In vitro & in silico: Two complementary approaches to elucidate the molecular mechanisms of Medium-Chain Acyl-CoA dehydrogenase deficiency (MCADD)

Cátia Alexandra Marques Bonito Ferreira

Master Course in Biopharmaceutical Sciences
2015
In vitro & in silico: Two complementary approaches to elucidate the molecular mechanisms of Medium-Chain Acyl-CoA dehydrogenase deficiency (MCADD)

Cátia Alexandra Marques Bonito Ferreira

Master Thesis supervised by
Prof. Dr. Fátima V. Ventura
Prof. Dr. Rita C. Guedes

Master Course in Biopharmaceutical Sciences
2015
The experimental work presented in this dissertation was carried out at the Department of Biochemistry and Human Biology, Metabolism and Genetics Group, Research Institute for Medicines, (iMed.ULisboa), Faculdade de Farmácia da Universidade de Lisboa. All computational studies were performed on the computer cluster of the Department of Pharmaceutical Chemistry and Therapeutics, Medicinal Chemistry Group, Research Institute for Medicines, (iMed.ULisboa), Faculdade de Farmácia da Universidade de Lisboa.
Acknowledgments

O primeiro agradecimento irá para as minhas orientadoras, Professora Fátima Ventura e Rita Guedes por todo o apoio ao longo desta tese, e não menos importante, pelo companheirismo e compreensão que demonstraram ao longo destes dois anos.

À professora Fátima Ventura, um agradecimento especial por ter aceite a minha proposta e me ter deixado introduzir estudos computacionais nunca antes realizados no grupo Met&Gen, na tentativa de obter respostas a muitas questões levantadas pela parte experimental.

À Professora Rita Guedes, agradeço a prontidão que demonstrou em orientar a parte computacional desta tese, tentando conciliar a linguagem computacional com a bioquímica que nem sempre foi fácil.

A ambas também não posso deixar de agradecer o encorajamento ao meu sentido crítico, deixando muitas vezes que a curiosidade me guiasse pela procura de mais dados, valorizando sempre o meu entusiasmo.

Quero também agradecer à Professora Paula Leandro pelo seu envolvimento e curiosidade pelo trabalho que se desenvolveu ao longo da tese bem como pela sua contagiente simpatia.

Não poderei deixar de agradecer à Professora Isabel Rivera pelo facto de me ter sugerido trabalhar com a professora Fátima Ventura na problemática da MCADD, ainda durante o ano curricular deste mestrado.

Um agradecimento especial à Professora Isabel Tavares de Almeida por ter autorizado a realização da tese no seu grupo de investigação ao qual me afeiçoei bastante, rendida à simpatia e à integridade das pessoas que o constituem.

Gostaria de agradecer ao João Leandro, ao Paulo Lino, ao Henrique Colaço, à Mariana Amaro e à Ana coelho por me terem apoiado no laboratório aquando da parte experimental. De igual forma, quero agradecer à D. Ana e à D. Isabel pelo apoio dentro do laboratório e pela boa disposição de ambas que se tornou contagiente.

Um agradecimento deveras especial ao meu marido, Ricardo Ferreira, por todo o apoio incondicional que me têm dado ao longo destes quase três anos de curso. Nunca vou esquecer o apoio que me deu quando decidi sair da farmácia e mudar completamente de área, para algo que verdadeiramente gosto. O caminho desde então não têm sido nada fácil, pensei algumas
vezes em desistir mas o seu incansável apoio nunca me levou a concretizar tais pensamentos, tendo sido o maior apoio de sempre durante este mestrado.

Quero agradecer à minha filha, Leonor Ferreira, embora ainda com 4 anos, foi fonte de inspiração para a mãe, no sentido de me fazer ver que tinha de tentar mudar de vida, para algo que gostasse, mas que me desse mais tempo para disfrutar de todos os momentos do seu crescimento.

Quero agradecer à minha irmã, Cristina Bonito por ser a rapariga mais querida do mundo, encorajando-me nesta aventura, dizendo muitas vezes que gostava de ser tão corajosa como eu e por me ter deixado muitos chocolates em cima da mesa durante a escrita desta tese.

Um agradecimento especial aos meus sogros, Carolina Ferreira e João Ferreira pelo apoio incondicional durante esta tese de mestrado a vários níveis, salientando a sempre prontidão para ajudar em todas as tarefas necessárias sem nunca negar auxílio. A eles agradeço do fundo do meu coração!

Por último, um enorme agradecimento aos meus pais, Helena Bonito e Alexandre Bonito por me terem dado as ferramentas necessárias para prosseguir os meus estudos. Agradeço a forma como sempre me tem apoiado mesmo nas decisões mais difíceis e nem sempre idealizadas por eles. A eles ser-lhe-ei eternamente grata!
# TABLE OF CONTENTS

Abstract ........................................................................................................................................... x
Keywords ......................................................................................................................................... xi
Resumo ........................................................................................................................................... xii
Palavras-chave ................................................................................................................................... xiii
Index of Figures ............................................................................................................................... xiv
Index of Tables ............................................................................................................................... xviii
Index of Annexes ............................................................................................................................ xx
Abbreviations .................................................................................................................................... xxi

1. Introduction ................................................................................................................................. page 1
   1.1 Mitochondrial fatty acid β-oxidation (mFAO) pathway ....................................................... page 1
   1.2 Medium-chain fatty acid acyl-CoA dehydrogenase (MCAD) enzyme ................ page 2
   1.3 Medium-chain acyl-CoA dehydrogenase deficiency (MCADD) ................ page 5
2. Goals and working plan ................................................................................................................ page 7
3. Experimental methods ................................................................................................................ page 9
   3.1 Expression and purification of recombinant hMCAD proteins ................................ page 9
      3.1.1 Protein heterologous expression ........................................................................ page 9
      3.1.2 Protein purification ............................................................................................. page 10
      3.1.3 Protein quantification ........................................................................................ page 10
   3.2 Structural characterization of recombinant hMCAD proteins ................................ page 11
      3.2.1 Determination of the oligomeric profile of recombinant hMCAD proteins by Size
            Exclusion Chromatography (SEC) ................................................................ page 11
      3.2.2 Assessment of the conformation flexibility of recombinant hMCAD proteins by Limited
            proteolysis with trypsin ........................................................................................ page 11
      3.2.3 Determination of in vitro stability of recombinant hMCAD proteins by thermal
            denaturation using Differential Scanning Fluorimetry (DSF) .................... page 12
      3.2.4 Determination of in vitro stability of recombinant hMCAD proteins by thermal
            denaturation using DSF in the presence of chaperones ..................................... page 12
      3.2.5 Assessment of the aggregation profile and aggregation kinetics of recombinant
            hMCAD proteins by Dynamic Light Scattering (DLS) ..................................... page 13
   3.3 Functional characterization of recombinant hMCAD proteins ................................ page 14
3.3.1 Enzymatic activity and determination of the kinetic parameters

3.3.2 Determination of in vitro stability of recombinant hMCAD proteins by thermal enzymatic inactivation

4. Computational methodology

4.1 Molecular dynamics (MD) method – a brief overview

4.2 Methodology

4.2.1 Initial structures

4.2.2 Preparation of input data

4.2.3 Equilibration and production runs

4.2.4 Analysis of MD trajectories

5. Experimental results

5.1 Expression and purification of MCAD wild-type and mutant forms of recombinant hMCAD: yield and purity grade

5.2 Structural characterization of recombinant hMCAD proteins

5.2.1 Determination of the oligomeric profile of recombinant hMCAD proteins by Size Exclusion Chromatography (SEC)

5.2.2 Assessment of the conformation flexibility of recombinant hMCAD proteins by Limited proteolysis with trypsin

5.2.3 Determination of in vitro stability of recombinant hMCAD proteins by thermal denaturation using Differential Scanning Fluorimetry (DSF)

5.2.4 Determination of in vitro stability of recombinant hMCAD proteins by thermal denaturation using DSF in the presence of chaperones

5.2.5 Assessment of the aggregation profile and aggregation kinetics of recombinant hMCAD proteins by Dynamic Light Scattering (DLS)

5.3 Functional characterization of recombinant hMCAD proteins

5.3.1 Enzymatic activity and determination of the kinetic parameters

5.3.2 Determination of in vitro stability of recombinant hMCAD proteins by thermal enzymatic inactivation

6. Computational results

6.1 Characterization of Sus scrofa (pig) MCAD wild-type enzyme (pMCADwt)
6.1.1 Analysis of protein stability ......................................................... page 40
6.1.2 Characterization of the catalytic pocket ........................................ page 42
6.1.3 Active site and molecular mechanism ........................................... page 46
6.2 Characterization of human MCAD wild-type enzyme (hMCADwt) ...... page 49
   6.2.1 Analysis of protein stability ....................................................... page 49
   6.2.2 Characterization of the catalytic pocket ....................................... page 51
   6.2.3 Active site and molecular mechanism ......................................... page 54
6.3 Characterization of Sus scrofa (pig) MCAD p.K304E variant (pMCAD p.K304E) ........................................................................ page 58
   6.3.1 Analysis of protein stability ....................................................... page 58
   6.3.2 Characterization of the catalytic pocket ....................................... page 61
   6.4.1 Analysis of protein stability ....................................................... page 67
   6.4.2 Characterization of the catalytic pocket ....................................... page 70
7. Discussion/Conclusion ....................................................................... page 76
8. Future perspectives ............................................................................ page 83
9. References .......................................................................................... page 88
Abstract

The medium-chain acyl-CoA dehydrogenase deficiency (MCADD) is the most common genetic disorder affecting the mitochondrial fatty acid β-oxidation (mFAO) pathway, in particular the first dehydrogenation reaction of C4-C14 fatty acyl substrates. The mature MCAD enzyme is a homotetramer with Flavin Adenine Dinucleotide (FAD) as its natural cofactor. The most common mutation found in MCADD patients (c.985A>G in ACADM gene) is translated in the substitution of lysine by a glutamic acid residue at position 304 (p.K304E) of the mature protein. Accordingly, this substitution has been widely related with protein misfolding and tetramer instability. To better understand the molecular basis of the pathogenicity of the p.K304E variant in MCADD, in vitro and in silico approaches were undertaken.

Experimental in vitro studies comprised the heterologous expression of human MCAD wild-type (hMCADwt) and p.K304E variant proteins, followed by the recombinant proteins’ purification and functional and structural characterization. The results obtained show that, when compared to hMCADwt, the p.K304E variant is structurally unstable, presenting an alteration in its oligomeric profile, thermal instability and a higher susceptibility to trypsin limited proteolysis, the latter pointing to a higher conformational flexibility of the variant protein. The addition of FAD in the initial step of the purification of the recombinant proteins seems to increase the tetrameric fraction of the variant protein and recover its deficient enzymatic activity. Moreover, FAD also incremented the thermal stability of both hMCADwt and p.K304E proteins, as expected being FAD the MCAD’s cofactor. Promising results were also obtained with glycerol and TMAO tested, among other compounds, as chemical chaperones for the p.K304E variant.

To complement the in vitro data, molecular dynamics (MD) simulations of the MCAD enzyme were also performed and both porcine and human structures were studied. The initial coordinates were obtained, respectively, from the crystallographic structure of the porcine MCAD protein (PDB: 1UDY; pMCADwt), and by reversing the crystallographic structure of the E376G/T255E mutant of the human MCAD protein (PDB: 1EGC; hMCADwt). In a following step, both pMCADwt and hMCADwt proteins were used to generate the corresponding p.K304E variants. The results obtained showed that both wild-type proteins were stable during the simulation time. The FAD seems to have a structural role in the catalytic pockets and in the tetramer stability. The catalytic pockets are highly flexible and dynamic to better accommodate the substrate. The p.K304E variant seems to induce a different behavior over MD simulations regarding both pMCADwt and hMCADwt. The p.K304E mutation seems also to induce structural modifications in the catalytic pockets, ultimately affecting their volume. Finally, the mutation seems to asymmetrically affect the FAD and substrate binding affinities in both pMCAD and
hMCAD variants, which may explain the decrease of the enzymatic activity and catalytic efficiency observed in the p.K304E/hMCAD variant (~50% residual activity in respect to hMCADwt).

Altogether, the in vitro and in silico data obtained contribute to better clarify the structural and functional abnormalities of the p.K304E variant of hMCAD thereby opening perspectives for the search for small molecules as protein stabilizers towards the development of pharmacological strategies for the treatment of MCADD.

**Keywords:** Mitochondrial fatty acid beta-oxidation; MCAD; p.K304E; Molecular dynamic simulations; chaperones.
Resumo

A deficiência da desidrogenase dos ácidos gordos de cadeia média (MCADD) é a doença genética mais comum da beta-oxidação mitocondrial dos ácidos gordos (mFAO). Esta alteração é causada por um defeito na enzima MCAD que catalisa o primeiro passo da mFAO. A MCAD é um homotetrâmero, tendo como cofator a Flavina Adenina Dinucleótido (FAD), que desempenha um papel fundamental na catálise da desidrogenação dos ácidos gordos de cadeia média (C4-C14). Nos doentes com MCADD, a mutação mais frequentemente descrita no gene ACADM (c.985A>G), que codifica para a MCAD, resulta na substituição, na proteína madura, de uma lisina por um ácido glutâmico na posição 304 (p.K304E). Esta mutação que tem sido associada a alterações estruturais da proteína variante e instabilidade da sua forma tetramérica com perda de função da enzima. Para uma melhor compreensão das bases moleculares subjacentes à patogenicidade da variante p.K304E associada à MCADD, foram efetuados vários estudos in vitro e in silico.

Os estudos experimentais envolveram a expressão heteróloga da proteína MCAD selvagem (hMCADwt) e da sua variante p.K304E, seguida da purificação e caracterização funcional e estrutural das proteínas recombinantes. Os dados obtidos confirmam a instabilidade estrutural da variante p.K304E relativamente à forma selvagem, com redução da fração tetramérica e incremento do conteúdo em formas agregadas, de elevada massa molecular, diméricas e monoméricas. Foi igualmente observada instabilidade térmica e uma maior susceptibilidade à ação de proteases que aponta para uma maior flexibilidade conformacional da variante p.K304E. A adição de FAD no passo inicial da purificação da proteína parece contribuir para uma melhoria do perfil de oligomerização da variante bem como uma recuperação da sua deficiente actividade enzimática. A presença de FAD parece igualmente promover a estabilidade térmica, tanto a hMCADwt como a p.K304E, o que seria de esperar, já que se trata do cofator da MCAD. Resultados promissores foram igualmente obtidos com o glicerol e TMAO, testados juntamente com outros compostos, como chaperones químicos para a variante p.K304E.

