Natl. Acad. Sci. USA*, 1985. **82**(8): p. 2225-2229.
141. Wakil, S.J., J.K. Stoops, and V.C. Joshi, *Fatty acid synthesis and its regulation*. *Annu Rev. Biochem.*, 1983. **52**: p. 537-579.
142. Batke, J., *Channeling of glycolytic intermediates by temporary, stationary bi-enzyme complexes is probable in vivo*. *Trends Biochem Sci.*, 1989. **14**(12): p. 481-482.
143. Mathews, C.K. and N.H. Sinha, *Are DNA precursors concentrated at replication sites?* *Proc. Natl. Acad. Sci. USA*, 1982. **79**(2): p. 302-306.
144. Ovádi, J. and T. Keleti, *Kinetic evidence for interaction between aldolase and D-glyceraldehyde-3-phosphate dehydrogenase*. *Eur. J. Biochem.*, 1978. **85**(1): p. 157-161.
145. Welch, G.R., *On the role of organized multienzyme systems in cellular metabolism: a general synthesis*. *Prog. Biophys. Mol. Biol.*, 1977. **32**(2): p. 103-191.
146. Keleti, T. and J. Ovádi, *Control of metabolism by dynamic macromolecular interactions*. *Curr. Top Cell Regul.*, 1988. **29**: p. 1-33.
147. Vértessy, B. and J. Ovádi, *A simple approach to detect active-site-directed enzyme-enzyme interactions. The aldolase/glycerol-phosphate-dehydrogenase enzyme system*. *Eur. J. Biochem.*, 1987. **164**(3): p. 655-659.

148. Huang, X., H.M. Holden, and F.M. Raushel, *Channeling of substrates and intermediates in enzyme-catalyzed reactions*. Annu. Rev. Biochem., 2001. **70**: p. 149-180.
149. Miles, E.W., *Tryptophan synthase: structure, function, and subunit interaction*. Adv Enzymol Relat Areas Mol Biol, 1979. **49**: p. 127-86.
150. Kirschner, K., et al., *The Tryptophan Synthase from Escherichia coli*. Eur. J. Biochem, 1975. **60**: p. 511-523.
151. Miles, E.W., *Structural basis for catalysis by tryptophan synthase*. Adv Enzymol Relat Areas Mol Biol, 1991. **64**: p. 93-172.
152. Ruvinov, S.B., et al., *Monovalent Cations Partially Repair a Conformational Defect in a Mutant Tryptophan Synthase Complex (-E109A)*. The Journal of Biological Chemistry, 1995. **270**(29): p. 17333-17338.
153. Pan, P., E. Woehl, and M.F. Dunn, *Protein architecture, dynamics and allostery in tryptophan synthase channeling*. Trends Biochem Sci, 1997. **22**(1): p. 22-7.
154. Yanofsky, C. and I.P. Crawford, *The Enzymes*. Boyer, P. D., ed. Academic Press, New York, 1972. **7**: p. 1-31.
155. Rhee, S., et al., *Crystal structures of a mutant (betaK87T) tryptophan synthase alpha2beta2 complex with ligands bound to the active sites of the alpha- and beta-subunits reveal ligand-induced conformational changes*. Biochemistry, 1997. **36**(25): p. 7664-80.
156. Yanofsky, C. and M. Rachmeler, *The exclusion of free indole as an intermediate in the biosynthesis of tryptophan in Neurospora crassa*. Biochimica et biophysica acta, 1958. **28**(3): p. 640-641.
157. Demoss, J.A., *Studies on the mechanism of the tryptophan synthetase reaction*. Biochimica et Biophysica Acta, 1962. **62**(2): p. 279-293.
158. Creighton, T.E., *A steady-state kinetic investigation of the reaction mechanism of the tryptophan synthetase of Escherichia coli*. Eur J Biochem, 1970. **13**(1): p. 1-10.
159. Matchet, W.H., *Indole Channeling by Tryptophan Synthase of Neurospora*. THE JOURNAL of Biological CHEMISTRY, 1973. **249**(13): p. 4041-4049.
160. Rhee, S., et al., *Exchange of K<sup>+</sup> or Cs<sup>+</sup> for Na<sup>+</sup> induces local and long-range changes in the three-dimensional structure of the tryptophan synthase alpha2beta2 complex*. Biochemistry, 1996. **35**(13): p. 4211-21.
161. Rhee, S., E.W. Miles, and D.R. Davies, *Cryo-crystallography of a true substrate, indole-3-glycerol phosphate, bound to a mutant (alphaD60N) tryptophan synthase alpha2beta2 complex reveals the correct orientation of active site alphaGlu49*. J Biol Chem, 1998. **273**(15): p. 8553-5.
162. Rhee, S., et al., *Cryocrystallography and microspectrophotometry of a mutant (alpha D60N) tryptophan synthase alpha 2 beta 2 complex reveals allosteric roles of alpha Asp60*. Biochemistry, 1998. **37**(30): p. 10653-9.
163. Schneider, T.R., et al., *Loop closure and intersubunit communication in tryptophan synthase*. Biochemistry, 1998. **37**(16): p. 5394-406.
164. Anderson, P.M. and A. Meister, *Control of Escherichia coli carbamyl phosphate synthetase by purine and pyrimidine nucleotides*. Biochemistry, 1966. **5**(10): p. 3164-9.
165. Anderson, P.M. and A. Meister, *Evidence for an Activated Form of Carbon Dioxide in the Reaction Catalyzed by Escherichia coli Carbamyl Phosphate Synthetase\**. Biochemistry, 1965. **4**(12): p. 2803-2809.
166. Sauers, C.K., W.P. Jencks, and S. Groh, *Alcohol-bicarbonate-water system. Structure-reactivity studies on the equilibriums for formation of alkyl*