Por forma a complementar os dados obtidos in vitro, simulações de Dinâmica Molecular (MD) foram efectuadas, tendo sido estudadas as estruturas das formas selvagens da MCAD porcina (pMCADwt) e da MCAD humana (hMCADwt). As coordenadas atómicas iniciais da pMCADwt foram obtidas a partir da correspondente estrutura cristalográfica disponível (PDB: 1UDY). Relativamente à hMCAD, as coordenadas atómicas tridimensionais foram obtidas após a reversão das mutações existentes na estrutura cristalográfica da variante E376G/T255E da proteína humana (PDB: 1EGC). Ambas as proteínas foram obtidas na presença de FAD e do
substrato octanoil-CoA, tendo sido adicionalmente efetuadas simulações na ausência de ligandos. Posteriormente, ambas as proteínas pMCADwt e hMCADwt foram utilizadas como modelo para obter as correspondentes variantes p.K304E. Os resultados obtidos demonstram que ambas as proteínas pMCADwt e hMCADwt se mostraram estáveis durante os 100 ns de simulação computacional. O centro ativo da enzima (que é igualmente o local de ligação do cofator) demonstrou ter um comportamento dinâmico e altamente flexível para uma melhor adaptação da estrutura à presença do substrato. Relativamente à variante p.K304E, foi observado um comportamento distinto das estruturas cristalográficas quando comparado com as respetivas pMCADwt e hMCADwt. A presença da mutação parece induzir modificações estruturais específicas no local de ligação do substrato. Finalmente, esta mutação também aparenta afetar as afinidades de ligação do cofator e do substrato em ambas as variantes pMCAD e hMCAD, embora de uma forma assimétrica, o que pode explicar a diminuição da atividade enzimática e da eficiência catalítica observadas na variante p.K304E (cerca de ~50% da atividade da forma selvagem).

Os dados in vitro e in silico obtidos contribuem para uma melhor elucidação das alterações estruturais e funcionais da variante p.K304E da proteína MCAD abrindo assim perspetivas para a descoberta de pequenas moléculas como estabilizadores proteicos e para o desenvolvimento de estratégias farmacológicas para o tratamento da MCADD.

**Palavras-chave:** Beta-oxidação mitocondrial dos ácidos gordos; MCAD; p.K304E variante; simulações de dinâmica molecular; chaperones.
## Index of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Fatty acid cellular uptake, activation and transport into the mitochondrial matrix for further fatty acid oxidation and ATP production</td>
<td>1</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Structural representation of MCAD wild-type monomer with localization of the 2 α-helical domains (purple), 1 β-sheet domain (yellow), FAD cofactor (dark yellow) and the substrate 3-thioctanoyl-CoA (green) and MCAD wild-type tetramer with localization of FAD (green) and substrate 3-thioctanoyl-CoA (dark yellow)</td>
<td>3</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Representation of the FAD and octanoyl-CoA moieties</td>
<td>3</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Representation of the catalytic pocket highlighting the FAD binding site, highlighting the substrate binding site and with both ligands demonstrating that the FAD and substrate binding sites are well defined inside the pocket</td>
<td>4</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Schematic representation of the α,β-dehydrogenation reaction</td>
<td>4</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Molecular surface representation of the electrostatic potential for the pMCADwt and its p.K304E variant</td>
<td>6</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Map of pETMCADwt construct</td>
<td>9</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Simplified flowchart of a typical molecular dynamics simulation</td>
<td>16</td>
</tr>
<tr>
<td>Figure 9</td>
<td>The substrate 3-thioctanoyl-CoA represented in all-atom or united-atom force-fields</td>
<td>18</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Representation of rhombic dodecahedral and a truncated octahedral simulation boxes when compared with the cubic box</td>
<td>19</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Schematic representation of periodic boundary conditions</td>
<td>19</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Electrophoretic profile obtained by SDS-PAGE at 12% for the fractions eluted during IMAC from 0.5 L culture for the recombinant proteins hMCADwt and p.K304E variant, purified in the absence of FAD and presence of FAD in cell lysis buffer, after expression in E.coli BL21(DE3) cells, at 27°C, for 21h, in the presence of 1mM IPTG</td>
<td>24</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Representative elution profiles of the recombinant hMCADwt and p.K304E variant, purified in the absence and presence of the FAD cofactor in the lysis buffer</td>
<td>25</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Degradation profile of the recombinant protein hMCADwt and the p.K304E variant, obtained by limited proteolysis by trypsin at 37°C, for 60 min</td>
<td>27</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Thermal stress induced denaturation monitored by differential scanning fluorimetry probing global protein unfolding from 20°C to 70°C</td>
<td>28</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Thermal stress induced denaturation monitored by differential scanning fluorimetry (DSF) probing global protein unfolding from 20°C to 70°C. Effect of glycerol, TMAO, Arginine and Taurine supplementation</td>
<td>30</td>
</tr>
</tbody>
</table>
Figure 17  Thermal stress induced denaturation monitored by differential scanning fluorimetry (DSF) probing global protein unfolding from 20°C to 70°C. Effect of FAD concentration ........................................... page 31

Figure 18  Derivatives of the data obtained for thermal stress induced denaturation monitored by differential scanning fluorimetry (DSF) probing global protein unfolding from 20°C to 70°C. Effect of glycerol, TMAO, Arginine and Taurine supplementation ........................................ page 32

Figure 19  Thermal aggregation profiles (from 20 to 70°C), of the recombinant protein hMCADwt, obtained by DLS .......................................................... page 33

Figure 20  Thermal aggregation profiles (from 20 to 70°C), of the recombinant protein p.K304E, obtained by DLS .......................................................... page 34

Figure 21A  Thermal aggregation kinetics of the recombinant hMCADwt, assessed by DLS at 37°C and 42°C .......................................................... page 35

Figure 21B  Thermal aggregation kinetics of the recombinant protein p.K304E, assessed by DLS at 37°C (Graphs A and C) and 42°C (Graph B and D) .......................................................... page 36

Figure 22  Variation in time of the absorbance at 600 nm of the pair PMS/DCPIP in the absence and presence of 33 μM octanoyl-CoA (C8-CoA) using 8 μg purified protein after SEC (tetrameric fraction) ........................................ page 37

Figure 23  Activity of hMCADwt and p.K304E tetrameric forms with C8-CoA as substrate (0 to 100 μM) and in the absence of FAD in the lysis buffer. .......................................................... page 38

Figure 24  Thermal inactivation profiles of the hMCADwt and p.K304E variant tetrameric fractions obtained in the absence of FAD in the lysis buffer during purification .......................................................... page 39

Figure 25  Root mean square deviation (RMSD) and radius of gyration (Rg) of pMCADwt tetramer MD simulations against the crystallographic coordinates (PDB ID: 1UDY) for the APO, FAD and LIPID systems .......................................................... page 41

Figure 26  Root mean square fluctuation (RMSF) of pMCADwt monomers’ MD simulations against the crystallographic coordinates (PDB ID: 1UDY) for the APO, FAD and LIPID systems .......................................................... page 42

Figure 27  Comparison between the β-factors retrieved from the crystallographic structure (PDB ID: 1UDY) and from MD simulations of the pMCADwt in the presence of FAD/SCOA (LIPID system) .......................................................... page 42

Figure 28  Probability distribution function of the average volumes of the catalytic pockets (A, B, C and D) of pMCADwt .......................................................... page 43

Figure 29  Representation of the ligand interactions with the protein residues lining the catalytic pocket of pMCADwt .......................................................... page 46

Figure 30  Representation of the catalytic pocket of pMCADwt in two different perspectives, highlighting the location of the residues that are important for the MCAD catalysis .......................................................... page 47

Figure 31  Root mean square deviation (RMSD) and radius of gyration (Rg) of hMCADwt tetramers’ MD simulations against the crystallographic coordinates (PDB ID: 1EGC - reversed) .......................................................... page 49
Figure 32  Root mean square fluctuation (RMSF) of hMCADwt monomers’ MD simulations against crystallographic coordinates (PDB ID: 1EGC – reversed) ................................................................. page 50
Figure 33  Comparison between the β-factors retrieved from the crystallographic structure (PDB ID: 1EGC) and from MD simulations of the hMCADwt ................................................................. page 51
Figure 34  Probability distribution function of the mean volumes of the catalytic pockets (A, B, C and D) of hMCADwt ................................................................. page 52
Figure 35  Representation of the ligand interactions with protein residues lining the catalytic pocket of hMCADwt (observed in all monomers) ........... page 53
Figure 36  Representation of the catalytic pocket of hMCADwt in two different perspectives, highlighting the location of residues that are important for MCAD catalysis ................................................................. page 55
Figure 37  Representation of the misalignment observed in the hMCADwt (LIPID system) .......................................................................................................................... page 56
Figure 38  Root mean square deviation (RMSD) and radius of gyration (Rg) of p.K304E/pMCAD variant MD simulations against the crystallographic coordinates (PDB ID: 1UDY-K304E) ................................................................. page 59
Figure 39  Root mean square fluctuation (RMSF) values obtained for the p.K304E/pMCAD variant for the APO, FAD and LIPID systems ........... page 60
Figure 40  Comparison between the β-factors retrieved from the crystallographic structure of p.K304E/pMCAD variant (PDB ID: 1UDY-K304E) and from MD simulations of the p.K304E/pMCAD ................................................................. page 61
Figure 41  Probability distribution function of the average volumes of the catalytic pockets (A, B, C and D) of the p.K304E/pMCAD variant ................. page 62
Figure 42  Free energy of binding (∆Gbind) in kJ.mol⁻¹ for the FAD and SCOA ligands in each monomer (A, B, C, D) of the p.K304E/pMCAD variant, in the LIPID system .......................................................................................................................... page 64
Figure 43  Representation of the ligand interactions with the residues lining the catalytic pocket of p.K304E/pMCAD variant ................................................................. page 66
Figure 44  Depiction of the catalytic pockets found in the MD simulation (LIPID system) of the p.K304E/pMCAD variant ................................................................. page 67
Figure 45  Root mean square deviation (RMSD) and radius of gyration (Rg) of p.K304E/hMCAD variant MD simulations against the reversed crystallographic coordinates (PDB 1EGC – K304E) ................................................................. page 68
Figure 46  Root mean square fluctuation (RMSF) values obtained for the p.K304E/pMCAD variant for the APO, FAD and LIPID systems ........... page 69
Figure 47  Comparison between the β-factors retrieved from the reversed crystallographic structure of p.K304E/hMCAD variant (left) (PDB ID: 1EGC – K304E) and from the structure obtained from MD simulations of p.K304E/hMCAD .......................................................................................................................... page 69
Figure 48  Probability distribution function of the average volumes of the catalytic pockets (A, B, C and D) p.K304E/hMCAD variant ................................................................. page 71
Free energy of binding ($\Delta G_{\text{bind}}$) in kJ.mol$^{-1}$ for the FAD and CCOA ligands in each monomer (A, B, C, D) of the p.K304E/hMCAD variant, in the LIPID system ................................................................. page 73

Representation of the ligand interactions with the residues lining the catalytic pocket of p.K304E/hMCAD variant: representation of the FAD interactions ................................................................. page 74

Representation of the ligand interactions with the residues lining the catalytic pocket of p.K304E/hMCAD variant: representation of the CCOA interactions ................................................................. page 75

Depiction of the catalytic pockets found in the p.K304E/hMCAD simulation ................................................................................................................................. page 76
Index of Tables

Table 1  Yield and purity grade of the recombinant hMCADwt and p.K304E variant, purified in the absence (-FAD) and presence of FAD (+ FAD) in the lysis buffer ............................................................. page 25

Table 2  Distribution of the oligomeric forms obtained by SEC of the recombinant hMCADwt and p.K304E mutant forms, in the absence (A) and presence (B) of FAD in the lysis buffer during purification ........................................ page 26

Table 3  Thermal induced protein melting point ($T_m$) of the recombinant wt and p.K304E variant .......................................................... page 28

Table 4  Melting temperature ($T_m$) determined by differential scanning fluorimetry (DSF) of tetrameric hMCADwt and p.K304E variant in the absence and presence of chemical and pharmacological chaperones ...... page 30

Table 5  Temperature midpoint of aggregation ($T_{agg}$) and aggregation transition points of the recombinant hMCADwt and p.K304E variant obtained by Dynamic Light Scattering (increasing temperature within the range of 20 to 70°C) ............................................................. page 34

Table 6  Aggregation kinetics (min) at 37°C and 42°C, by DLS of recombinant hMACDwt and p.K304E variant purified in the absence and presence of FAD in the lysis buffer before IMAC purification ........................................ page 35

Table 7  Specific enzymatic activity, determined as a function of the variation of the DCPIP concentration in the reaction ($\mu$mol DCPIP.min$^{-1}$.mg$^{-1}$), for the hMCADwt and the p.K304E variant after SEC purification (tetrameric fraction), in the absence and presence of the cofactor FAD .............. page 37

Table 8  Kinetic parameters of recombinant hMCADwt and p.K304E variant (purified in the absence of FAD cofactor) using octanoyl-CoA as substrate after purification by SEC (tetrameric fraction) ................... page 38

Table 9  Hydrogen bonds between the FAD of one monomer and the neighboring monomer of the dimer unit in pMCADwt – FAD system (A) and LIPID system (B) .................................................. page 44

Table 10  Hydrogen bonds between the protein residues located at the active site of pMCADwt (APO system), that are described to be important on MCAD catalysis. (A) APO system; (B) FAD system and (C) LIPID system ........ page 48

Table 11  Hydrogen bonds between the FAD of one monomer and the neighboring monomer of the dimer unit in hMCADwt – FAD system (A) and LIPID system (B) ................................................................. page 11

Table 12  Hydrogen bonds between the residues located in the active site of hMCADwt (APO system), that are described to be important on MCAD catalysis. (A), APO system; (B), FAD system and (C), LIPID system ........ page 57

Table 13  Hydrogen bonds between the FAD of one monomer and the neighboring monomer of the dimer unit in p.K304E/pMCAD – FAD system (A) and LIPID system (B) ............................................................................. page 62

Table 14  Free energy of binding ($\Delta G_{bind}$) in kJ.mol$^{-1}$ for the FAD and SCOA ligands in each monomer (A, B, C, D) of the p.K304E/pMCAD variant, in the LIPID system .................................................................................. page 63
Table 15  Hydrogen bonds between the FAD of one monomer and the neighboring monomer of the dimer unit in p.K304E/hMCAD – FAD system (A) and LIPID system (B) ................................................................. page 72

Table 16  Free energy of binding ($\Delta G_{\text{bind}}$) in kJ.mol$^{-1}$ for the FAD and CCOA ligands in each monomer (A, B, C, D) of the p.K304E/hMCAD variant, in the LIPID system ................................................................................................................................. page 72
Index of Annexes

S1. Molecular dynamics simulation parameters
   a) Steepest descent algorithm (energy minimization) .............................................. a-ii
   b) Temperature equilibration (NVT ensemble) ....................................................... a-ii
   c) Pressure equilibration (NpT ensemble) ............................................................. a-ii
   d) Production run (NpT ensemble) ........................................................................... a-iii
   e) Specific parameters when considering the FAD cofactor and the substrate .......... a-iii

S2. Characterization of the *Sus scrofa* (pig) MCAD wild-type enzyme (pMCADwt) ....... a-iv

S3. Characterization of the human MCAD wild-type enzyme (hMCADwt) ....................... a-vii

S4. Characterization of the *Sus scrofa* (pig) pK304E mutated MCAD enzyme
    (p.K304E/pMCAD) ................................................................................................. a-x