- monocarbonates and on the rates of their decomposition in aqueous alkali.* Journal of the American Chemical Society, 1975. **97**(19): p. 5546-5553.
167. Anderson, P.M., *Nitrogen Metabolism and Excretion*, ed. PJ Walsh, P Wright. NewYork: CRC Press, 1995: p. 33-49.
168. Jones, M.E., *Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis.* Annu Rev Biochem, 1980. **49**: p. 253-79.
169. Meister, A., *Adv. Enzymol. Relat. Areas. Mol. Biol.*, 1989. **62**: p. 315-746.
170. Thoden, J.B., et al., *Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product.* Biochemistry, 1997. **36**(21): p. 6305-16.
171. Thoden, J.B., et al., *The structure of carbamoyl phosphate synthetase determined to 2.1 Å resolution.* Acta crystallographica. Section D, Biological crystallography, 1999. **55**(Pt 1): p. 8-24.
172. Raushel, F.M., J.B. Thoden, and H.M. Holden, *The Amidotransferase Family of Enzymes: Molecular Machines for the Production and Delivery of Ammonia.* Biochemistry, 1999. **38**(25): p. 7891-7899.
173. Holden, H.M., J.B. Thoden, and F.M. Raushel, *Carbamoyl phosphate synthetase: a tunnel runs through it.* Current Opinion in Structural Biology, 1998. **8**(6): p. 679-685.
174. Holden, H.M., J.B. Thoden, and F.M. Raushel, *Carbamoyl phosphate synthetase: an amazing biochemical odyssey from substrate to product.* Cellular and Molecular Life Sciences CMLS, 1999. **56**(5): p. 507-522.
175. Tesmer, J.J.G., et al., *The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families.* Nature Structural Biology, 1996. **3**(1): p. 74-86.
176. Larsen, T.M., et al., *Three-dimensional structure of Escherichia coli asparagine synthetase B: a short journey from substrate to product.* Biochemistry, 1999. **38**(49): p. 16146-57.
177. Kim, J.H., et al., *Structure and function of the glutamine phosphoribosylpyrophosphate amidotransferase glutamine site and communication with the phosphoribosylpyrophosphate site.* J Biol Chem, 1996. **271**(26): p. 15549-57.
178. Krahn, J.M., et al., *Coupled formation of an amidotransferase interdomain ammonia channel and a phosphoribosyltransferase active site.* Biochemistry, 1997. **36**(37): p. 11061-8.
179. O'Neil, R.H., et al., *Phylogenetic classification of protozoa based on the structure of the linker domain in the bifunctional enzyme, dihydrofolate reductase-thymidylate synthase.* J Biol Chem, 2003. **278**(52): p. 52980-7.
180. Hirst, M., et al., *Human GMP synthetase. Protein purification, cloning, and functional expression of cDNA.* J Biol Chem, 1994. **269**(38): p. 23830-7.
181. Nakamura, J., et al., *The glutamine hydrolysis function of human GMP synthetase. Identification of an essential active site cysteine.* J Biol Chem, 1995. **270**(40): p. 23450-5.
182. Zalkin, H. and J.L. Smith, *Enzymes utilizing glutamine as an amide donor.* Adv Enzymol Relat Areas Mol Biol, 1998. **72**: p. 87-144.
183. Kohls, D., et al., *The crystal structure of the formiminotransferase domain of formiminotransferase-cyclodeaminase: implications for substrate channeling in a bifunctional enzyme.* Structure, 2000. **8**(1): p. 35-46.

184. Lapierre, P., et al., *Formiminotransferase cyclodeaminase is an organ-specific autoantigen recognized by sera of patients with autoimmune hepatitis*. Gastroenterology, 1999. **116**(3): p. 643-9.
185. Liberal, R., et al., *Diagnostic criteria of autoimmune hepatitis*. Autoimmunity Reviews, 2014. **13**(4): p. 435-440.
186. Benkovic, S.J. and R.L. Blakley, *Folates and Pterins, Chemistry and Biochemistry of Folates*. Vol. 1. 1984.
187. Paquin, J., C.M. Baugh, and R.E. MacKenzie, *Channeling between the active sites of formiminotransferase-cyclodeaminase. Binding and kinetic studies*. J Biol Chem, 1985. **260**(28): p. 14925-31.
188. Appling, D.R., *Compartmentation of folate-mediated one-carbon metabolism in eukaryotes*. Faseb j, 1991. **5**(12): p. 2645-51.
189. Zalkin, H., *The amidotransferases*. Adv Enzymol Relat Areas Mol Biol, 1993. **66**: p. 203-309.
190. Hutson, R.G., et al., *Amino acid control of asparagine synthetase: relation to asparaginase resistance in human leukemia cells*. Am J Physiol, 1997. **272**(5 Pt 1): p. C1691-9.
191. Balasubramanian, M.N., E.A. Butterworth, and M.S. Kilberg, *Asparagine synthetase: regulation by cell stress and involvement in tumor biology*. American journal of physiology. Endocrinology and metabolism, 2013. **304**(8): p. E789-E799.
192. Richards, N.G. and M.S. Kilberg, *Asparagine synthetase chemotherapy*. Annu Rev Biochem, 2006. **75**: p. 629-54.
193. Gaufichon, L., et al., *Biological functions of asparagine synthetase in plants*. Plant Science, 2010. **179**(3): p. 141-153.
194. Senecoff, J.F., E.C. McKinney, and R.B. Meagher, *De novo purine synthesis in Arabidopsis thaliana. II. The PUR7 gene encoding 5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole synthetase is expressed in rapidly dividing tissues*. Plant Physiol, 1996. **112**(3): p. 905-17.
195. Herz, S., S. Eberhardt, and A. Bacher, *Biosynthesis of riboflavin in plants. The ribA gene of Arabidopsis thaliana specifies a bifunctional GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase*. Phytochemistry, 2000. **53**(7): p. 723-31.
196. Moffatt, B.A. and H. Ashihara, *Purine and pyrimidine nucleotide synthesis and metabolism*. Arabidopsis Book, 2002. **1**: p. e0018.
197. Hung, W.F., et al., *Characterization of Arabidopsis glutamine phosphoribosyl pyrophosphate amidotransferase-deficient mutants*. Plant Physiol, 2004. **135**(3): p. 1314-23.
198. van der Graaff, E., et al., *Molecular analysis of "de novo" purine biosynthesis in solanaceous species and in Arabidopsis thaliana*. Front Biosci, 2004. **9**: p. 1803-16.
199. Smith, J.L., *Glutamine PRPP amidotransferase: snapshots of an enzyme in action*. Curr Opin Struct Biol, 1998. **8**(6): p. 686-94.
200. Smith, J.L., et al., *Structure of the allosteric regulatory enzyme of purine biosynthesis*. Science, 1994. **264**(5164): p. 1427-33.
201. Walsh, T.A., et al., *Chemical genetic identification of glutamine phosphoribosylpyrophosphate amidotransferase as the target for a novel bleaching herbicide in Arabidopsis*. Plant Physiol, 2007. **144**(3): p. 1292-304.
202. Rudolph, J. and J. Stubbe, *Investigation of the Mechanism of Phosphoribosylamine Transfer from Glutamine Phosphoribosylpyrophosphate*