S5. Characterization of the human K304E mutated MCAD enzyme (p.K304E/hMCAD) ... a-xiii

S6. Free energies of binding
   a) $g_{mmpbsa}$ run parameters ................................................................................ a-xvi

S7. Sequence alignment for human MCAD (SWISS-PROT P11310), pig (P41367),
    *C. elegans* (Q22347), mouse (P45952), rat (P08053) and bovine (Q3SZB4). .... a-xvii
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td><strong>ACADM</strong></td>
<td>gene Acyl-Coenzyme Dehydrogenase, C-4 to C-12 straight chain</td>
</tr>
<tr>
<td>ACS</td>
<td>Fatty acyl-CoA synthase</td>
</tr>
<tr>
<td>ATB</td>
<td>Automated Topology Builder and Repository</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td><strong>BamHI</strong></td>
<td><em>Bacillus amyloli</em> type II restriction endonuclease</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>C8-carnitine</td>
<td>Octanoyl-CoA acylcarnitine</td>
</tr>
<tr>
<td>CACT</td>
<td>Carnitine/Acyl-Carnitine translocase</td>
</tr>
<tr>
<td>CC</td>
<td>Chemical Chaperones</td>
</tr>
<tr>
<td>CCOA</td>
<td>Octanoyl-CoA</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CoASH</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>CPT1</td>
<td>Carnitine palmitoyltransferase 1</td>
</tr>
<tr>
<td>CPT2</td>
<td>Carnitine palmitoyltransferase 2</td>
</tr>
<tr>
<td>D</td>
<td>Aspartic acid</td>
</tr>
<tr>
<td>Da</td>
<td>Daltons</td>
</tr>
<tr>
<td>DCPIP</td>
<td>2,6-dichlorophenolindophenol</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic Light Scattering</td>
</tr>
<tr>
<td>DSF</td>
<td>Differential Scanning Fluorimetry</td>
</tr>
<tr>
<td>E</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>ETF</td>
<td>Electron-Transfer Flavoprotein</td>
</tr>
<tr>
<td><strong>ETF-QO</strong></td>
<td>Electron-transfer Flavoprotein ubiquinone oxidoreductase or ETF-ubiquinone oxidoreductase or ETF-Dehydrogenase</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin Adenine Dinucleotide (fully oxidized form; quinone form)</td>
</tr>
<tr>
<td>FADH2</td>
<td>Flavin Adenine Dinucleotide (reduced form; hydroquinone form)</td>
</tr>
<tr>
<td><strong>FAT/CD36</strong></td>
<td>Fatty Acyl Translocase</td>
</tr>
<tr>
<td>FATP</td>
<td>Fatty acid transport proteins</td>
</tr>
<tr>
<td>Fig.</td>
<td>Figure</td>
</tr>
<tr>
<td>G</td>
<td>Guanine (cDNA); Glycine (protein)</td>
</tr>
<tr>
<td>H</td>
<td>Histidine</td>
</tr>
</tbody>
</table>
HB Hydrogen bond

\textit{Hind}III \textit{Haemophilus influenza} type II site-specific deoxyribonuclease restriction enzyme

hMCAD human Medium-Chain Acyl-CoA Dehydrogenase

IMAC Immobilized Metal Affinity Chromatography

IMM Inner Mitochondrial Membrane

IPTG Isopropyl-\beta-D-1-thiogalactopyranoside

K Lysine (protein)

\( K \) Kelvin degrees

Kan Kanamycin

Kb kilobase

\( K_m \) Michaelis-Menten kinetic constant or affinity constant

\( K_{\text{cat}} \) turnover number or catalytic constant

Kpi Phosphate buffer

L Leucine

LB Luria Bertani liquid

M Molar mass

\( M_r \) Relative molecular weight

MCAD Medium-Chain Acyl-CoA dehydrogenase

MCADD Medium-Chain Acyl-CoA dehydrogenase Deficiency

MCS Multiple Cloning Site

MD Molecular Dynamics

mFAO Mitochondrial Fatty Acid Oxidation

NaCl Sodium chloride

NBS Newborn Screening

\( \text{Ni}^{2+} \) Nickel divalent cation

ng nanogram

nm Nanometer

OCTN2 organic cation/carnitine transporter

OMIM Online Mendelian Inheritance in Man

ON Overnight

PAGE Polyacrylamide Gel Electrophoresis

PC Pharmacological Chaperones

PDB Protein Data Bank

pMCAD pig Medium-Chain Acyl-CoA Dehydrogenase
PME  Particle Mesh Ewald
PMS  Phenazine methosulfate
PMSF  Phenylmethylsulfonyl fluoride
ps  Picoseconds
PstI  Providencia stuartii type II restriction endonuclease
R  Arginine
RC  Respiratory Chain
rpm  Rotations per minute
RT-PCR  Real-Time Polymerase Chain Reaction
SCOA  3-thiaoctanoyl-CoA
SDS  Sodium Dodecyl Sulfate
SEC  Size Exclusion Chromatography
s  Seconds
$t_{1/2}$  half-life
T  Threonine
$T_{agg}$  Aggregation temperature
TaqI  Thermus aquaticus restriction enzyme
TCA  Tricarboxylic Acid Cycle
TMAO  Trimethylamine N-oxide
$Tm$  melting temperature
V  Valine
$V_{max}$  Maximum velocity (enzyme kinetics)
$V_r$  Retention volume
wt  Wild-type
Y  Tyrosine
$\beta$-ME  Beta-mercaptopethanol
6xHis  Six histidines codon sequence
1. Introduction

1.1. The mitochondrial fatty acid β-oxidation (mFAO) pathway. Fatty acids are the major sources of energy in organs/tissues as liver, heart and skeletal muscle and mitochondrial fatty acid β-oxidation (mFAO) their main metabolic pathway (Fig. 1). Fatty acids, released from fatty depots or obtained from diet, are translocated into the cell by a mechanism mediated by various protein transporters, including the fatty acid translocase (FAT/CD36), the tissue specific fatty acid transport proteins (FATPs) and the plasma membrane fatty acid-binding protein (FABPpm). Previous to breakdown within the mitochondria, fatty acids have to be activated to coenzyme A thioesters by substrate specific fatty acyl-CoA syntheses (ACS).

To pass through the two mitochondrial membrane barriers, the acyl-CoA thioesters produced in the cytosol (particularly the long-chain acyl-CoA thioesters) have to be firstly converted by Carnitine palmitoyltransferase 1 (CPT1), located within the inner leaflet of the outer mitochondrial membrane, into their carnitine derivatives which are further translocated across the inner mitochondrial membrane (IMM) by Carnitine/Acyl-Carnitine Translocase (CACT) in exchange for free carnitine. Once inside the mitochondria they are reconverted by Carnitine...
Palmitoyltransferase 2 (CPT2), associated with the inner leaflet of the IMM, back to the original acyl-CoA thioesters, the true substrates of mFAO. The mFAO pathway is composed by four consecutive reactions (Fig. 1) catalyzed by various isoenzymes with substrate specificity. Each cycle yields an acyl-CoA thioester shortened in two carbon atoms, which returns to the β-oxidation cycle, and an acetyl-CoA molecule, that enters into the tricarboxylic acid cycle (TCA) to further produce Adenosine Triphosphate (ATP) in the Respiratory chain (RC) (Fig. 1). The acetyl-CoA may also be converted into steroid derivatives or used in the synthesis of ketone bodies (energetic alternative substrates).

1.2. Medium-chain fatty acid acyl-CoA dehydrogenase (MCAD) enzyme. The medium-chain acyl-CoA dehydrogenase (MCAD) enzyme (EC 1.3.99.3) belongs to the acyl-CoA dehydrogenases family of flavoproteins and catalyzes the first step of mFAO. It is a mitochondrial protein encoded by the nuclear ACADM gene, located in the short arm of chromosome 1 at position 31 (1p31) (Genetics Home Reference http://ghr.nlm.nih.gov/gene/ACADM). This gene has more than 44 kb organized in 12 exons. The MCAD protein is translated in the cytosol as a precursor with 421 amino acids holding an N-terminal peptide signaling sequence of 25 amino acids, which directs the protein to mitochondria where it acquires the correct folding. The mature protein, with 396 amino acids per monomer is formed when the signal peptide is cleaved off (Fig. 2A). MCAD is a homotetramer (Fig. 2B) being each monomer divided into three structural domains of approximately equal size: the N-terminal α-domain (from N-terminus to L129) and the C-terminal α-domain (from E240 to C-terminus) flanking an intermediate β–sheet domain (from M130 to G239). The N- and C-terminal α-domains consist of six α-helices (A-F and G-L, respectively) that together seem to form a single α-domain. The β-domain is located at the protein surface and is composed by two orthogonal β-sheets. The Flavin Adenine Dinucleotide (FAD) is the natural cofactor of MCAD which binds non-covalently in the active site, being critical for the enzyme’s function. It seems also to have a structural role as it has been described that upon incorporation of FAD in each monomer, dimerization is promoted followed by the assembly of two homodimers into the final tetrameric form. The tetramer may be considered a dimer of dimers (northern and southern hemispheres) and the interface between them involves groups of four α-helices from the C-terminal α-domain of each monomer (protein core) (Fig. 2B).

The FAD and the fatty acyl-substrate moieties (Fig. 3) as well as their positions inside the catalytic pocket of MCAD are shown in Figures 4 and 5, respectively. The flavin ring along with the ribityl
group (riboflavin moiety) of FAD are located between the β-domain and the C-terminal α-domain of one monomer with the sinister (si)-face of flavin ring directed to β-domain.\textsuperscript{5}

**Figure 2:** Structural representation of: (A) MCAD wild-type monomer with localization of the 2 α-helical domains (purple), 1 β-sheet domain (yellow), FAD cofactor (dark yellow) and the substrate 3-thiaoctanoyl-CoA (green); (B) MCAD wild-type tetramer with localization of FAD (green) and substrate 3-thiaoctanoyl-CoA (dark yellow). The dashed line represents the interface between two dimers: monomers A and B (left) and monomers C and D (right). Figures were created with MOE v2013.08 from PDB ID: 1UDY.

The pyrophosphate portion of FAD lies between the β-domain of one monomer and the C-terminal α-domain of the other, while the adenosine group is located at the interface between the two monomers of the dimer (Fig. 4A).\textsuperscript{6-8} The binding site for the fatty-acyl substrate is mainly formed at the interface between the β-domain and C-terminal α-domain of the monomer (Fig. 4B). The terminal acyl group is deeply buried between the N- and C-terminal α-domains facing the rectus (re)-face of the flavin ring of FAD (Fig. 4C).

**Figure 3:** Representation of the FAD and octanoyl-CoA moieties (obtained from www.chemicalize.org).
The 3’-phosphoadenosine of the CoA moiety is located at the interface between the two neighboring monomers of the dimer and partially exposed to solvent (Fig. 4C). The catalytic pocket can accommodate acyl-chains from C4 to C14 with higher affinity for C6 to C10 fatty acyl-groups.

The Cα-Cβ bond of the substrate is sandwiched between the carboxylate group (side-chain) of the glutamic acid residue E376 (catalytic residue) and the flavin ring of FAD (Fig. 4C), which favors the α,β-dehydrogenation reaction (Fig. 5).

Mechanistically, the catalytic residue E376 abstracts the α-hydrogen of the acyl moiety of the substrate as a proton and the β-hydrogen is transferred as a hydride onto the N(5) of the flavin ring of FAD, yielding a trans-Δ2-enoyl-CoA derivative and FADH₂, respectively.

![Figure 4](image)

**Figure 4:** Representation of: (A) the catalytic pocket highlighting the FAD binding site (FAD; dark yellow); (B) the catalytic pocket, highlighting the substrate binding site (3-thiaoctanoyl-CoA; dark green); and (C) the catalytic pocket with both ligands demonstrating that the FAD and substrate binding sites are well defined inside the pocket. The purple regions represent more hydrophilic regions of the pocket and the green regions represented the most hydrophobic. Figures created with MOE v2013.08 from PDB ID: 1UDY.

The Cα-Cβ bond of the substrate is sandwiched between the carboxylate group (side-chain) of the glutamic acid residue E376 (catalytic residue) and the flavin ring of FAD (Fig. 4C), which favors the α,β-dehydrogenation reaction (Fig. 5).

Mechanistically, the catalytic residue E376 abstracts the α-hydrogen of the acyl moiety of the substrate as a proton and the β-hydrogen is transferred as a hydride onto the N(5) of the flavin ring of FAD, yielding a trans-Δ2-enoyl-CoA derivative and FADH₂, respectively.

![Figure 5](image)

**Figure 5:** Schematic representation of the α,β-dehydrogenation reaction. The catalytic residue E376 removes the α-hydrogen of the acyl moiety of the substrate as a proton with the formation of an intermediate species. In a following step, the β-hydrogen is transferred as a hydride onto the N(5) of the flavin ring of FAD, yielding an enoyl-CoA derivative and FADH₂, respectively. (adapted from Rudik et al.11)
The redox regeneration of FAD releases electrons into the electron transfer flavoprotein (ETF) which in turn are transferred to Ubiquinone (coenzyme Q), within the RC, in a reaction mediated by the electron transfer flavoprotein ubiquinone oxidoreductase (ETF-QO) enzyme.3,6

1.3. The medium-chain acyl-CoA dehydrogenase deficiency (MCADD). The medium-chain acyl-CoA dehydrogenase deficiency (MCADD; OMIM #201450; ORPHA42) is the most common genetic disorder of mFAO.3 It is an autosomal recessive inborn metabolic error with heterogeneous clinical, biochemical and mutational profiles, which difficult the diagnosis, the establishment of genotype/phenotype correlations and the treatment.12 Patients with MCADD have recurrent metabolic crisis with hypoketotic hypoglycemia and lethargy, due to rapid exhaustion of glycogen depots.3 They can also present Reye-like syndrome with seizures, coma and unexpected death. The symptoms may be triggered by situations that induce cellular catabolism as prolonged fasting, intense exercise or illness with fever episodes.3,13 The determination of the MCADD biochemical profile is crucial for the correct diagnosis of the disease. It is characterized by high levels of blood/plasma fatty acylcarnitines (C6-C10), with particular reference to the levels of octanoyl-CoA acylcarnitine (C8-carnitine), an essential metabolite present in symptomatic or asymptomatic individuals.14 In case of a positive result, other metabolites are measured as well (e.g. plasmatic free fatty acids, urinary medium-chain dicarboxylic acids, resultant from alternative metabolic pathways, and acylglycine conjugates), however, they are only detected in the acute phases of the disease.14

The MCADD presents a high rate of morbidity and mortality, especially in childhood.9,15 Children with MCADD have 4% risk of mortality within the first 3 days of life, rising up to 5-7% until 6 years of age.9 However, individuals can be asymptomatic until their first metabolic crisis which may occur only in adulthood.16 Epidemiological data reveal that the worldwide prevalence of MCADD is around 1:30,000 to 1:45,000 (95% CI) based on clinical onset.9 Nevertheless, after the global expansion of the newborn screening (NBS) programs to include the diagnosis of MCADD the prevalence of the disease increased to 1:15,000.17 In Portugal, according to the 2014 report from the Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA,IP), the MCADD birth prevalence since 2004 – when the disease was introduced in the NBS Portuguese program – is 1:8,804 (> 500,000 screened newborns), making this disease the most frequent inborn error of metabolism in the country.18