- Amidotransferase to Glycinamide Ribonucleotide Synthetase. Biochemistry*, 1995. **34**(7): p. 2241-2250.
203. Brannigan, J.A., et al., *A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. Nature*, 1995. **378**(6555): p. 416-9.
204. Smith, J.L., *Structures of glutamine amidotransferases from the purine biosynthetic pathway. Biochem Soc Trans*, 1995. **23**(4): p. 894-8.
205. Nelson, K., et al., *Sequence of the dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene of Leishmania amazonensis. Nucleic Acids Res*, 1990. **18**(9): p. 2819.
206. Ackermann, W.W. and V.R. Potter, *Enzyme inhibition in relation to chemotherapy. Proc Soc Exp Biol Med*, 1949. **72**(1): p. 1-9.
207. Ferone, R. and S. Roland, *Dihydrofolate reductase: thymidylate synthase, a bifunctional polypeptide from Crithidia fasciculata. Proc Natl Acad Sci U S A*, 1980. **77**(10): p. 5802-6.
208. Ivanetich, K.M. and D.V. Santi, *Bifunctional thymidylate synthase-dihydrofolate reductase in protozoa. Faseb j*, 1990. **4**(6): p. 1591-7.
209. Cella, R., et al., *Proteolytic and partial sequencing studies of the bifunctional dihydrofolate reductase-thymidylate synthase from Daucus carota. Plant Mol Biol*, 1991. **16**(6): p. 975-82.
210. Lazar, G., H. Zhang, and H.M. Goodman, *The origin of the bifunctional dihydrofolate reductase-thymidylate synthase isogenes of Arabidopsis thaliana. Plant J*, 1993. **3**(5): p. 657-68.
211. Meek, T.D., E.P. Garvey, and D.V. Santi, *Purification and characterization of the bifunctional thymidylate synthetase-dihydrofolate reductase from methotrexate-resistant Leishmania tropica. Biochemistry*, 1985. **24**(3): p. 678-86.
212. Liang, P.H. and K.S. Anderson, *Substrate channeling and domain-domain interactions in bifunctional thymidylate synthase-dihydrofolate reductase. Biochemistry*, 1998. **37**(35): p. 12195-205.
213. Trujillo, M., et al., *Heterologous expression and characterization of the bifunctional dihydrofolate reductase-thymidylate synthase enzyme of Toxoplasma gondii. Biochemistry*, 1996. **35**(20): p. 6366-74.
214. Knighton, D.R., et al., *Structure of and kinetic channelling in bifunctional dihydrofolate reductase-thymidylate synthase. Nat Struct Biol*, 1994. **1**(3): p. 186-94.
215. Stroud, R.M., *An electrostatic highway. Nat Struct Biol*, 1994. **1**(3): p. 131-4.
216. Trujillo, M., R. Duncan, and D.V. Santi, *Construction of a homodimeric dihydrofolate reductase-thymidylate synthase bifunctional enzyme. Protein Eng*, 1997. **10**(5): p. 567-73.
217. Reed, L.J., *Multienzyme complexes. Accounts of Chemical Research*, 1974. **7**(2): p. 40-46.
218. Perham, R.N., *Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein. Biochemistry*, 1991. **30**(35): p. 8501-12.
219. Izard, T., et al., *Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. Proc Natl Acad Sci U S A*, 1999. **96**(4): p. 1240-5.
220. Berg, A., et al., *Kinetics and specificity of reductive acylation of wild-type and mutated lipoyl domains of 2-oxo-acid dehydrogenase complexes from Azotobacter vinelandii. Eur J Biochem*, 1998. **252**(1): p. 45-50.



221. Berg, A. and A. de Kok, *2-Oxo acid dehydrogenase multienzyme complexes. The central role of the lipoyl domain*. Biol Chem, 1997. **378**(7): p. 617-34.
222. Wallis, N.G., et al., *Recognition of a surface loop of the lipoyl domain underlies substrate channelling in the pyruvate dehydrogenase multienzyme complex*. J Mol Biol, 1996. **263**(3): p. 463-74.
223. Aevarsson, A., et al., *Crystal structure of 2-oxoisovalerate and dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes*. Nat Struct Biol, 1999. **6**(8): p. 785-92.
224. Kis, K. and A. Bacher, *Substrate channeling in the lumazine synthase/riboflavin synthase complex of Bacillus subtilis*. J Biol Chem, 1995. **270**(28): p. 16788-95.
225. Ladenstein, R., et al., *Heavy riboflavin synthase from Bacillus subtilis. Crystal structure analysis of the icosahedral beta 60 capsid at 3.3 Å resolution*. J Mol Biol, 1988. **203**(4): p. 1045-70.
226. Ladenstein, R., et al., *The lumazine synthase/riboflavin synthase complex of Bacillus subtilis. X-ray structure analysis of hollow reconstituted beta-subunit capsids*. Eur J Biochem, 1994. **223**(3): p. 1007-17.
227. Ritsert, K., et al., *Studies on the lumazine synthase/riboflavin synthase complex of Bacillus subtilis: crystal structure analysis of reconstituted, icosahedral beta-subunit capsids with bound substrate analogue inhibitor at 2.4 Å resolution*. J Mol Biol, 1995. **253**(1): p. 151-67.
228. Bacher, A., et al., *Biosynthesis of riboflavin: lumazine synthase and riboflavin synthase*. Methods Enzymol, 1997. **280**: p. 389-99.
229. Neuberger, G. and A. Bacher, *Biosynthesis of riboflavin. An aliphatic intermediate in the formation of 6,7-dimethyl-8-ribityllumazine from pentose phosphate*. Biochemical and Biophysical Research Communications, 1985. **127**(1): p. 175-181.
230. Volk, R. and A. Bacher, *Studies on the 4-carbon precursor in the biosynthesis of riboflavin. Purification and properties of L-3,4-dihydroxy-2-butanone-4-phosphate synthase*. J Biol Chem, 1990. **265**(32): p. 19479-85.
231. Kis, K., R. Volk, and A. Bacher, *Biosynthesis of riboflavin. Studies on the reaction mechanism of 6,7-dimethyl-8-ribityllumazine synthase*. Biochemistry, 1995. **34**(9): p. 2883-92.
232. Kasaragod, P., et al., *Crystal Structure of Liganded Rat Peroxisomal Multifunctional Enzyme Type 1: A FLEXIBLE MOLECULE WITH TWO INTERCONNECTED ACTIVE SITES*. J. Biol. Chem., 2010. **285**(31): p. 24089-24098.
233. Kasaragod, P., et al., *Structural enzymology comparisons of multifunctional enzyme, type-1 (MFE1): the flexibility of its dehydrogenase part*. FEBS Open Bio, 2017. **7**(12): p. 1830-1842.
234. Tsuchiya, D., et al., *Ligand-induced domain rearrangement of fatty acid beta-oxidation multienzyme complex*. Structure, 2006. **14**(2): p. 237-46.
235. Wheeldon, I., et al., *Substrate channelling as an approach to cascade reactions*. Nat Chem, 2016. **8**(4): p. 299-309.
236. Alder, B.J. and T.E. Wainwright, *Studies in Molecular Dynamics. I. General Method*. J. Chem. Phys., 1959. **31**(2): p. 459.
237. van der Ploeg, P. and H.J.C. Berendsen, *Molecular dynamics simulation of a bilayer membrane*. Journal of Chemical Physics, 1982. **76**(6): p. 3271-3276.
238. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins*. Nature, 1977. **267**: p. 585-590.