In many countries, the diagnosis of MCADD through NBS programs, revealed a wide genotypic heterogeneity of the human MCAD (hMCAD) encoding gene (ACADM). At present, about 100 different mutations have been reported in MCADD patients (Public Human Gene Mutation Database®, Cardiff, UK). Mostly, they are point mutations resulting in single amino acid changes
(missense mutations) in the hMCAD sequence, which can alter the enzyme’s function and/or structure impairing or fully abolishing its activity.\textsuperscript{5,15} The most common mutation found in MCADD patients is the transition in exon 11 of the \textit{ACADM} gene of an adenine to a guanine in nucleotide 985 (c.985A>G). This mutation, in mature hMCAD protein sequence, is translated in the substitution of the residue lysine 304 by a glutamic acid (p.K304E) which is described to promote mild to severe alterations in the enzyme’s function and/or structure.\textsuperscript{5,15} At the protein level, this mutation is located in the middle of α-helix H, which makes part of the dimer-dimer interface of the tetramer (Fig. 6). The change of a lysine (positively charged amino acid) (Fig. 7A) by a glutamic acid (negatively charged amino acid) in this critical region, leads to a negative charge concentration (Fig. 7B) which may affect the folding and the tetramer assembly.\textsuperscript{4,5,7,15,19}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{Molecular surface representation of the electrostatic potential (blue, electron donor; white, neutral; red, electron acceptor) for the pMCADwt (A) and its p.K304E variant (B). The substitution of a lysine by a glutamic acid in the α-helix H (left helix) causes an excess of negative charge in this region that could affect the tetramer stability. Figures created with MOE v2013.08 from PDB ID: 1UDY.}
\end{figure}

The majority of North-Western European MCADD patients are homozygous for the p.K304E mutation. However, many others are compound heterozygotes for this mutation, i.e., they have the most common mutation in one allele and a rare mutation in the other.\textsuperscript{15} In Portugal, France and Spain the disease is more prevalent in Gypsies, in particular due to homozygosity for p.K304E mutation.\textsuperscript{9} Historical data reveal a Gypsy migratory flow, in large scale, from India to Europe during the Byzantine Empire (900-1,100 years ago), becoming one of the first Europeans.\textsuperscript{20} Population-genetic studies between European countries colonized by Gypsies and European countries without gypsy influence reveal a heterogenic population with different genetic patrimonies. A recent study from Ventura \textit{et al}.\textsuperscript{9} found the same \textit{ACADM} genetic background between MCADD patients with gypsy ancestry and MCADD patients descendent from North-Western European Caucasians – homozygosity for p.K304E mutation and for the H1
haplotype of *ACADM* (based in the intragenic bi-allelic polymorphisms for *BamHI, PstI* and *TaqI*).\textsuperscript{9,21} This strongly supports the hypothesis that p.K304E is an ancient mutation originated in a common ancestor of the first Europeans.

Currently, the long-term treatment of MCADD patients consists mainly in the avoidance of fasting through frequent feeding, although in many countries a high-carbohydrate and a reduced-fat diet, supplemented with oral L-carnitine, is also recommended.\textsuperscript{15,22} Patients with severe hypoglycemia must be treated immediately with intravenous administration of glucose solution to avoid episodes of coma and death.\textsuperscript{16}

Despite under dietary control, many patients present acute life-threatening decompensation episodes, being necessary the search for novel therapeutic strategies for MCADD treatment.\textsuperscript{23} Pharmacological (PC) and chemical chaperones (CC) have been studied in the context of some inborn errors of metabolism, such as Phenylketonuria and Lysosomal Storage Diseases, as possible therapeutic approaches for disorders associated with misfolded proteins resulting from missense mutations (conformational disorders).\textsuperscript{24} As MCADD is also considered a conformational disorder,\textsuperscript{23} characterized by protein misfolding with loss-of-function, PCs and CCs may also constitute promising therapeutic strategies for the treatment of MCADD patients.\textsuperscript{24} The CCs do not bind directly to misfolded proteins and are unspecific for the target protein. Their mechanism of action involves changes in the protein environment by increasing the protein hydration and subsequently its compactness.\textsuperscript{25,26} The CCs are mostly osmolytes such as polyols (e.g. glycerol, sorbitol), sugars (e.g. threalse), methylamines [e.g. trimethylamine-N-oxide (TMAO), betaine and glycerophosphorylcholine] or amino acids (e.g. arginine, glycine and taurine).\textsuperscript{25} The PCs act more specifically, binding to target misfolded protein and promoting its stabilization. Thereby, these small molecules avoid early protein degradation, increasing the pool of the functional protein inside the cell.\textsuperscript{26} The PCs can be competitive inhibitors, ligands, agonists/antagonists, and protein cofactors, including metal ions.\textsuperscript{25}

2. Goals and working plan

A recent study involving MCADD patients and respective families revealed in Portugal a higher percentage (94%) of homozygotes for the most common variant p.K304E.\textsuperscript{9} However, among the cohort studied there were also patients detected with the p.K304E in one allele and a rare mutation in the other allele (compound heterozygotes) as observed in other countries.\textsuperscript{5,9} Due to a highly heterogeneous profile of MCADD, the molecular mechanisms underlying this enzymatic deficiency remains poorly understood, hindering the search for novel therapeutic strategies for the treatment of these patients.
To better understand the molecular mechanisms underlying MCADD associated with the most common variant p.K304E, a combination of in vitro and in silico approaches was undertaken. The experimental studies aimed: 1) the functional and structural characterization of the wild-type (wt) and p.K304E forms of hMCAD and 2) to explore the effect of a selected group of compounds (small molecules) in the structure and function of the p.K304E variant. These compounds are currently being tested as potential modifiers of misfolded proteins involved in other metabolic genetic diseases, and may also constitute pharmacological strategies for the treatment of MCADD patients carrying the commonest mutation.

To complement and provide insights into the experimental data, molecular dynamics (MD) simulations aiming the structural characterization and the dynamic behavior of the MCADwt enzyme as well as its p.K304E variant were carried out.

**Working plan**

1) **Experimental studies**

The hMCADwt and the p.K304E variant were obtained using an *E. coli* expression system previously developed in the Met&Gen group (Metabolism and Genetics Group, Research Institute for Medicines, iMed.ULisboa, Faculty of Pharmacy, University of Lisbon). The recombinant proteins were purified by Immobilized Metal Affinity Chromatography (IMAC) followed by analysis of the electrophoretic profiles of the eluted fractions with a polyacrylamide denaturing gel (SDS-PAGE at 12%) and protein quantification by the Bradford assay. By these mean it was possible to determine the yield and purity grade of the produced recombinant proteins. The tetrameric forms of each recombinant protein were isolated from IMAC eluted fractions by Size Exclusion Chromatography (SEC) and used for further functional and structural characterization.

The functional characterization of the recombinant hMCAD proteins comprised the determination of the enzymatic activity and the respective kinetic parameters (*V*<sub>max</sub>, *K*<sub>m</sub>, *K*<sub>cat</sub> and catalytic efficiency). To study the structure and stability of the recombinant hMCAD proteins several cutting to edge methodologies were applied: Size Exclusion Chromatography (SEC), Trypsin limited proteolysis, Thermal inactivation profile, Differential Scanning Fluorimetry (DSF) in the absence and presence of chaperones and Dynamic Light Scattering (DLS).

2) **Computational studies**

In these work, MCAD dynamics was studied for the first time using molecular dynamics (MD) simulations under different conditions. The crystallographic structure of the *Sus scrofa* (pig) MCAD enzyme (PDB ID: 1UDY), which has more than 90% of homology with the hMCAD, was
used as the initial structure to perform MD simulations. In order to relate the computational with the experimental data, MD simulations of the hMCAD were also performed.

Five crystallographic structures of the hMCAD enzyme are available in the Protein Data Bank (PDB) website, however, these structures were obtained either with the protein modified with point mutations, namely in the catalytic residue E376, or in complex with the ETF flavoprotein. Among the available structures, the crystallographic structure of the E376G/T255E double mutant of hMCAD enzyme (PDB ID: 1EGC) was used as template to be reverted into the hMCADwt. In a following step, the p.K304E mutation was inserted into the pig MCAD (pMCAD) and the obtained hMCADwt structures and the eventual modifications in the protein dynamics were characterized.

3. Experimental Methods

3.1 Expression and purification of recombinant hMCAD proteins.

3.1.1 Protein heterologous expression: Two constructs pETMCAD (pETMCAD_{wt} and pETMCADd_{K304E}) were previously prepared for the heterologous expression of hMCAD_{wt} and p.K304E variant proteins (Fig. 7). To distinguish the protein of interest from the bacterial endogenous proteins, the cDNA of the recombinant protein has at its C-terminus a six histidine codon sequence (6xHis-tag) that is crucial for protein purification. The transformation of 100 μL of competent *E. coli* BL21 (DE3) cells was made by heat-shock (42°C, 90 s) with 10 ng of the plasmid DNA containing the cDNA of interest (pETMCAD construct).

After transformation, the cells were incubated at 37°C in liquid Luria Bertani (LB) cell culture medium (1% tryptone; 0.5% yeast extract; 0.17 M NaCl; pH 7.0) with 20 mM glucose for one hour with agitation (160 rpm). The cells were then plated onto solid LB medium (liquid LB medium plus agar 1.5%) with Kanamycin (Kan) at a final concentration of 100 μg/mL (solid LB/Kan) and incubated overnight (ON) at 37°C. From the obtained transformants, an isolated colony was inoculated in LB medium, in the presence of 50 μg/mL Kan, and further incubated ON at 37°C with agitation (160 rpm).

In the following day a 1/100 dilution was performed in LB medium with 50 μg/mL Kan and further incubated at 37°C as above until an
optical density at 600 nm of 0.6. At this point, protein expression was induced by adding 0.5 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubating over 21 hours at 27°C with agitation (140 rpm). The cells were isolated by centrifugation (4000 rpm/10 min/4°C) and the resulting pellet used immediately or alternatively stored at -20°C up to a month.

3.1.2 Protein purification: The recombinant hMCAD proteins were purified by Immobilized Metal Affinity Chromatography (IMAC) based in the work of Zeng and Li. In order to first lyse the bacterial cells, a lysis buffer was prepared [20 mM phosphate buffer (Kpi) pH 7.4; 500 mM NaCl; 1 mM phenylmethylsulfonyl fluoride (PMSF); 1 mg/mL of lysozyme; Dnase] in the presence and absence of 0.1 mg/mL of FAD cofactor. The cells were further incubated in this buffer at 4°C (30 min) and sonicated (4°C; 4 X 60 s; 50% of duty free cycle). A centrifugation (10,000 rpm/40 min/4°C) was followed to separate the soluble and insoluble cellular fractions. The IMAC resin (Ni2+-NTA) was equilibrated (3 X MilliQ water; 3 X 20 mM Kpi, pH 7.4; 500 mM NaCl plus 10 mM imidazole) and added to the soluble fraction (supernatant). The mixture (resin plus lysate) was incubated with β-mercaptoethanol (10 mM) and imidazole (10 mM) for an hour at 4°C with agitation. The IMAC purification protocol comprised an imidazole gradient: 2 x 5 ml of 20 mM; 2 x 2.5 ml of 40 mM; 5 x 0.5 ml of 500 mM imidazole in elution buffer (20 mM Kpi pH 7.4; 500 mM NaCl).

In order to identify the fractions where the recombinant protein were eluted as well as to evaluate their purity grade, IMAC eluted fractions were analyzed by SDS-PAGE at 12 %, after Coomassie brilliant blue R staining. Protein purity grade was assessed by densitometric scanning of major and minor bands using the Image J® 1.43U NIH software (Image Processing and Analysis in Java). The fractions containing the recombinant protein were combined in a dialysis membrane (cut-off of 12-14 kDa) and incubated ON at 4°C with agitation, in the dialysis buffer (20 mM Kpi, pH 7.4; 5 mM β-mercaptoethanol; 5% glycerol). Alternatively, the IMAC fractions containing the purified enzyme were pooled and loaded into a desalting gravity-feed PD-10 column with Sephadex™ G-25 Medium (exclusion limit: 5 kDa; GE Healthcare Life Sciences; Uppsala, Sweden) and eluted with SEC buffer (20 mM Kpi, pH 7.4; 200 mM NaCl). In this case the proteins were further concentrated using a ViVaSpin 15R (Sartorius Stedim Biotech, GmbH, Goettingen, Germany) concentrator by centrifugation (3,000g/10 min/4°C). The purified proteins were used immediately or stored at -80°C or in liquid nitrogen (N2) in 20% glycerol.

3.1.3 Protein quantification: In order to analyze protein expression levels and thus determine the yield of the expression and purification steps, the Bradford assay was performed using the...
Bio-Rad Protein assay (Bio-Rad® Laboratories, Inc., EUA) and a calibration curve (1 and 10 μg/mL) of bovine serum albumin (BSA). The assays were made, at least, in duplicated.

3.2 Structural characterization of recombinant hMCAD proteins.

3.2.1 Determination of the oligomeric profile of recombinant hMCAD proteins by Size Exclusion Chromatography (SEC). In order to isolate the hMCAD tetrameric forms, the purified proteins were submitted to purification by SEC. In addition, SEC was also used for structural analysis of the recombinant hMCAD proteins by the determination of their oligomeric profile, in the presence and absence of FAD in the lysis buffer during purification. The SEC separates molecules according with their size, using a porous gel inserted into a column. Smaller molecules than the pore diameter are retained in the gel and eluted slowly when compared with larger molecules that are not retained in the gel and thus eluted faster. After a centrifugation step (10,000g/10 min/4°C) to eliminate potential solid aggregates, the resulting supernatant was injected in a protein purification system (Fast Protein Liquid Chromatography; AKTA®primeplus, GE Healthcare, EUA). The separation of the protein oligomeric forms was made in a molecular exclusion column HiLoad™ 16/60, Superdex™ 200 prep grade (cut-off of 10-600 kDa) with a constant flow of 0.7 mL/min of SEC buffer (20 mM Kpi, pH 7.4; 200 mM NaCl) as eluent at 4°C. The particles were detected by spectrophotometry in a UV spectrum (280 nm). The identification of the molecular mass of each peak corresponding to the different oligomeric forms of the recombinant hMCAD proteins was performed by comparing the retention volume of the obtained peaks with a calibration curve (Log Mr versus Vr/total column volume), prepared after the separation of a proteins mixture with known molecular weight (cytochrome c, 12.4 kDa; ribonuclease, 13.7 kDa; myoglobin, 17.6 kDa; ovalbumin, 43 kDa; bovine serum albumin (BSA), 66 kDa; alcohol dehydrogenase, 150 kDa; β-amylase, 200 kDa and apoferritin, 440 kDa). Blue dextran 2000 kDa was used to determine the void column i.e. the elution volume of molecules that are excluded from the column due to their large size, and tyrosine was used to determine the total column volume. The oligomeric profile of each recombinant protein is represented by the content (%) of aggregates, high molecular weight, tetrameric, dimeric and monomeric forms in respect to total protein. The fractions corresponding to the tetrameric forms (Mr 176.8 kDa) were collected and concentrated in a ViVaSpin 15R concentrator (Sartorius Stedim Biotech, GmbH, Goettingen, Germany) by centrifugation (3,000g/10 min/4°C) for immediate use or to be stored at -80°C or in liquid N₂.

3.2.2 Assessment of the conformation flexibility of recombinant hMCAD proteins by Limited Proteolysis with trypsin. This method is based in the susceptibility of each recombinant protein
to protease degradation. A reaction mixture with a total volume of 150 µL containing 0.3 mg/mL of purified protein (tetrameric form) purified in the presence and absence of FAD in the lysis buffer, and 0.012 mg/mL trypsin [1:25 trypsin: protein ratio (by mass)] was incubated at 37°C in SEC buffer. At timed intervals (0-60 min), aliquots of the reaction were collected and mixed with the soybean trypsin inhibitor at 1:1.5 protease to inhibitor ratio (by mass), in order to stop the proteolysis reaction, and denaturated at 95°C in the presence of SDS-PAGE loading buffer. The degradation profile was evaluated by SDS-PAGE at 12%. The changes in the electrophoretic profile of each protein were quantified by densitometry of the various obtained bands using the Image J® 1.43U NIH program (Image Processing and Analysis in Java). The half-life ($t_{1/2}$), i.e. the incubation time necessary for 50% proteolytic degradation, or residual protein (%) after 30 min incubation, was determined considering 100% of full-length protein at t₀.

3.2.3 Determination of the in vitro stability of recombinant hMCAD proteins by thermal denaturation using Differential Scanning Fluorimetry (DSF). This method was used to evaluate the stability of the recombinant hMCAD proteins through the determination of their melting temperature. It uses the SYPRO Orange fluorophore which emits fluorescence when binds to hydrophobic residues exposed upon protein thermal unfolding ($\lambda_{exc}$ 395 nm and $\lambda_{em}$ 450 nm). In RT-PCR plates, SEC buffer was firstly distributed per well followed by the protein (20 µg of isolated tetramers) and finally the Sypro Orange (5x). The RT-PCR plates were sealed with Optical-Quality Sealing Tape (BioRad® Laboratories, Inc., EUA), centrifuged (500g/1 min) and run in a Real-Time PCR system (C1000 Touch thermal cycler; BioRad). After an incubation period of 10 min at 20°C, the temperature was increased at a rate of 0.2°C/12 s ranging from 20°C to 90°C. The fluorescence intensity is plotted as a function of the temperature generating a sigmoidal curve. The inflection point of the curve represents the thermal induced protein melting point ($T_m$), i.e. the temperature at which the free energy of the native and the unfolded forms of the protein are equivalent, and was calculated using the Boltzmann equation

$$I(x) = I_0 + \frac{(I_1-I_0)}{1+e^{\frac{(T_m-T(x))}{\text{slope}}}} \quad (1).$$

The $T_m$ values obtained for the recombinant hMCAD proteins were further compared and conclusions on the stability of the p.K304E variant protein were inferred when the difference between the $T_m$ of the hMCADwt and the p.K304E variant was > 2°C.

3.2.4 Determination of the in vitro stability of recombinant hMCAD proteins by thermal denaturation using DSF in the presence of small molecules. The DSF can also be used as a powerful tool for the high throughput search for small molecular weight compounds with
pharmacological interest. As mentioned above, CCs and PCs are promising strategies for the rescue of misfolded proteins, acting as stabilizers of the defective but functional protein, promoting its correct folding and preventing its early degradation by the cellular quality control system.\textsuperscript{25,35}

The DSF was performed as described above (point 3.2.3) and the effect of some CCs, already studied in the laboratory for misfolded proteins associated with other inborn errors of metabolism was assessed: Glycerol (final concentration 2.5%), TMAO (final concentration 100 µM), Arginine (final concentration 100 µM) and Taurine (final concentration 100 µM). The FAD as the MCAD cofactor may be a potential PC\textsuperscript{32}; therefore its effect on the recombinant hMCAD proteins was evaluated as well (final concentration 5-80 µM). The $T_m$ values obtained for the recombinant hMCAD proteins in the presence and absence of these compounds were compared and a $T_m$ increase > 2°C in the presence of the tested molecule indicates its influence on the protein stability.

3.2.5 \textbf{Assessment of the aggregation profile and aggregation kinetics of recombinant hMCAD proteins by Dynamic Light Scattering (DLS).} The stability of the recombinant hMCAD proteins (tetrameric forms) was also assessed by following the thermal aggregation profiles of the recombinant proteins purified in the presence and absence of FAD in the lysis buffer. The DLS is a laser scattering technique based on the particles’ Brownian motion in a colloidal solution. The light scattering from the Brownian particles allows the measurement of the particles’ size by measuring the intensity of the light dispersion.\textsuperscript{36} Before DLS analysis the tetrameric fraction of each recombinant hMCAD protein was centrifuged (10.000g/15 min/4°C), diluted to 0.15 mg/mL with SEC buffer and filtered through 0.2 µm membranes to remove larger soluble aggregates. Three different assays were made at least in duplicate: thermal aggregation with ramping of the temperature from 20 to 70°C and collection of the data on particle size distribution and total scattering intensity; and the aggregation profile at 37°C and 42°C. The data were processed using the Zetasizer Nano DTS Software V7.10 and the GraphPad Prism software V6.00. The thermal aggregation was analyzed by a linear fit to the scattering intensity signal to determine the proteins’ aggregation point temperature ($T_{agg}$), i.e. the temperature at which the protein starts to aggregate. The aggregation kinetics were calculated based on a non-linear regression of the data obtained using the exponential growth curve described by Golub \textit{et al.}\textsuperscript{37}, modified to include the aggregation \textit{lag phase}

\begin{equation}
I = I_0 + (I_{lim} - I_0)\left[1 - e^{-k_1(t-t_0)}\right]
\end{equation}
where $I_{lim}$ is the intensity limit ($I$) when $t \to \infty$, $k_i$ is the first order kinetic constant and $t_0$ is the lag phase ($t_{lag}$). By comparing the $T_{agg}$ and $t_{lag}$ obtained for the hMCADwt and p.K304E variant was possible to infer on the structure and stability of the latter. For the $T_{agg}$ parameter a result was considered significantly different when the $T_{agg}$ obtained for both recombinant hMCAD proteins differ $>2^\circ C$.

3.3 Functional characterization of recombinant hMCAD proteins.

3.3.1 Enzymatic activity and determination of the kinetic parameters. The enzymatic activity of the recombinant hMCAD proteins, purified in the presence and absence of FAD in the lysis buffer, was measured spectrophotometrically in the presence of the substrate octanoyl-CoA as previously described. Briefly, the activity assay is based in the monitoring of the decrease at 600 nm of the absorbance of 2,6-dichlorophenolindophenol (DCPIP) in the presence of Phenazine methosulfate (PMS). The DCPIP/PMS constitutes an artificial redox pair, which mimics the reaction between the FAD and ETF protein after the substrate oxidation by MCAD. The PMS is an intermediate electron acceptor ("FAD") and the DCPIP is the final receptor ("ETF protein"). The kinetic parameters ($V_{max}$, $K_m$, $K_{cat}$, and catalytic efficiency) were also determined. The enzymatic activity of the recombinant hMCAD proteins was determined in the isolated tetrameric forms of the purified proteins. The initial step was the determination of the concentration of the DCPIP solution at 600 nm ($\varepsilon_{600} = 21.0 \text{ mM}^{-1}\text{.cm}^{-1}$) and of the octanoyl-CoA solution (substrate) at 259 nm ($\varepsilon_{259} = 15.9 \text{ mM}^{-1}\text{.cm}^{-1}$), using the Lambert-Beer’s law. The standard reaction mixture (20 mM KPi, pH 7.4; 1.5 mM PMS; 48 μM DCPIP; 33 μM octanoyl-CoA) was prepared in a final volume of 1 mL. The enzymatic assay began with the addition of 8 μg of the purified enzyme to the standard reaction mixture (1 min/25°C). A control assay was performed for each protein in the same conditions in the absence of substrate. The hMCAD activity was calculated as μmol of DCPIP reduced per mg of protein per min. The protein activity data was processed by subtracting the control data. The kinetic parameters $V_{max}$ and $K_m$ are derived from the Michaelis-Menten equation and were determined by assessing the enzymatic activity of hMCAD at increasing substrate concentrations, ranging from 0 μM to 100 μM (at least 12 different concentrations). The catalytic constant ($K_{cat}$) was calculated from the product of the $V_{max}$ value with the molecular mass (Mr) of the recombinant hMCAD monomer (44.2 kDa) and further used to determine the enzymes’ catalytic efficiency ($K_{cat}/K_m$).

3.3.2 Determination of the in vitro stability of recombinant hMCAD proteins by thermal enzymatic inactivation. To assess the effect of temperature on MCAD activity, thermal inactivation profiles of recombinant hMCAD proteins were obtained. For that purpose, 8 μg of
the SEC purified tetrameric fraction, were pre-incubated, in 20 mM Kpi, pH 7.4, in a final volume of 100 μL, at different temperatures (25, 30, 35, 37, 39, 43, 45, 47, 50, 55 and 60 °C) for 10 min and cooled on ice for the same period. Enzymatic activity was determined as described above using 33 μM octanoyl-CoA as substrate. The enzymatic activity obtained after incubation at 25°C was considered 100% and the relative enzyme activity was determined for the rest of the tested temperature points. The relative enzymatic activity (%) was further plotted against incubation temperature and the thermal midpoint of enzyme inactivation (T_{1/2}), i.e. the temperature at which is reached 50% of relative enzymatic activity, was calculated for both hMCADwt and p.K304E variant proteins and compared. The results allowed to infer the stability of the p.K304E variant protein when the difference between the T_{1/2} obtained for the wild-type and the p.K304E variant was > 2°C.

4. Computational methodology

4.1 Molecular dynamics (MD) simulations – a brief overview

In the past years, computational techniques as molecular dynamics (MD) or docking have been successfully applied to the study of many diseases, providing new insights to better understand structural and functional features of proteins. Moreover, such techniques may complement experimental data by giving additional information that allows new experimental approaches or by selecting compounds that can be further tested in vitro and in vivo. Currently, as all major work developed on MCADD is experimental, computational methods can be applied to provide a deeper knowledge about MCAD variants and to clarify the molecular mechanisms underlying MCADD, aiming for the discovery and development of compounds that may be potential candidates for the treatment of this disease.

**General concepts.** Molecular Dynamics (MD) is a computer technique widely used to understand the structure and dynamics behavior of a system at the atomic level. For this reason, MD simulations have been a powerful tool and a strong complement to experiments, especially in studies of protein-ligand interactions or protein folding/unfolding processes. By this method, the time dependent behavior of a molecular system is calculated from its atoms and/or molecules by solving the Newton’s motion equations which relates the forces acting on each atom with its positions during the simulation. A global flow scheme for molecular dynamics simulations is given by Figure 8, with each run requiring an initial coordinate set and, optionally, the initial velocities of all particles.
In MD simulations, the total energy of the system can be divided into kinetic (K) and potential energies (V) as described by the Hamiltonian equation (\( \mathcal{H} \)).

\[
\mathcal{H}(p, rpm, s) = K(p, m) + V(r, s) \quad (3)
\]

The first term of equation represents the kinetic energy (K) of the system and is a function of the momentum (p) and the mass (m) of the atoms, being represented as

\[
K(p, m) = \frac{1}{2} \sum_{i=1}^{N} m_i v_i^2 \quad (4)
\]

where \( m_i \) and \( v_i \) are the mass and velocity of the atom \( i \), respectively.

The second term of the equation represents the total potential energy (V) of the system as a function of the atom coordinates (r) and the force field parameters (s). It can be calculated as
the sum of bonded (bond) and non-bonded (nbond) interactions and is represented by the equation

\[ V(r, s) = V^{\text{bond}}(r, s) + V^{\text{nbond}}(r, s) \]  

(5)

Accordingly, bonded interactions are used to describe the relations existing between two or more atoms that are directly connected by rigid bonds. Thus, the energetic terms describing bond lengths (stretching), angles (bending) and torsions (dihedrals) can be given by

\[ V^{\text{bond}} = \sum_{\text{bond}} k_d \left( d - d_0 \right)^2 + \sum_{\text{ang}} k_\theta \left( \theta - \theta_0 \right)^2 + \sum_{\text{tors}} k_\phi \left( 1 + \cos(n\phi - \phi_0) \right) + \sum_{\text{imp}} k_\psi \left( \varphi - \varphi_0 \right)^2 \]  

(6)

While bond lengths and angles are described by harmonic potentials, dihedrals can be calculated by periodic, Ryckaert-Bellemans or Fourier functions. Improper dihedrals are also used to maintain a given molecular structure planar (e.g. aromatic groups) or to maintain the molecules’ chirality (a mirror image).41,43–46

In the remaining term (eq. 3), non-bonded (nbond) interactions take into account the repulsion, dispersion and Coulomb terms. While repulsion and dispersion terms are often combined in the Lennard-Jones (LJ) potential, the Coulomb (C) term includes all total or partially charged atoms. Both terms can by combined in the equation

\[ V^{\text{nbond}} = \sum_{LJ} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{C} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \]  

(7)

Since the first MD simulation of biomolecules and water, performed in 1970’s, multiple force fields, have been developed, with increasing accuracy for proteins and nucleic acids.38 The term ‘force field’ describes a set of equations and parameters used to calculate forces and potential energies generated along the simulation, regarding interactions between atoms and molecules within the system. The force field is chosen depending on the system to simulate, biological structure, atom types and, often, on the available computational power.41,43–46 The most common biomolecular force fields available are AMBER, CHARMM, OPLS and GROMOS, although coarse grained force fields as MARTINI47 are increasingly being used to simulate very large systems as lipid bilayers or micelles. While the three first are “all atom” force fields, i.e. each atom within the system is represented including the nonpolar hydrogens (aliphatic CHn), the GROMOS is a typical example of an “united atom” force field where the aliphatic carbon atom and the nonpolar hydrogens bonded to it are considered as a single unit (Fig. 9).38,41,44–46,48
Figure 9: The substrate 3-Thioctanoyl-CoA (PDB ID: 1UDY) represented in all-atom (A) or united-atom (B) force-fields. The nonpolar hydrogens (aliphatic CH\textsubscript{n} - light gray) are considered in an all-atom force field but not in a united-atom force field. Figures were created with MOE v2013.08 from PDB ID: 1UDY.