- 
239. Roccatano, D., A. Barthel, and M. Zacharias, *Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation*. Biopolymers, 2007. **85**(5-6): p. 407-421.
240. Sharma, S., F. Ding, and N.V. Dokholyan, *Multiscale modeling of nucleosome dynamics*. Biophys J., 2007. **92**(5): p. 1457-1470.
241. Tinoco, I.J. and J.D. Wen, *Simulation and analysis of single-ribosome translation*. Phys Biol., 2009. **6**(2): p. 025006.
242. Brandman, R., Y. Brandman, and V.S. Pande, *A-site residues move independently from P-site residues in all-atom molecular dynamics simulations of the 70S bacterial ribosome*. PLoS One, 2012. **7**(1): p. e29377.
243. van Gunsteren, W.F., et al., *A molecular dynamics computer simulation of an eight-base-pair DNA fragment in aqueous solution: comparison with experimental two-dimensional NMR data*. Ann N Y Acad Sci, 1986. **482**: p. 287-303.
244. Duan, Y. and P.A. Kollman, *Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution*. Science, 1998. **282**: p. 740-4.
245. Jayachandran, G., V. Vishal, and V.S. Pande, *Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece*. J Chem Phys, 2006. **124**: p. 194,902.
246. Freddolino, P.L., et al., *Molecular dynamics simulations of the complete satellite tobacco mosaic virus*. Structure, 2006. **14**: p. 437-449.
247. Brooks, B., et al., *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations*. J. Comp. Chem., 1983. **4**: p. 187-217.
248. Weiner, P.K. and P.A. Kollman, *Assisted model building with energy refinement. A general program for modeling molecules and their interactions*. J. Comput. Chem., 1981. **2**: p. 287-303.
249. Nelson, M., et al., *NAMD—A parallel, object-oriented molecular dynamics program*. International Journal of Supercomputer Applications and High Performance Computing, 1996. **10**: p. 251-268.
250. Berendsen, H., D. van der Spoel, and R. van Drunen, *GROMACS: A messagepassing parallel molecular dynamics implementation*. Comp. Phys., 1995. **91**: p. 43-56.
251. Diehl, M. and T. Fischer, *Skript zur Mathematik fuer die Molekulare Biotechnologie*. 2003, Universitaet Heidelberg.
252. Leach, A.R., *Molecular Modelling*. 2001.
253. McQuaerrie, D.A., *Statistical Mechanics*, University Science Books. 2000.
254. Berendsen, H.J.C., et al., *Intermolecular Forces*. Reidel, Dodrecht, 1981.
255. Nosé, S., *A Unified Formulation of the Constant Temperature Molecular Dynamics Methods*. J. Chem. Phys., 1984. **81**: p. 511-519.
256. Hoover, W.G., *Canonical dynamics: Equilibrium phase-space distributions*. Phys. Rev. A., 1985. **31**: p. 1695-1697.
257. Andersen, H.C., *Molecular dynamics simulations at constant pressure and/or temperature*. J. Chem. Phys., 1980. **72**: p. 2384-2393.
258. Ruslan, L. and D. Chack, *Langevin thermostat for rigid body dynamics*. J. Chem. Phys., 2009. **130**: p. 234101.
259. Born, M. and R. Oppenheimer, *Zur Quantentheorie der Molekeln*, Annalen der Physik. Annalen der Physik, 1927. **84**: p. 457.
260. Vollhardt, K.P.C. and N.E. Schore, *Organic Chemistry*, Freeman. 1999: p. Chap.: 2.5-2.7.
-

- 
261. Sugitaand, Y. and Y. Okamoto, *Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test*. Chem. Phys. Lett., 1999. **314**: p. 141.
262. Wu, X. and S. Wang, *Enhancing systematic motion in molecular dynamics simulation*. J. Chem. Phys., 1999. **110**: p. 9401.
263. Torrie, G.M. and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*. J. Comput. Phys., 1977. **23**: p. 187.
264. Voter, A.F., *Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events*. Phys. Rev. Lett., 1997. **78**: p. 3908.
265. Voter, A.F., *A method for accelerating the molecular dynamics simulation of infrequent events*. J. Chem. Phys., 1997. **106**: p. 4665.
266. Steiner, M.M. and P.A. Genilloud, *Simple bias potential for boosting molecular dynamics with the hyperdynamics scheme*. Phys. Rev. B, 1998. **57**: p. 10236.
267. Rahman, J.A. and J.C. Tully, *Puddle-skimming: An efficient sampling of multidimensional configuration space*. J. Chem. Phys., 2002. **116**: p. 8750.
268. Hamelberg, D., J. Mongan, and J.A. McCammon, *Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules*. J. Chem. Phys., 2004. **120**: p. 11919.
269. Pal, S. and K.A. Fichthorn, *Accelerated molecular dynamics of infrequent events*. Chem. Engineering J., 1999. **74**: p. 77-83.
270. Gong, X.G. and J.W. Wilkins, *Hyper molecular dynamics with a local bias potential*. Phys. Rev. B, 1999. **59**: p. 54.
271. Henzler-Wildman, K.A. and D. Kern, *Dynamic personalities of proteins*. Nature, 2007. **450**(7172): p. 964-972.
272. Tama, F., *Normal mode analysis with simplified models to investigate the global dynamics of biological systems*. Protein Peptide Lett., 2003. **10**(2): p. 119-132.
273. Ma, J.P., *Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes*. Structure, 2005. **13**(3): p. 373-380.
274. Bahar, I. and A.J. Rader, *Coarse-grained normal mode analysis in structural biology*. Curr. Opin. Struc. Biol., 2005. **15**(5): p. 586-592.
275. Bahar, I., et al., *Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins*. Chem. Rev., 2010. **110**(3): p. 1463-1497.
276. Krebs, W.G., et al., *Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic*. Proteins, 2002. **48**(4): p. 682-695.
277. Petrone, P.P. and S. Vijay, *Can conformational change be described by only a few normal modes?* Biophys. J., 2006. **90**(5): p. 1583-1593.
278. Dobbins, S.E., V.I. Lesk, and M.J.E. Sternberg, *Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking*. Proc. Natl. Acad. Sci. USA, 2008. **105**(30): p. 10390-10395.
279. Wako, H. and S. Endo, *Ligand-induced conformational change of a protein reproduced by a linear combination of displacement vectors obtained from normal mode analysis*. Biophys. Chem., 2011. **159**(2-3): p. 257-266.
280. Skjaerven, L., A. Martinez, and N. Reuter, *Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit*. Proteins, 2010. **79**(1): p. 232-243.
-