All these force fields have parameters to successfully simulate proteins and nucleic acids, but cofactors, lipids or other small molecules, required in studies of protein-peptide, protein-ligand or protein-lipid interactions, are frequently absent. Nonetheless, parameters for such molecules (bonded and non-bonded) can be obtained by using web servers that can generate topologies accordingly with the chosen force field.\textsuperscript{49,50} The PRODRG\textsuperscript{51} or the Automated Topology Builder (ATB) and Repository\textsuperscript{49,52} are examples of web servers designed to create small molecules topologies able to be used together with the selected force field.\textsuperscript{50}

**Methodology to setting up a MD simulation.** Before starting a MD simulation it is necessary to construct the system to be simulated.

**Starting the structures.** The initial structures to simulate may be obtained through crystallographic techniques, nuclear magnetic resonance (NMR) spectroscopy or homology models. The most accurate structures are resolved through crystallography methods, providing a higher structure resolution when compared with those resolved by NMR spectroscopy or resultant from the homology models.\textsuperscript{54} The homology models are useful when there is no crystallographic structure available for the target protein but there is a homologue (called templates). The accuracy of the predicted structure depends on the percentage of homology between the target and the template. The model may be considered accurate for studies of drug discovery if the sequence identity (SI) between the target and its homologues is over than 60% or contribute to design mutagenesis experiments when the homology is between 25 and 50%. The model is considered merely speculative when the SI between the target protein and the template is between 10 and 25%, which may introduce serious errors in the protein secondary structure.\textsuperscript{54}

**Defining the box and solvating the system.** The definition of a suitable simulation box is important because: i) it defines a system comprising a finite number of atoms; ii) it prevents the protein to 'see' its periodic image across the boundary of the box by applying periodic boundary
conditions (see below); and iii) it allows the optimization of the total number of atoms present in the system in order to improve computational performance. The most common box shapes are the triclinic, the cubic, the rhombic dodecahedral and the truncated octahedral (Fig. 10). The latter are smallest that the cubic box and are more suitable to simulate spherical macromolecules in solution.\textsuperscript{41,43–46}

![Figure 10: Representation of rhombic dodecahedral (left) and a truncated octahedral (right) simulation boxes when compared with the cubic box [adapted from Gromacs User Manual version 4.5.6].](image)

**Periodic Boundary conditions (PBC).** Most of the MD simulations use periodic boundary conditions (PBC) to avoid artifacts caused by the finite size of the simulation box, i.e. the box is surrounded by periodic images of itself and if one atom leaves the simulation box its periodic image enters in the opposite side of the box simulating an infinite system (Fig. 11).\textsuperscript{39,41,44–46}

![Figure 11: Schematic representation of periodic boundary conditions: one atom that leaves the simulation box, enters again by the opposite side [reprinted from http://isaacs.sourceforge.net](image)

**Solvating the system.** The last step before the simulation is to fill the box with water and eventually add counter ions (e.g. sodium and chlorine) to compensate the system final charge.\textsuperscript{41,43–46}

**Energy minimization.** The protein solvation may introduce large forces in the system whereby an energy minimization run is recommended to relax the protein and avoid distortions in the structure.\textsuperscript{41,43–46}

**Equilibration and production runs.** The equilibration runs aim to heat the system to the target temperature, generate atomic velocities (\textit{NVT} ensemble) and stabilize the system’s pressure (\textit{NpT} ensemble). The \textit{NVT} ensemble is the first equilibration run in which the number of atoms
(N), Volume (V) and temperature (T) of the system are constant, followed by the \( NpT \) ensemble wherein the number of atoms (N), pressure (p) and Temperature (T) are kept constant. Position-restraints of heavy atoms are applied in both equilibration steps to avoid protein distortions.\(^{41,43-46}\) Finally, the production run is similar to the previous simulations, being again employed an \( NpT \) ensemble in which the number of atoms (N), pressure (p) and temperature (T) are kept constant. Usually, these simulations are longer than the equilibration runs, with no positions restraint on heavy atoms.\(^ {41,43-46}\)

**Simulation parameters – Algorithms.** A set of algorithms are applied to control the simulation parameters. The bond lengths can be fixed applying constrained algorithms such LINCS\(^ {55,56}\) (default) or SHAKE.\(^ {57}\) To calculate the short-range electrostatic and short-range van der Walls interactions, cut-offs are applied but for the long-range electrostatic interactions a particle mesh Ewald (PME) method is used instead.\(^ {58,59}\) The weak-coupling scheme of Berendsen\(^ {60}\), the velocity-rescale (V-rescale)\(^ {61}\) and the extended ensemble Nosé-Hoover\(^ {62,63}\) are the thermostats used to control the system’s temperature in the MD simulations. The weak-coupling scheme of Berendsen\(^ {60}\) and the V-rescale (modified Berendsen thermostat)\(^ {61}\) are used when an efficient control of temperature and a uniform system heating are required (\( NVT \)). The Nosé-Hoover thermostat\(^ {62,63}\) allows a better control of the temperature oscillations around an average value of the target temperature, being more appropriate for long simulations (production runs). In the early stages of the equilibrium, the system’s pressure can be controlled by a weak-coupling scheme of Berendsen barostat being more resistant to the pressure variations, more suitable for these stages. In the MD run (production run) the Parrinello-Rahman barostat\(^ {27}\) is the most used, being more suitable to resist to pressure oscillations throughout an extended simulation.

### 4.2 Methodology

**Molecular dynamics (MD) simulations of MCAD wt proteins and p.K304E variants.**

**4.2.1 Initial structures.** The starting coordinates of pMCADwt were obtained from the crystallographic structure of the *Sus scrofa* (pig) MCAD protein (PDB ID: 1UDY; 2.4 Å resolution),\(^ 6\) in complex with FAD cofactor and the lipid substrate analogue 3-thiaoctanoyl-CoA (SCOA). In order to relate the computational with the experimental data and also validate the MD method, the hMCADwt protein was obtained by reverting the crystallographic structure of the E376G/T255E double mutant of the hMCAD protein (PDB ID: 1EGC; 2.6 Å resolution),\(^ {64}\) in complex with FAD and the lipid substrate octanoyl-CoA (CCOA) into the hMCADwt. In a following step, both pMCAD and hMCAD wild-type proteins were used as templates to generate the
respective p.K304E variants (p.K304E/pMCAD and p.K304E/hMCAD, respectively). The mutants were constructed by substituting the lysine (K) by a glutamic acid (E) residue at position 304 using the MOE 2013.08\textsuperscript{65} software.

4.2.2 Preparation of input data. The water molecules within both MCAD crystallographic structures were removed and the protonation of the initial structures followed at $T = 310$ K (37°C) and pH = 7.4 (see Experimental Section) using the Protonate 3D module available in MOE 2013.08\textsuperscript{65} software. In order to use these structures with the GROMOS96\textsuperscript{42,66} 53a6\textsuperscript{38} force field, all the non-polar hydrogens were removed, also with MOE 2013.08\textsuperscript{65} software.

The protein topologies were generated according with the GROMOS96\textsuperscript{42,66} 53a6\textsuperscript{38} force field by the \textit{pdb2gmx} module available in GROMACS 4.6.3\textsuperscript{41,44–46} package. The topologies of the FAD, SCOA and CCOA were built using the Automated Topology Builder and Repository (ATB)\textsuperscript{49,52} web server according with the selected force field.

Three systems were built for both MCAD wt proteins and p.K304E variants: APO (no substrate, no FAD), FAD (no substrate, with FAD), and LIPID (with substrate and FAD) systems. To avoid interactions with the respective periodic images, the protein was kept at the center of a cubic simulation box at a distance of 1.0 nm from the box edge. Then, all systems for pMCADwt protein were solvated with 66462 (APO system), 66354 (FAD system) and 66208 (LIPID system) water molecules with Simple Point Charge (SPC)\textsuperscript{67} parameterization. In all systems periodic boundary conditions (PBC) were applied to all dimensions. As both the FAD cofactors and the acyl-CoA substrates are negatively charged (charge of -2 and -3 respectively), the system’s total charge was neutralized by replacing an adequate number of water molecules with 1 (APO system), 9 (FAD system) or 25 (LIPID system) sodium ions. Overall, the number of atoms in each system was 214676 (APO system), 214620 (FAD system) and 214458 (LIPID system) respectively.

The systems for hMCADwt were solvated with 64182 (APO system), 64079 (FAD system) and 63941 (LIPID system) SPC water molecules after centering the protein at a cubic simulation box, maintaining a distance of 1.0 nm from the box edge to allow periodic boundary conditions. The system’s total charge was neutralized with 4 (APO system), 12 (FAD system) and 28 (LIPID system) sodium ions by replacing the same amount of water molecules. Overall, the number of atoms in each system was 207910 (APO system), 207869 (FAD system) and 207731 (LIPID system) respectively.

The systems comprising the p.K304E variants were built as previously described. In the p.K304E/pMCAD variant, 66454 (APO system), 66346 (FAD system) and 66200 (LIPID system) SPC water molecules were used to solvate the box while the system’s total charge was neutralized with 9 (APO system), 17 (FAD system) and 33 (LIPID system) sodium ions. At the end,
the number of atoms in each system was 214.648 (APO system), 214.592 (FAD system) and 214430 (LIPID system), respectively.

For the p.K304E/hMCAD variant, 64.179 (APO system), 64.076 (FAD system) and 63.938 (LIPID system) SPC water molecules were used. To neutralize all systems charge, 12 (APO system), 20 (FAD system) and 36 (LIPID system) sodium ions were added to the system. Overall, the number of atoms in each system was 207.897 (APO system), 207.856 (FAD system) and 207718 (LIPID system), respectively.

After the systems preparation, and to remove clashes between the atoms, energy minimization runs were performed using the steepest descent algorithm. The minimization stopped when the maximum force acting on an atom was < 1000 kJ.mol⁻¹ nm⁻¹.

4.2.3 Equilibration and production runs: The MD simulations were performed using the GROMACS 4.6.3⁴¹,⁴⁴–⁴⁶ package and the GROMOS96⁴²,⁶⁶ 53a6³⁸ force field. The systems were previously equilibrated under a NVT run (100 ps) followed by 2 ns of an NpT ensemble. The V-rescale weak coupling method⁶¹ was used to heat the systems at 310 K (37°C) and the Nosé-Hoover thermostat⁶²,⁶³/Parrinello-Rahman barostat⁶⁸ were used to generate a rigorous NpT ensemble (T = 310 K, P = 1 bar, respectively). Positional restraints on heavy atoms were applied in NVT and NPT (1 ns) runs followed by an unrestrained NPT (1ns) on heavy atoms of the FAD and substrates (FAD and LIPID systems). Bond lengths were constrained with LINCS⁵⁵,⁵⁶ and SETTLE⁵⁷ (water molecules) algorithms. A cut-off of 1.0 nm was used to compute the short-range electrostatic and the short-range van der Waals interactions. Long-range electrostatic interactions were calculated with the Particle Mesh Ewald (PME) method⁵⁸,⁵⁹. A 100 ns of fully unrestrained production runs were performed under the same conditions, using the Nosé-Hoover thermostat⁶²,⁶³/Parrinello-Rahman barostat⁶⁸ to control the temperature and the pressure, respectively.

4.2.4 Analysis of the MD trajectories: Several modules available in GROMACS⁴¹,⁴⁴–⁴⁶ package were used to analyze the MCAD structural stability and its dynamic behavior. For each system, the root mean square deviation (RMSD) and radius of gyration (Rg) of the Cα atoms in respect to the crystallographic structure were calculated, using the equilibrated structure (last 40 ns), by the modules g_rms and g_gyrate. The evolution of the secondary structure in function of time was calculated by DSSP package⁶⁹,⁷⁰ using the do_dssp module. To assess the protein flexibility the root mean square fluctuation (RMSF) and the β-factors of the protein’s residues were calculated by the g_rmsf module. As the experimental results point to a reduced enzymatic activity for the p.K304E variant, the structure and dynamics of the catalytic pocket were assessed to investigate how the mutation may affect the MCAD function. The volumes of the catalytic
pocket were estimated along the simulation using the VOIDOO\textsuperscript{71} software. The amino acids lining the catalytic pocket were determined using the EPOS\textsuperscript{BP,72,73} and the FAD and substrate binding regions were assessed by the \texttt{g\_contacts}\textsuperscript{74} module added to GROMACS. To estimate the free energies of binding for the FAD and the substrate, the \texttt{g\_mmpbsa} software\textsuperscript{75} was used to solve the Poisson-Boltzmann equation numerically by linking GROMACS with APBS.\textsuperscript{76} The distances between FAD, substrate and protein’s residues in the catalytic pocket were calculated with \texttt{g\_dist} and \texttt{g\_mindist} modules in GROMACS. Hydrogen bonds were calculated with \texttt{g\_hbond}\textsuperscript{77} and the side-chain conformations of protein’s residues were analyzed by the \texttt{g\_angle} module. Visual inspections was performed with VMD 1.9.1\textsuperscript{78} and MOE 2013.08\textsuperscript{65} software.

5. Experimental results

5.1 Expression and purification of recombinant hMCAD proteins: yield and purity grade. The constructs pETMCAD\textsubscript{dwt} and pETMCAD\textsubscript{dK304E} were previously prepared for the heterologous expression of the wt and p.K304E variant of hMCAD, respectively, in \textit{E.coli} BL21(DE3) cells. Protein expression was induced with 0.5 mM of IPTG, for 21 hours at 27°C. The recombinant hMCAD proteins were further purified by IMAC with an imidazole gradient (20 mM to 500 mM) in order to elute the His-tagged recombinant proteins. The drastic procedures that a recombinant protein undertakes during its purification such as cell lysis and differential centrifugation, used to separate the soluble and insoluble fractions, as well as the IMAC purification steps may affect the binding of FAD to the protein. In order to assess the true impact of the isolation procedures on the stability and function of the recombinant proteins, the addition of 0.1 mg/ml FAD to the lysis buffer was tested. The results obtained were compared with the recombinant protein prepared in the absence of the cofactor.\textsuperscript{32} To identify the fractions where the His-tagged proteins were eluted and evaluate the yield and purity grade of the recombinant proteins, the IMAC eluted fractions were analyzed by SDS-PAGE at 12 %.

The electrophoretic profiles obtained after purification of 0.5 L bacterial culture (Fig. 12A and 12B), revealed that both hMCAD\textsubscript{wt} and p.K304E variant were eluted mostly in the first fraction of 500 mM of Imidazole. Both recombinant proteins showed high expression levels. The presence of FAD in the lysis buffer during the purification, and solely by observation of the electrophoretic profiles, apparently increased the expression levels of both the hMCAD\textsubscript{wt} and the pK304E variant.

Moreover, the densitometric analysis of the SDS-PAGE gels (ImageJ 1.49k; http://imagej.nih.gov/ij/) revealed that both purified recombinant hMCAD proteins were expressed with high purity grade (Table 1).
Figure 12: Electrophoretic profile obtained by SDS-PAGE at 12% for the fractions eluted during IMAC purification (second fraction of 40 mM and the first three fractions of 500 mM imidazole) from 0.5 L culture for the recombinant proteins hMCADwt and p.K304E variant, purified in the absence of FAD (A) and presence of FAD (B) in cell lysis buffer, after expression in E.coli BL21(DE3) cells, at 27°C, for 21h, in the presence of 1mM IPTG. Molecular weight marker NZYcolour Protein Marker II (NZYTech®, Lda., Portugal).