281. Rueda, M., P. Chacón, and M. Orzoco, *Thorough validation of protein normal mode analysis: a comparative study with essential dynamics*. *Structure*, 2007. **15**(5): p. 565-575.
282. Fuglebakk, E., N. Reuter, and K. Hinsen, *Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions*. *J. Chem. Theory Comput.*, 2013: p. 5618-5628.
283. Romo, T.D. and A. Grossfield, *Validating and improving elastic network models with molecular dynamics simulations*. *Proteins*, 2011. **79**(1): p. 23-34.
284. Ahmed, A., S. Villinger, and H. Gohlke, *Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses*. *Proteins*, 2010. **78**(16): p. 3341-3352.
285. Dykeman, E.C. and O.F. Sankey, *Normal mode analysis and applications in biological physics*. *J Phys Condens Matter*, 2010. **22**(42): p. 423202.
286. Tiwari, S., et al., *WEBnm@ v2.0: Web server and services for comparing protein flexibility*. *BMC Bioinformatics*, 2014. **15**: p. 427.
287. Hinsen, K., et al., *Harmonicity in slow protein dynamics*. *Chem. Phys.*, 2000. **261**: p. 25-37.
288. Maguid, S., et al., *Exploring the common dynamics of homologous proteins. Application to the globin family*. *Biophys. J.*, 2005. **89**(1): p. 3-13.
289. Leo-Macias, A., et al., *An analysis of core deformations in protein superfamilies*. *Biophys. J.*, 2005. **88**(2): p. 1291-1299.
290. Leo-Macias, A., et al., *Core deformations in protein families: a physical perspective*. *Biophys. Chem.*, 2005. **115**(2-3): p. 125-128.
291. Zen, A., et al., *Correspondences between low-energy modes in enzymes: Dynamics-based alignment of enzymatic functional families*. *Protein Sci.*, 2008. **17**(5): p. 918-929.
292. Zen, A., et al., *Using dynamics-based comparisons to predict nucleic acid binding sites in proteins: an application to OB-fold domains*. *Bioinformatics*, 2009. **25**(15): p. 1876-1883.
293. Fuglebakk, E., J. Echave, and N. Reuter, *Measuring and comparing structural fluctuation patterns in large protein datasets*. *Bioinformatics*, 2012. **28**(19): p. 2431-2440.
294. Zheng, W. and S. Doniach, *A comparative study of motor-protein motions by using a simple elastic-network model*. *Proc. Natl. Acad. Sci. USA*, 2003. **100**(23): p. 13253-13258.
295. Zheng, W. and B.R. Brooks, *Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin*. *Biophys. J.*, 2005. **89**: p. 167-178.
296. Hollup, S.M., et al., *Exploring the factors determining the dynamics of different protein folds*. *Protein Sci.*, 2011. **20**(1): p. 197-209.
297. Rodgers, T.L., et al., *Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors*. *PLoS Biol.*, 2013. **11**(9): p. e1001651.
298. Marcos, E., R. Crehuet, and I. Bahar, *Changes in dynamics upon oligomerization regulate substrate binding and allostery in amino acid kinase family members*. *PLoS Comput. Biol.*, 2011. **7**(9): p. e1002201.
299. Katebi, A.R. and R.L. Jernigan, *The critical role of the loops of triosephosphate isomerase for its oligomerization, dynamics, and functionality*. *Protein Sci.*, 2014. **23**(2): p. 213-228.

300. Kolan, D., G. Fonar, and A.O. Samson, *Elastic network normal mode dynamics reveal the GPCR activation mechanism*. Proteins, 2014. **82**(4): p. 579-586.
301. Lukman, S. and G.H. Grant, *A network of dynamically conserved residues deciphers the motions of maltose transporter*. Proteins, 2009. **76**(3): p. 588-597.
302. Seckler, J.M., et al., *The interplay of structure and dynamics: insights from a survey of HIV-1 reverse transcriptase crystal structures*. Protein, 2013. **81**(10): p. 1792-1801.
303. Yang, L., et al., *Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes*. Structure, 2008. **16**(2): p. 321-330.
304. van Wynsberghe, A.W. and Q. Cui, *Conservation and variation of structural flexibility in protein families*. Structure, 2010. **18**(3): p. 281-283.
305. Micheletti, C., *Comparing proteins by their internal dynamics: Exploring structure–function relationships beyond static structural alignments*. Phys. Life Rev., 2012. **10**: p. 1-26.
306. <http://enm.lobos.nih.gov>. [cited 2020 26.08.2020].
307. Seo, S. and M.K. Kim, *KOSMOS: a universal morph server for nucleic acids, proteins and their complexes*. Nucleic Acids Res., 2012. **40**: p. W531-536.
308. Kruger, D.M., A. Ahmed, and H. Gohlke, *NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins*. Nucleic Acids Res., 2012. **40**: p. W310-316.
309. Suhre, K. and Y. Sanejouand, *ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement*. Nucleic Acid Res., 2004. **32**: p. 610-614.
310. Lindahl, E., et al., *NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis*. Nucleic Acids Res., 2006. **34**: p. W52-56.
311. Eyal, E., L.W. Yang, and I. Bahar, *Anisotropic network model: systematic evaluation and a new web interface*. Bioinformatics, 2006. **22**: p. 2619-2627.
312. Hollup, S.M., G. Salensminde, and N. Reuter, *WEBnm@: a web application for normal mode analyses of proteins*. BMC Bioinformatics, 2005. **6**: p. 52.
313. Huber, R. and W.S. Bennett, Jr., *Functional significance of flexibility in proteins*. Biopolymers, 1983. **22**: p. 261-279.
314. McCammon, J.A. and S.C. Harvey, *Dynamics of proteins and nucleic acids*. Biochem. Educ. Cambridge University Press, 1987. **17**: p. 220-220 pp 234.
315. Hünenberger, P.H., A.E. Mark, and W.F. van Gunsteren, *Fluctuation and Cross-correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations*. Journal of Molecular Biology, 1995. **252**(4): p. 492-503.
316. Kasahara, K., I. Fukuda, and H. Nakamura, *A Novel Approach of Dynamic Cross Correlation Analysis on Molecular Dynamics Simulations and Its Application to Ets1 Dimer–DNA Complex*. PLOS ONE, 2014. **9**(11): p. e112419.
317. McCammon, J.A., *Protein dynamics*. Reports on Progress in Physics, 1984. **47**: p. 1.
318. Cattell, R.B., *The Scree Test For The Number Of Factors*. Multivariate Behavioral Research, 1966. **1**(2): p. 245-276.
319. Kieseritzky, G. and E.W. Knapp, *Optimizing pK<sub>a</sub> computation in proteins with pH adapted conformations*. Proteins Struct. Funct. Genet., 2008. **71**: p. 1335-1348.

- 
320. Meyer, T. and E.W. Knapp, *pKa Values in Proteins Determined by Electrostatics Applied to Molecular Dynamics Trajectories*. J. Chem. Theory Comput., 2015. **11**: p. 2827-2840.
321. Rabenstein, B. and E.W. Knapp, *Calculated pH-Dependent Population and Protonation of Carbon-Monoxo-Myoglobin Conformers*. Biophys. J., 2001. **80**: p. 1141-1150.
322. You, T.J. and D. Bashford, *Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility*. Biophys. J., 1995. **69**: p. 1721-1733.
323. Ullmann, G.M. and E.W. Knapp, *Electrostatic models for computing protonation and redox equilibria in proteins*, Eur. Biophys. Biophys. J., 1999. **28**: p. 533-551.
324. Bashford, D. and M. Karplus, *pKa's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model*. Biochemistry, 1990. **29**: p. 10219-10225.
325. Berman, H.M., *The Protein Data Bank*. Nucleic Acids Res., 2000. **28**: p. 235-242.
326. Lopes, P.E., O. Guvench, and A.D. MacKerell, Jr., *Current status of protein force fields for molecular dynamics simulations*. Methods Mol Biol, 2015. **1215**: p. 47-71.
327. Dastmalchi, S., M. Hamzeh-Mivehroud, and B. Sokouti, *Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery*. 2016, IGI Global AMTCP: Medical Information Science Reference (an imprint of IGI global).
328. Saldivar-Gonzalez, F., F. Prieto-Martínez, and J. Medina-Franco, *Descubrimiento y desarrollo de fármacos: un enfoque computacional*. Educación Química, 2016. **28**.
329. Houston, D.R. and M.D. Walkinshaw, *Consensus docking: improving the reliability of docking in a virtual screening context*. J Chem Inf Model, 2013. **53**(2): p. 384-90.
330. Tuccinardi, T., et al., *Extensive consensus docking evaluation for ligand pose prediction and virtual screening studies*. J Chem Inf Model, 2014. **54**(10): p. 2980-6.
331. Halperin, I., et al., *Principles of docking: An overview of search algorithms and a guide to scoring functions*. Proteins, 2002. **47**(4): p. 409-43.
332. Agarwal, S. and R. Mehrotra, *Mini Review\_ An overview of Molecular Docking*. JSM Chemistry, 2016. **2**: p. 1024.
333. Ferreira, L.G., et al., *Molecular docking and structure-based drug design strategies*. Molecules, 2015. **20**(7): p. 13384-421.
334. Warren, G.L., et al., *Essential considerations for using protein-ligand structures in drug discovery*. Drug Discov Today, 2012. **17**(23-24): p. 1270-81.
335. Feher, M. and C.I. Williams, *Numerical errors and chaotic behavior in docking simulations*. J Chem Inf Model, 2012. **52**(3): p. 724-38.
336. Stewart, J.J., *Application of the PM6 method to modeling the solid state*. J Mol Model, 2008. **14**(6): p. 499-535.
337. Wada, M. and M. Sakurai, *A quantum chemical method for rapid optimization of protein structures*. J Comput Chem, 2005. **26**(2): p. 160-8.
338. Hostaš, J., J. Řezáč, and P. Hobza, *On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions*. Chemical Physics Letters, 2013. **568-569**: p. 161-166.
339. Nikitina, E., et al., *Semiempirical Calculations of Binding Enthalpy for Protein-Ligand Complexes*. International Journal of Quantum Chemistry, 2004. **97**: p. 747-763.
340. Ohno, K., et al., *Application of an integrated MOZYME+DFT method to pKa calculations for proteins*. Chemical Physics Letters, 2001. **341**(3): p. 387-392.
-

- 
341. Marzaro, G., et al., *Psoralen derivatives as inhibitors of NF- $\kappa$ B/DNA interaction: synthesis, molecular modeling, 3D-QSAR, and biological evaluation*. J Med Chem, 2013. **56**(5): p. 1830-42.
342. Espinoza-Fonseca, L.M. and J.G. Trujillo-Ferrara, *Identification of multiple allosteric sites on the M1 muscarinic acetylcholine receptor*. FEBS Lett, 2005. **579**(30): p. 6726-32.
343. Goodford, P.J., *A computational procedure for determining energetically favorable binding sites on biologically important macromolecules*. J Med Chem, 1985. **28**(7): p. 849-57.
344. Feinstein, W.P. and M. Brylinski, *Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets*. J Cheminform, 2015. **7**: p. 18.
345. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*. J Comput Chem, 2010. **31**(2): p. 455-61.
346. Brint, A.T. and P. Willett, *Algorithms for the identification of three-dimensional maximal common substructures*. Journal of Chemical Information and Computer Sciences, 1987. **27**(4): p. 152-158.
347. Fischer, D., et al., *Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition*. Proteins, 1993. **16**(3): p. 278-92.
348. Norel, R., et al., *Molecular surface recognition by a computer vision-based technique*. Protein Eng, 1994. **7**(1): p. 39-46.
349. Rarey, M., et al., *A fast flexible docking method using an incremental construction algorithm*. J Mol Biol, 1996. **261**(3): p. 470-89.
350. DesJarlais, R.L., et al., *Docking flexible ligands to macromolecular receptors by molecular shape*. J Med Chem, 1986. **29**(11): p. 2149-53.
351. Leach, A.R. and I.D. Kuntz, *Conformational analysis of flexible ligands in macromolecular receptor sites*. Journal of Computational Chemistry, 1992. **13**(6): p. 730-748.
352. Ewing, T.J., et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. J Comput Aided Mol Des, 2001. **15**(5): p. 411-28.
353. Schnecke, V. and L. Kuhn, *Virtual screening with solvation and ligand-induced complementarity*. 2007. p. 171-190.
354. Goodsell, D.S., et al., *Automated docking in crystallography: analysis of the substrates of aconitase*. Proteins, 1993. **17**(1): p. 1-10.
355. Hart, T.N. and R.J. Read, *A multiple-start Monte Carlo docking method*. Proteins, 1992. **13**(3): p. 206-22.
356. Goodsell, D.S. and A.J. Olson, *Automated docking of substrates to proteins by simulated annealing*. Proteins, 1990. **8**(3): p. 195-202.
357. McMartin, C. and R.S. Bohacek, *QXP: powerful, rapid computer algorithms for structure-based drug design*. J Comput Aided Mol Des, 1997. **11**(4): p. 333-44.
358. Accelrys Inc., S.D., CA, USA., *Affinity*. Accelrys Inc., San Diego, CA, USA.
359. Kollman, P., *Free energy calculations: Applications to chemical and biochemical phenomena*. Chemical Reviews, 1993. **93**(7): p. 2395-2417.
360. Aqvist, J., V.B. Luzhkov, and B.O. Brandsdal, *Ligand binding affinities from MD simulations*. Acc Chem Res, 2002. **35**(6): p. 358-65.
-

- 
361. Carlson, H.A. and W.L. Jorgensen, *An Extended Linear Response Method for Determining Free Energies of Hydration*. The Journal of Physical Chemistry, 1995. **99**(26): p. 10667-10673.
362. Srinivasan, J., et al., *Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices*. Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.
363. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models*. Acc Chem Res, 2000. **33**(12): p. 889-97.
364. Still, W.C., et al., *Semianalytical treatment of solvation for molecular mechanics and dynamics*. Journal of the American Chemical Society, 1990. **112**(16): p. 6127-6129.
365. Guimarães, C.R. and A.M. Mathiowetz, *Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations*. J Chem Inf Model, 2010. **50**(4): p. 547-59.
366. Singh, N. and A. Warshel, *Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions*. Proteins, 2010. **78**(7): p. 1705-23.
367. Zhang, X., H. Perez-Sanchez, and F.C. Lightstone, *A Comprehensive Docking and MM/GBSA Rescoring Study of Ligand Recognition upon Binding Antithrombin*. Curr Top Med Chem, 2017. **17**(14): p. 1631-1639.
368. Böhm, H.J., *Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs*. J Comput Aided Mol Des, 1998. **12**(4): p. 309-23.
369. Gehlhaar, D.K., et al., *Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming*. Chem Biol, 1995. **2**(5): p. 317-24.
370. Verkhivker, G.M., et al., *Deciphering common failures in molecular docking of ligand-protein complexes*. J Comput Aided Mol Des, 2000. **14**(8): p. 731-51.
371. Jain, A.N., *Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities*. J Comput Aided Mol Des, 1996. **10**(5): p. 427-40.
372. Head, R.D., et al., *VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands*. Journal of the American Chemical Society, 1996. **118**(16): p. 3959-3969.
373. Eldridge, M.D., et al., *Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes*. J Comput Aided Mol Des, 1997. **11**(5): p. 425-45.
374. Pason, L.P. and C.A. Sotriffer, *Empirical Scoring Functions for Affinity Prediction of Protein-ligand Complexes*. Mol Inform, 2016. **35**(11-12): p. 541-548.
375. Exner, T., O. Korb, and T. Brink, *New and improved features of the docking software PLANTS*. Vol. 3. 2009.
376. Meng, X.Y., et al., *Molecular docking: a powerful approach for structure-based drug discovery*. Curr Comput Aided Drug Des, 2011. **7**(2): p. 146-57.
377. Murray, C.W., T.R. Auton, and M.D. Eldridge, *Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model*. J Comput Aided Mol Des, 1998. **12**(5): p. 503-19.
-