Nevertheless, and as shown in the Table 1, a significant lower purity grade was observed for the p.K304E variant when compared with the wt, being independent of the addition of FAD in the lysis buffer (p<0.05). Furthermore, the presence of the cofactor did not contributed to increase the purity grade of the wt, however, affected negatively, the purity grade of the p.K304E variant. The fractions with higher amount of protein (fractions 500₁ and 500₂) were combined and further purified by dialysis (cut-off of 12-14 kDa) or gel filtration (exclusion limit 5 kDa) in order to remove the imidazole used in the elution process. In a following step, the purified recombinant proteins were quantified by the Bradford method. Based in the purity grade previously determined and in the protein quantification, it was possible to calculate the yield obtained for each recombinant protein (Table 1). Similar yield values were observed for both hMCADwt and p.K304E forms in the absence of FAD in the lysis buffer, nonetheless, the presence of the cofactor affected positively (although not significantly) the yield of the hMCADwt but not of the p.K304E variant.
Table 1: Yield and purity grade of the recombinant hMCADwt and p.K304E variant, purified in the absence (-FAD) and presence of FAD (+FAD) in the lysis buffer.

<table>
<thead>
<tr>
<th></th>
<th>hMCADwt</th>
<th>p.K304E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>- FAD</strong></td>
<td>Yield (mg/L)</td>
<td>4.6 ± 3.2 (n=21)</td>
</tr>
<tr>
<td></td>
<td>Purity grade (%)</td>
<td>87.0 ± 9.1 (n=25)</td>
</tr>
<tr>
<td><strong>+ FAD</strong></td>
<td>Yield (mg/L)</td>
<td>5.9 ± 4.6 (n=10)</td>
</tr>
<tr>
<td></td>
<td>Purity grade (%)</td>
<td>87.0 ± 7.0 (n=11)</td>
</tr>
</tbody>
</table>

The data represents the mean ± SD; n - number of independent determinations performed in duplicate.

5.2 Structural characterization of recombinant hMCAD proteins.

5.2.1 Determination of the oligomeric profile of recombinant hMCAD proteins by Size Exclusion Chromatography (SEC). After the purification processes, the purified recombinant hMCAD proteins were submitted to a centrifugation step to eliminate potential solid aggregates of the protein solution and further injected in a protein purification system. The identification of the molecular mass of each peak corresponding to the different oligomeric forms of the recombinant proteins was performed by comparing the retention volume of the obtained peaks with a calibration curve (Log Mr versus Vr/total column volume), prepared after the separation of a mixture of proteins with known molecular weight (point 3.2.1 of the methods chapter). The oligomeric profile of each recombinant protein is represented by the content (%) of each oligomeric form in respect to total protein and is given by the ratio between the respective peak area and the sum of all the peaks’ areas.

The analysis of the SEC profiles for both hMCADwt and p.K304E variant revealed that the tetramers are the major oligomeric form present in both recombinant proteins (Fig. 13).

Figure 13: Representative elution profiles of the recombinant hMCADwt and p.K304E variant, purified in the absence (blue lines) and presence (green lines) of the FAD cofactor in the lysis buffer. The different oligomeric species identified are indicated: aggregates (A), Mr > 1000 kDa; high molecular weight forms (H), Mr > 400 kDa; tetramers (T), Mr ~170 kDa; dimers (D), Mr ~90 kDa, and monomers (M), Mr ~44 kDa. The apparent molecular weight (Mr) was estimated from the elution volume of proteins of known molecular weight. The dotted lines indicate the elution volume for the tetramers (left line) and the dimers (right line) of recombinant hMCAD proteins.
Nevertheless, in the p.K304E variant, a drastic reduction of the tetrameric fraction was observed in respect to the hMCADwt protein (Table 2), mainly resultant from the increase in the content of aggregates, HMW forms, dimers and monomers.

The presence of the cofactor during protein purification led to a slight increase of the tetrameric fraction in the hMCADwt, however an increase of aggregates and monomeric forms were also observed. In contrast, the presence of FAD improved the oligomeric profile of the p.K304E variant, by increasing the tetrameric fraction and reducing the dimers and monomers’ content, although not significantly due to wide variability of the obtained data.

Table 2: Distribution of the oligomeric forms obtained by SEC of the recombinant hMCADwt and p.K304E mutant forms, in the absence (A) and presence (B) of FAD in the lysis buffer during purification.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>Oligomeric forms (%)</th>
<th></th>
<th>B</th>
<th>Oligomeric forms (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(without FAD)</td>
<td></td>
<td></td>
<td>(with FAD)</td>
</tr>
<tr>
<td>hMCAD</td>
<td>wt</td>
<td>n</td>
<td>Aggregates</td>
<td>HMW</td>
<td>Tetramers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>1.57±1.63</td>
<td>0.85±1.25</td>
<td>84.2±5.53</td>
</tr>
<tr>
<td></td>
<td>p.K304E</td>
<td>4</td>
<td>7.81±3.50</td>
<td></td>
<td>6.26±5.09</td>
</tr>
</tbody>
</table>

Statistical significance determined using the Student unpaired t test and comparing the variant and the wt forms of hMCAD (* p<0.05; ** p<0.005); n – number of independent determinations performed in duplicate. The data represents the mean ± SD.

**5.2.2 Assessment of the conformation flexibility the recombinant hMCAD proteins by Limited Proteolysis with trypsin.** The reaction mixture containing the tetrameric forms of the recombinant proteins and trypsin [1:25 trypsin: protein ratio (by mass)] was incubated at 37°C in SEC buffer for 60 min. The trypsin reaction was stopped at different time points by adding the soybean trypsin inhibitor [1:1.5 trypsin: protein ratio (by mass)]. The samples were denatured and further applied in a SDS-PAGE gel at 12% in order to assess the degradation profile (Fig. 14A). The densitometric analysis of each protein band (software ImageJ™ 1.43U NIH) allowed the determination of the degradation rate of the recombinant proteins (Fig. 14B). Considering that at t = 0 min, the protein is in its full-length (100%), both hMCADwt and p.K304E variant did not reached 50% of degradation during the maximum reaction time and thus, the half-life (t1/2), could not be determined. Alternatively, the residual protein (%) after 50% of the reaction time (i.e. 30 min) was measured.
Figure 14: Limited proteolysis by trypsin. (A) Degradation profile of the recombinant protein hMCADwt and the p.K304E variant, obtained by limited proteolysis by trypsin at 37°C, for 60 min; M – NZYTech LMW Protein Marker (NZYTech®, Lda. Portugal). (B) Representation of the degradation rate (%) determined by the non-linear fit of the mean ± SD of the percentage of full-length protein in respect to t0 (100%) of two to five and two to four independent experiments for hMCADwt and p.K304E, respectively.

The data obtained from the degradation profiles after proteolysis by trypsin showed that the residual amount of protein (%) at 30 min was 74.8% for the hMCADwt, indicating higher resistance to the protease action when compared with the p.K304E variant with 56.1% of the residual protein at 30 min.

5.2.3 Determination of the in vitro stability of recombinant hMCAD proteins by thermal denaturation using Differential Scanning Fluorimetry (DSF). In Real-Time PCR plates 20 μg of isolated tetramers were mixed with the SYPRO Orange fluorophore (5x) in a final volume of 50 μL and further subjected to thermal denaturation. The SYPRO Orange is a dye that emits fluorescence when binds to hydrophobic regions of a protein exposed upon protein thermal unfolding. The fluorescence intensity is plotted as a function of the temperature generating a
sigmoidal curve representing its inflection point the protein’s melting point ($T_m$). A difference in the $T_m$ of different proteins or in the $T_m$ obtained with the same protein under different conditions is considered to be significant when it is higher than 2 °C.

The thermal protein unfolding profiles obtained for the hMCADwt and the p.K304E variant are represented in Figure 15A and the correspondent derivative shown in Figure 15B for more clear comparison. Interestingly, for both proteins two transitions were observed, being this more pronounced in the case of p.K304E (Fig. 15B and Table 3).

**Figure 15:** Thermal stress induced denaturation monitored by differential scanning fluorimetry probing global protein unfolding from 20°C to 70°C with a rate of 0.2°C/12 s. Unfolding profiles of hMCADwt and p.K304E variant were determined as changes in the SYPRO Orange fluorescence ($\lambda_{exc}$ 395 nm and $\lambda_{em}$ 450 nm). In Graph A data are given as non-linear fit of the mean of four and two independent experiments, respectively, for the wt and the mutant protein. In Graph B the corresponding derivative results are represented. Both proteins revealed a biphasic unfolding mechanism.

<table>
<thead>
<tr>
<th>Table 3: Thermal induced protein melting point ($T_m$) of the recombinant wt and p.K304E variant.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$T_{m1}$ (°C)</strong></td>
</tr>
<tr>
<td>hMCAD wt</td>
</tr>
<tr>
<td>p.K304E</td>
</tr>
</tbody>
</table>

The data represents the mean ± SD; n - number of independent determinations.
The results showed that these recombinant proteins, present at least two different domains with different thermal stability. A significant difference in $T_m$ ($P<0.005$) was observed between the hMCADwt and the p.K304E variant but only in the first transition. In respect to the second transition no difference was observed in the calculated $T_m$ values for hMCADwt and p.K304E variant which may be related with a protein region that is not destabilized in the mutant protein.

5.2.4 Determination of the in vitro stability of recombinant hMCAD proteins by thermal denaturation using DSF in the presence of chaperones. The DSF in the presence of small molecules was performed as described in the methods chapter (point 3.2.3). As mentioned, the CCs and PCs are promising strategies to stabilize and rescue misfolded proteins resultant from missense mutations. In this work some CCs, already studied in the laboratory for the treatment of other inborn errors of metabolism, were applied to assess their effects on hMCAD stability. Thus, Glycerol (final concentration 2.5%), TMAO (final concentration 100 μM), Arginine (final concentration 100 μM) and Taurine (final concentration 100 μM) were added to the recombinant proteins and the thermal denaturation profiles further assessed. As the cofactor of hMCAD, FAD may be a potential PC, therefore its effect on hMCAD stability was also evaluated (final concentration 5-80 μM). An increment higher than 2°C in the $T_m$ values of the protein in the presence of the tested molecules indicates their potential impact upon the protein stability.

As shown in Figure 16 and Table 4, glycerol led to a significant increase in the $T_{m1}$ values of both hMCADwt ($p<0.05$) and p.K304E variant ($p<0.01$). Nevertheless while this increase was > 5.9°C in the case of the variant protein, it was only equivalent to a difference in $T_{m1}$ in respect to absence of chaperone of less than 2°C in the wt protein, and thus not relevant in terms of the increment in the stability of the hMCADwt.

The presence of TMAO did not change the $T_{m1}$ values of the hMCADwt, nevertheless, led to an increase in the $T_{m1}$ value of the p.K304E variant. Despite this increment, the difference in the $T_{m1}$ values for the p.K304E variant in the absence and presence of the chaperone is not statistical significant.

The addition of 100 μM of the amino acids Arginine and Taurine did not reveal to be stabilizing to none of the proteins. Furthermore, the data gathered also indicate that the cofactor is the only molecule that really increased significantly the $T_{m1}$ values of both hMCADwt and p.K304E proteins. As shown in Figure 17A and Table 5, the hMCADwt protein increased its $T_{m1}>14.5^\circ$C ($P<0.0001$) in the presence of 80 μM of FAD.
The same dramatic effect was even more pronounced for the p.K304E mutant which reached with 80 \( \mu M \) FAD the \( T_{m1} \) observed for the wt protein, with an increment of > 17.5°C (P<0.0005). The presence of the cofactor also led to an abolishment of the second transition (\( T_{m2} \)) which indicates that it must be a domain that is not destabilized by the mutation.

**Table 4:** Melting temperature (\( T_{m} \)) determined by differential scanning fluorimetry (DSF) of tetrameric hMCADwt and p.K304E variant in the absence and presence of chemical and pharmacological chaperones.

<table>
<thead>
<tr>
<th>Chaperones</th>
<th>hMCADwt</th>
<th>p.K304E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( T_{m1} (°C) )</td>
<td>( T_{m2} (°C) )</td>
</tr>
<tr>
<td>-</td>
<td>45.8 ± 0.6 (n=4)</td>
<td>50.3 ± 0.3 (n=4)</td>
</tr>
<tr>
<td>FAD (80 ( \mu M ))</td>
<td>60.4 ± 0.3 (n=4)</td>
<td>-</td>
</tr>
<tr>
<td>Glycerol (2.5%)</td>
<td>47.1 ± 0.9 (n=6)</td>
<td>52.2 ± 1.7 (n=7)</td>
</tr>
<tr>
<td>TMAO (100 ( \mu M ))</td>
<td>45.8 ± 0.3 (n=5)</td>
<td>51.5 ± 0.5 (n=6)</td>
</tr>
<tr>
<td>Arg (100 ( \mu M ))</td>
<td>45.8 ± 0.6 (n=4)</td>
<td>51.3 ± 0.5 (n=4)</td>
</tr>
<tr>
<td>Tau (100 ( \mu M ))</td>
<td>45.9 ± 1.3 (n=4)</td>
<td>51.7 ± 2.3 (n=7)</td>
</tr>
</tbody>
</table>

The statistical significance determined using the unpaired Student t test and comparing the wt and the variant protein (\# p<0.05; ## p<0.005) and the presence and absence of chaperones (* p<0.05; ** p<0.01; *** p<0.0005; 4* p<0.0001); n – number of independent determinations performed in duplicate. The data represents the mean ± SD.

**Figure 16:** Thermal stress induced denaturation monitored by differential scanning fluorimetry (DSF) probing global protein unfolding from 20°C to 70°C with a rate of 0.2°C/12 s. Unfolding profiles of recombinant hMCADwt and p.K304E variant were determined as changes in SYPRO Orange fluorescence (\( \lambda_{exc} \) 395 nm and \( \lambda_{em} \) 450 nm). Data are given as non-linear fit of the mean of four and two independent experiments, respectively, for the wt and the variant proteins in the presence and absence of 2.5% Glycerol and 100 \( \mu M \) of TMAO, Arginine and Taurine.
Figure 17: Thermal stress induced denaturation monitored by differential scanning fluorimetry (DSF) probing global protein unfolding from 20°C to 70°C with a rate of 0.2°C/12 s. Unfolding profiles of recombinant hMCADwt and p.K304E variant were determined as changes in SYPRO Orange fluorescence ($\lambda_{\text{exc}}$ 395 nm and $\lambda_{\text{em}}$ 450 nm). In Graph A data are given as non-linear fit of the mean of four and two independent experiments, respectively, for the wt and the variant proteins in the presence and absence of 80 µM of FAD, the hMCAD cofactor. In Graph B data is given as the non-linear fit data of one experiment performed in duplicate where thermal denaturation was monitored in hMCADwt and p.K304E by DSF in the presence of increasing concentrations of FAD (0 – 80 µM).