- 
378. Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: methods and applications*. Nat Rev Drug Discov, 2004. **3**(11): p. 935-49.
379. Huang, S.Y., S.Z. Grinter, and X. Zou, *Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions*. Phys Chem Chem Phys, 2010. **12**(40): p. 12899-908.
380. Muegge, I., *A knowledge-based scoring function for protein-ligand interactions: Probing the reference state*. Perspectives in Drug Discovery and Design, 2000. **20**(1): p. 99-114.
381. Debroise, T., E.I. Shakhnovich, and N. Chéron, *A Hybrid Knowledge-Based and Empirical Scoring Function for Protein-Ligand Interaction: SMOG2016*. J Chem Inf Model, 2017. **57**(3): p. 584-593.
382. Shoichet, B.K., et al., *Lead discovery using molecular docking*. Curr Opin Chem Biol, 2002. **6**(4): p. 439-46.
383. Gschwend, D.A., A.C. Good, and I.D. Kuntz, *Molecular docking towards drug discovery*. J Mol Recognit, 1996. **9**(2): p. 175-86.
384. Alogheli, H., et al., *Docking of Macrocycles: Comparing Rigid and Flexible Docking in Glide*. J Chem Inf Model, 2017. **57**(2): p. 190-202.
385. Guedes, I., C. Magalhães, and L. Dardenne, *Receptor-ligand molecular docking*. Biophysical Reviews, 2014. **6**: p. 75-87.
386. Kuntz, I.D., et al., *A geometric approach to macromolecule-ligand interactions*. J Mol Biol, 1982. **161**(2): p. 269-88.
387. Miller, M.D., et al., *FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure*. J Comput Aided Mol Des, 1994. **8**(2): p. 153-74.
388. Shoichet, B.K., et al., *Structure-based discovery of inhibitors of thymidylate synthase*. Science, 1993. **259**(5100): p. 1445-50.
389. Gabb, H.A., R.M. Jackson, and M.J. Sternberg, *Modelling protein docking using shape complementarity, electrostatics and biochemical information*. J Mol Biol, 1997. **272**(1): p. 106-20.
390. Koshland, D.E., Jr., *CORRELATION OF STRUCTURE AND FUNCTION IN ENZYME ACTION*. Science, 1963. **142**(3599): p. 1533-41.
391. Hammes, G.G., *Multiple conformational changes in enzyme catalysis*. Biochemistry, 2002. **41**(26): p. 8221-8.
392. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function*. Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.
393. Teague, S.J., *Implications of protein flexibility for drug discovery*. Nat Rev Drug Discov, 2003. **2**(7): p. 527-41.
394. Friesner, R.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*. J Med Chem, 2004. **47**(7): p. 1739-49.
395. Halgren, T.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening*. J Med Chem, 2004. **47**(7): p. 1750-9.
396. Friesner, R.A., et al., *Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes*. J Med Chem, 2006. **49**(21): p. 6177-96.
397. Zhu, K., et al., *Docking covalent inhibitors: a parameter free approach to pose prediction and scoring*. J Chem Inf Model, 2014. **54**(7): p. 1932-40.
-

- 
398. Toledo Warshaviak, D., et al., *Structure-Based Virtual Screening Approach for Discovery of Covalently Bound Ligands*. *Journal of Chemical Information and Modeling*, 2014. **54**(7): p. 1941-1950.
399. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. *Adv Drug Deliv Rev*, 2001. **46**(1-3): p. 3-26.
400. Leach, A.R., et al., *Three-dimensional pharmacophore methods in drug discovery*. *J Med Chem*, 2010. **53**(2): p. 539-58.
401. McInnes, C., *Virtual screening strategies in drug discovery*. *Curr Opin Chem Biol*, 2007. **11**(5): p. 494-502.
402. Hert, J., et al., *Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures*. *Org Biomol Chem*, 2004. **2**(22): p. 3256-66.
403. Sun, H., *Pharmacophore-based virtual screening*. *Curr Med Chem*, 2008. **15**(10): p. 1018-24.
404. Hawkins, P.C., A.G. Skillman, and A. Nicholls, *Comparison of shape-matching and docking as virtual screening tools*. *J Med Chem*, 2007. **50**(1): p. 74-82.
405. Laggner, C., et al., *Discovery of high-affinity ligands of sigma1 receptor, ERG2, and emopamil binding protein by pharmacophore modeling and virtual screening*. *J Med Chem*, 2005. **48**(15): p. 4754-64.
406. Wolber, G. and T. Langer, *LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters*. *J Chem Inf Model*, 2005. **45**(1): p. 160-9.
407. Baroni, M., et al., *A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application*. *J Chem Inf Model*, 2007. **47**(2): p. 279-94.
408. Sciabola, S., et al., *High-throughput virtual screening of proteins using GRID molecular interaction fields*. *J Chem Inf Model*, 2010. **50**(1): p. 155-69.
409. McGaughey, G.B., et al., *Comparison of topological, shape, and docking methods in virtual screening*. *J Chem Inf Model*, 2007. **47**(4): p. 1504-19.
410. Krüger, D.M. and A. Evers, *Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors*. *ChemMedChem*, 2010. **5**(1): p. 148-58.
411. Tan, L., et al., *Integrating structure- and ligand-based virtual screening: comparison of individual, parallel, and fused molecular docking and similarity search calculations on multiple targets*. *ChemMedChem*, 2008. **3**(10): p. 1566-71.
412. Muthas, D., et al., *Is it possible to increase hit rates in structure-based virtual screening by pharmacophore filtering? An investigation of the advantages and pitfalls of post-filtering*. *J Mol Graph Model*, 2008. **26**(8): p. 1237-51.
413. Kontoyianni, M., et al., *Theoretical and practical considerations in virtual screening: a beaten field?* *Curr Med Chem*, 2008. **15**(2): p. 107-16.
414. Calder, P.C., *Functional Roles of Fatty Acids and Their Effects on Human Health*. *JPEN J Parenter Enteral Nutr*, 2015. **39**(1 Suppl): p. 18s-32s.
415. Lazarow, P.B. and C. De Duve, *A fatty acyl-CoA oxidizing system in rat liver peroxisomes; enhancement by clofibrate, a hypolipidemic drug*. *Proc Natl Acad Sci U S A*, 1976. **73**(6): p. 2043-6.
416. Lazarow, P.B., *Rat liver peroxisomes catalyze the beta oxidation of fatty acids*. *J Biol Chem*, 1978. **253**(5): p. 1522-8.
-

417. Kunau, W.H., et al., *Comparative enzymology of beta-oxidation*. Biochem Soc Trans, 1988. **16**(3): p. 418-20.
418. Leenders, F., et al., *The sequence of porcine 80 kDa 17 beta-estradiol dehydrogenase reveals similarities to the short chain alcohol dehydrogenase family, to actin binding motifs and to sterol carrier protein 2*. Mol Cell Endocrinol, 1994. **104**(2): p. 127-31.
419. Novikov, D.K., et al., *Peroxisomal beta-oxidation. Purification of four novel 3-hydroxyacyl-CoA dehydrogenases from rat liver peroxisomes*. J Biol Chem, 1994. **269**(43): p. 27125-35.
420. Dieuaide-Noubhani, M., et al., *Further characterization of the peroxisomal 3-hydroxyacyl-CoA dehydrogenases from rat liver. Relationship between the different dehydrogenases and evidence that fatty acids and the C27 bile acids di- and tri-hydroxycoprostanic acids are metabolized by separate multifunctional proteins*. Eur J Biochem, 1996. **240**(3): p. 660-6.
421. Qin, Y.M., et al., *Peroxisomal multifunctional enzyme of beta-oxidation metabolizing D-3-hydroxyacyl-CoA esters in rat liver: molecular cloning, expression and characterization*. Biochem J, 1997. **321** ( Pt 1)(Pt 1): p. 21-8.
422. Hiltunen, J.K., et al., *Peroxisomal multifunctional beta-oxidation protein of Saccharomyces cerevisiae. Molecular analysis of the fox2 gene and gene product*. J Biol Chem, 1992. **267**(10): p. 6646-53.
423. Kasaragod, P., et al., *The isomerase and hydratase reaction mechanism of the crotonase active site of the multifunctional enzyme (type - 1), as deduced from structures of complexes with 3S - hydroxy - acyl - CoA*. FEBS Journal, 2013. **280**: p. 3160-3175.
424. Kasaragod, P., et al., *Crystal Structure of Liganded Rat Peroxisomal Multifunctional Enzyme Type 1: A FLEXIBLE MOLECULE WITH TWO INTERCONNECTED ACTIVE SITES*. J. Biol. Chem., 2010. **285**(31): p. 24089-24098.
425. Kieseritzky, G. and E.W. Knapp, *Optimizing pK A computation in proteins with pH adapted conformations*. Proteins Struct. Funct. Genet., 2008. **71**: p. 1335-1348.
426. Meyer, T. and E.W. Knapp, *pKa Values in Proteins Determined by Electrostatic Applied to Molecular Dynamics Trajectories*. Journal of Chemical Theory and Computation, 2015. **11**: p. 2827-2840.
427. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. J. Chem. Phys., 1983. **79**: p. 926.
428. Battocchio, G., et al., *Dynamic Properties of the Photosensory Domain of Deinococcus radiodurans Bacteriophytochrome*. J Phys Chem B, 2020. **124**(9): p. 1740-1750.
429. Takiden, A., et al., *Structural and Vibrational Characterization of the Chromophore Binding Site of Bacterial Phytochrome Agp1*. Photochem Photobiol, 2017. **93**(3): p. 713-723.
430. Mishra, S.P., et al., *Multivariate Statistical Data Analysis- Principal Component Analysis (PCA)*. International Journal of Livestock Research, 2017. **7**(5): p. 60-78.
431. Chovancová, E., et al., *CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures*. PLoS Computational Biology, 2012. **8**(10): p. e1002708.
432. Pavelka, A., et al., *CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2016. **13**(3).

- 
433. Xu, D., Y. Xu, and E.C. Uberbacher, *Computational Tools For Protein Modeling*. Current Protein & Peptide Science, 2000. **1**: p. 1.
434. Pettersen, E.F., et al., *UCSF Chimera - A visualization system for exploratory research and analysis*. J. Comput. Chem., 2004. **25**(13): p. 1605-12.
435. Schrödinger, L., *Maestro*. 2019: New York, NY.
436. Zhu, K., et al., *Docking covalent inhibitors: A parameter free approach to pose prediction and scoring*. J. Chem. Inf. Model, 2014. **54**: p. 1932-1940.
437. Jacobson, M.P., et al., *On the Role of Crystal Packing Forces in Determining Protein Sidechain Conformations*. Journal of Molecular Biology, 2002. **320**: p. 597-608.
438. Jacobson, M.P., et al., *Force Field Validation Using Protein Sidechain Prediction*. Journal of Physical Chemistry B, 2002. **106**: p. 11673-11680.
439. Jacobson, M.P., et al., *A Hierarchical Approach to All-Atom Loop Prediction*. Proteins, 2004. **55**: p. 351-367.
440. Li, X., M.P. Jacobson, and R.A. Friesner, *High Resolution Prediction of Protein Helix Positions and Orientations*. Proteins, 2004. **55**: p. 368-382.
441. Andrec, M., et al., *Complete Protein Structure Determination Using Backbone Residual Dipolar Couplings and Sidechain Rotamer Prediction*. Journal of Structural and Functional Genomics, 2002. **2**: p. 103-111.
442. Guallar, V., et al., *Computational Modeling of the Catalytic Reaction in Triose Phosphate Isomerase*. Journal of Molecular Biology, 2004. **337**: p. 227-239.
443. Kalyanaraman, C., K. Bernacki, and M.P. Jacobson, *Virtual screening against highly charged active sites: Identifying substrates of alpha-beta barrel enzymes*. Biochemistry, 2005. **44**: p. 2059-2071.
444. [http://www.jacobsonlab.org/plop\\_manual/plop\\_overview.htm](http://www.jacobsonlab.org/plop_manual/plop_overview.htm). 2020 [cited 2020 27.08.2020].
445. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Nucleic Acids Res., 2000. **28**: p. 27-30.
446. Kanehisa, M., et al., *New approach for understanding genome variations in KEGG*. Nucleic Acids Res., 2019. **47**: p. D590-D595.
447. Kanehisa, M., *Toward understanding the origin and evolution of cellular organisms*. Protein Sci., 2019. **28**: p. 1947-1951.
448. Schrödinger, L., *The PyMOL Molecular Graphics System*.
449. Painsar, T. and A. Poso, *Binding Affinity via Docking: Fact and Fiction*. Molecules (Basel, Switzerland), 2018. **23**(8): p. 1899.
450. Sridhar, S., et al., *Crystallographic binding studies of rat peroxisomal multifunctional enzyme type 1 with 3-ketodecanoyl-CoA: capturing active and inactive states of its hydratase and dehydrogenase catalytic sites*. Acta Crystallographica Section D, 2020. **76**(12): p. 1256-1269.
451. Tobias, D.J. and C.L. Brooks, *Molecular dynamics with internal coordinate constraints*. J. Chem. Phys., 1988. **89**: p. 5115.
452. Hünenberger, P.H., A.E. Mark, and W.F. van Gunsteren, *Fluctuation and Cross-correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations*. Journal of Molecular Biology, 1995. **252**(4): p. 492-503.
453. Tiwari, S., et al., *WEBnm@ v2.0: Web server and services for comparing protein flexibility*. BMC Bioinformatics, 2014. **15**: p. 427.
454. Maplesoft, *Maple*. 2019, <https://www.maplesoft.com/>: Ontario, USA.
455. Hollup, S.V., G. Salensminde, and R.e. al, *WEBnm@: a web application for normal mode analyses of proteins*. BMC Bioinformatics, 2005. **6**: p. 52.
-