For a more clear comprehension of the hMCADwt and the p.K304E variant thermal unfolding profiles, and the respective melting transitions ($T_{m1}$ and $T_{m2}$), a representation of the combined derivative results of the fluorescence intensity obtained in the absence and in the presence of the small molecules tested is also provided (Fig. 18).

The changes in $T_{m1}$ were also analyzed for hMCADwt and p.K304E proteins with increasing concentrations of FAD (Fig. 17B). From the $T_{m1}$ values obtained in function of the cofactor concentrations it was possible to calculate the $K_d$ constant for FAD. The data gathered suggest that the p.K304E variant has higher affinity for the cofactor (1.9 ± 0.3 µM) when compared with the wt protein ($K_d = 5.8 \pm 1.1$ µM) likely due to a lack of FAD in the mutant protein while the wt protein has already almost a full occupancy of the FAD binding sites. The intrinsic changes in the structure of the p.K304E variant may compromise the FAD binding to protein and a loss of the cofactor may occur during the isolation and purification steps.
It would be interesting to determine the $K_d$ value for FAD with proteins to which FAD was added to the lysis buffer or proteins expressed in a riboflavin (FAD precursor) supplemented culture medium. Curiously, the observation of Figure 17B leads to the suggestion that despite a loss of FAD which may occur during the purification steps of the p.K304E variant, an excess of the cofactor may destabilize the protein's conformation.

5.2.5 Assessment of the aggregation profile and aggregation kinetics of recombinant hMCAD proteins by Dynamic Light Scattering (DLS). In order to complement the structural studies of protein stability, thermal aggregation profiles for both hMCADwt and p.K304E variant were determined by Dynamic Light Scattering (DLS) in the absence and presence of FAD in the lysis buffer. The tetrameric fractions of the recombinant proteins were used for three different assays: a) thermal aggregation profile of the proteins with ramping of the temperature from 20 to 70°C; b) aggregation profile at physiological temperature (37°C); and c) aggregation profile at
42°C, in order to evaluate protein stability in MCADD patients harboring the p.K304E mutation during fever episodes. The temperature at which the protein starts to aggregate ($T_{agg}$) was given by the analysis of the signal corresponding to the size and/or intensity of the light dispersion. The $T_{agg}$ corresponds to the temperature at which the signal starts to significantly increase, due to the increment of the light dispersion as the result of the formed protein aggregates.

Thermal aggregation profile with ramping of the temperature (20 to 70°C) showed that the hMCADwt started to aggregate at about 40-41°C, independently of the addition of FAD in the lysis buffer during purification (Table 5). The determination of the protein melting point ($T_m$) corresponding to the inflection point of the curve (Fig. 19) revealed that the hMCADwt has two transitions, as observed in thermal denaturation profiles (DSF).

**Figure 19:** Thermal aggregation profiles (from 20 to 70°C), of the recombinant protein hMCADwt, obtained by DLS. The protein used (0.15 mg/mL) corresponds to the tetrameric fraction obtained by SEC purified without (A) or with (B) the addition of FAD to the lysis buffer before IMAC. Graphs represent the data obtained from a representative experiment performed in duplicate. Graphs and non-linear regression curve fitting were performed with the GraphPad Prism6® software.
Table 5: Temperature midpoint of aggregation ($T_{agg}$) and aggregation transition points of the recombinant hMCADwt and p.K304E variant obtained by Dynamic Light Scattering (increasing temperature within the range of 20 to 70°C).

<table>
<thead>
<tr>
<th></th>
<th>hMCAD wt</th>
<th>p.K304E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_{agg}$ (°C)</td>
<td>$T_{m1}$ (°C)</td>
</tr>
<tr>
<td>- FAD</td>
<td>39.6$^a$</td>
<td>43.8$^a$</td>
</tr>
<tr>
<td>+ FAD</td>
<td>41.3 ± 0.7$^b$</td>
<td>43.0 ± 1.5$^b$</td>
</tr>
</tbody>
</table>

The protein used (0.15 mg/mL) corresponds to the tetrameric fraction obtained by SEC, purified in the absence (-FAD) and presence (+ FAD) of FAD in the lysis buffer before IMAC. (a) Values represent the mean of duplicates of one experiment. b. Values represent the mean ± SD of two independent experiments performed in duplicate and were obtained by the non-linear regression of the intensity of the light dispersion and aggregated protein size (GraphPad Prism6®).

The p.K304E variant protein started to aggregate sooner than the wt form, around 35-36°C and also presented two thermal transitions states (Fig. 20). Interestingly in the variant protein, the presence of FAD in the purification steps led to an earlier protein aggregation when compared with the protein purified in the absence of the cofactor in the lysis buffer.

![Figure 20](image)

Figure 20: Thermal aggregation profiles (from 20 to 70°C), of the recombinant protein p.K304E, obtained by DLS. The protein used (0.15 mg/mL) corresponds to the tetrameric fraction obtained by SEC purified without (A) or with (B) the addition of FAD to the lysis buffer before IMAC. Graphs represent data obtained from a representative experiment performed in duplicate. Graphs and non-linear regression curve fitting were produced with the GraphPad Prism6® software.
The aggregation kinetics allowed the determination of the time at which the protein started to denature by aggregation. As observed, the hMCADwt protein started to aggregate at about 9.0 min at 37°C, independently of the addition of FAD during purification. At higher temperature (42°C), an earlier aggregation of the hMCADwt was observed, being also independent of the addition of the cofactor before the IMAC purification (Table 6 and Fig. 21A/21B). The aggregation kinetics data revealed that the p.K304E variant protein started to aggregate earlier than the wt protein, for both temperatures. Again, a negative effect of the addition of FAD during purification was observed for this variant, due to an earlier beginning of aggregation.

**Table 6**: Aggregation kinetics (min) at 37°C and 42°C, by DLS of recombinant hMCADwt and p.K304E variant purified in the absence and presence of FAD in the lysis buffer before IMAC purification.

<table>
<thead>
<tr>
<th></th>
<th>hMCAD wt</th>
<th>p.K304E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t(\text{lag}) (37°C)</td>
<td>t(\text{lag}) (42°C)</td>
</tr>
<tr>
<td>- FAD</td>
<td>9.0 min</td>
<td>4.1 min</td>
</tr>
<tr>
<td>+ FAD</td>
<td>9.4 ± 0.4 min</td>
<td>4.0 ± 0.2</td>
</tr>
</tbody>
</table>

Values represent the mean ± SD of two independent experiments performed in duplicate and were obtained by the non-linear regression of the values of intensity of the light dispersion and aggregated protein size (GraphPad Prism6®). \(t\text{lag}\) – lag phase. The statistical significance was determined using the multiple t tests and comparing the wt and the variant proteins ( * \(p<0.005\)).

**Figure 21A**: Thermal aggregation kinetics of the recombinant hMCADwt, assessed by DLS at 37°C (Graphs A and C) and 42°C (Graph B and D). The protein used (0.15 mg/mL) was the tetrameric fraction purified by SEC without addition (Graphs A and B) and with the addition of FAD (Graphs C and D) to the lysis buffer before IMAC purification. Graphs represent data obtained from a representative experiment performed in duplicate. Graphs and non-linear regression curve fitting were produced with the GraphPad Prism6® software.
5.3.1 Enzymatic activity and determination of the kinetic parameters. The enzymatic activity was measured in 8 μg of recombinant proteins, obtained in the absence and presence of FAD in the lysis buffer, after purification by IMAC and SEC (tetrameric forms), according with the method described by Zeng and Li. It is based in the decrease of the absorbance at 600 nm of the artificial redox pair PMS/DCPIP which represents the reaction between the FAD and ETF protein, respectively, after the substrate oxidation. All the determinations were performed in the presence of 33 μM of octanoyl-CoA (C8-CoA, CCOA) as substrate. The enzymatic activity determined in the absence of substrate was used as a control of the potential spontaneous reduction of the DCPIP component (Fig. 22).

As shown in the Figure 23 and Table 7, there was a reduction in the specific enzymatic activity of the p.K304E variant of hMCAD when compared with the wild-type form, with a mean of 46% of residual activity in the absence of FAD during purification.
Figure 22: Variation in time of the absorbance at 600 nm of the pair PMS/DCPIP in the absence (dotted line) and presence (full line) of 33 µM octanoyl-CoA (C8-CoA) using 8 µg purified protein after SEC (tetrameric fraction). The data shown is representative of the data obtained for hMCAD wt (blue) and p.K304E (black). The arrow indicates the starting of the reaction with the addition of the protein.

Nevertheless, in the presence of the cofactor before IMAC purification, the p.K304E seems to recover its specific enzymatic activity, abolishing the in vitro enzymatic activity deficiency characteristic of this mutant.

Table 7: Specific enzymatic activity, determined as a function of the variation of the DCPIP concentration in the reaction (µmol DCPIP.min⁻¹.mg⁻¹), for the hMCADwt and the p.K304E variant after SEC purification (tetrameric fraction), in the absence and presence of the cofactor FAD.

<table>
<thead>
<tr>
<th>hMCAD protein</th>
<th>Purification in the absence of FAD</th>
<th>Purification in the presence of FAD</th>
<th>P&lt;sub&gt;a&lt;/sub&gt;</th>
<th>P&lt;sub&gt;b&lt;/sub&gt;</th>
<th>Residual activity c (%) Wt</th>
<th>Residual activity c (%) Wt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Specific enzymatic activity&lt;sup&gt;a&lt;/sup&gt; (µmol.min⁻¹.mg⁻¹)</td>
<td>Residual activity&lt;sup&gt;c&lt;/sup&gt; (%) Wt</td>
<td>Specific enzymatic activity&lt;sup&gt;a&lt;/sup&gt; (µmol.min⁻¹.mg⁻¹)</td>
<td>Residual activity&lt;sup&gt;c&lt;/sup&gt; (%) Wt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wt</td>
<td>1.19 ± 0.47 (n=11)</td>
<td>--</td>
<td>1.73 ± 0.40 (n=2)</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p.K304E</td>
<td>0.53 ± 0.33 (n=5)</td>
<td>ns</td>
<td>45.6± 27.8 (n=5)</td>
<td>1.87 (n=1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Reaction conditions: 33 µM C8-CoA, 8 µg protein; <sup>b</sup> Statistical significance (P) determined by the unpaired Student’s t test (two-tailed P value) and compared the p.K304E variant with the wt protein: * p<0.05, ** p<0.01 and *** p<0.001; <sup>c</sup> Residual activity of the p.K304E variant in respect to the activity of the wt protein, thus considered as 100% activity; <sup>d</sup> Statistical significance was determined by the unpaired Student’s t test (two-tailed P value) and compared the effect of adding FAD to the lysis buffer during isolation of the recombinant proteins: # p<0.05 e ## p<0.01; ns – non significant; Data represent the mean ± SD; n – number of independent determinations performed in duplicate.

These results suggest that the structural alterations induced in the MCAD protein by the c.985G>A mutation may compromise the FAD binding to the catalytic pocket of the enzyme and thus, its enzymatic activity.

The kinetic parameters such as the maximum velocity (V<sub>max</sub>), affinity constant (K<sub>m</sub>), catalytic constant (K<sub>cat</sub>), and the catalytic efficiency (K<sub>cat</sub>/K<sub>m</sub>) were also measured by assessing the enzymatic activity of the tetrameric fraction of hMCAD proteins (purified without the addition
of FAD to the lysis buffer), at increasing concentrations of the substrate octanoyl-CoA (0 – 100 μM). The kinetic profiles revealed that both hMCAD recombinant proteins followed the Michaelis-Menten equation (Fig. 23).

The kinetic parameters were determined by the non-linear regression of the Michaelis-Menten equation and are indicated in Table 8.

The analysis of the kinetic parameters confirmed the enzymatic activity deficit reported for the p.K304E variant. The $V_{\text{max}}$ value for the mutant form is about 38% lower than the $V_{\text{max}}$ found for the wt protein, being translated in a decrease in the catalytic efficiency of the p.K304E variant (34%). Nevertheless, the affinity for octanoyl-CoA is similar for both recombinant proteins. As the experiments were only performed in the absence of FAD during purification it is not possible to assess the influence of the cofactor in the kinetic behavior of the hMCADwt and the p.K304E variant. This aspect should be further investigated.

**Table 8:** Kinetic parameters of recombinant hMCADwt and p.K304E variant (purified in the absence of FAD cofactor) using octanoyl-CoA as substrate after purification by SEC (tetrameric fraction).

<table>
<thead>
<tr>
<th>Recombinant proteins</th>
<th>$V_{\text{max}}$ (µmol. min$^{-1}$.mg$^{-1}$)</th>
<th>$K_{\text{m}}$ (µM)</th>
<th>$K_{\text{cat}}$ (min$^{-1}$)</th>
<th>$K_{\text{cat}}/K_{\text{m}}$ (µM$^{-1}$.min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hMCADwt</td>
<td>1.86±0.16</td>
<td>2.8±1.2</td>
<td>82.2</td>
<td>29.1</td>
</tr>
<tr>
<td>p.K304E</td>
<td>0.70±0.02</td>
<td>3.1±0.5</td>
<td>31.0</td>
<td>9.92</td>
</tr>
</tbody>
</table>

The data were analyzed with the GraphPad Prism (Prism 6 for Windows, version 6.01) software using a non-linear regression of the Michaelis-Menten equation. The catalytic activity was determined at room temperature (25°C), in the presence of 8 µg protein, the redox pair PMS/DCPIP and increasing concentration of the substrate octanoyl-CoA (0–100 µM), by the decrease in absorbance of DCPIP at 600 nm. $K_{\text{cat}}/K_{\text{m}}$ represents the catalytic efficiency. $K_{\text{cat}}$ was determined based upon the molecular weight of one recombinant hMCAD subunit (44.2 kDa). The data represent the mean ± SD of duplicates of two independent reactions for the hMCADwt and of one reaction for the p.K304E variant.
5.3.2 Determination of the in vitro stability of recombinant hMCAD proteins by thermal enzymatic inactivation. Thermal inactivation profiles of recombinant hMCAD proteins were determined by measuring the enzymatic activity, in the presence of 33 μM of octanoyl-CoA as substrate, after incubation of the proteins at increasing temperatures (25°C to 60°C). The thermal inactivation profiles of the hMCADwt and the p.K304E variant (tetrameric fraction; obtained without adding FAD in the lysis buffer) are depicted in the Figure 24A.

The data obtained showed a decrease in the enzymatic activity of the p.K304E variant when compared with the wt form, revealing thermal instability of the variant protein as previously observed in the DSF and DLS assays. The thermal midpoint of enzyme inactivation (T1/2) for each recombinant protein, i.e. the temperature at which is reached 50% of the enzymatic activity measured after incubation at 25°C, was calculated by the non-linear regression of the experimental data (Fig. 24B). The thermal instability of the p.K304E variant was confirmed with the p.K304E variant showing a T1/2 of 44.3°C against a T1/2 of 51.2°C for the hMCADwt protein. Nevertheless, these results did not reveal statistical significance due to the high variability of the experimental data, requiring further investigation.

Figure 24: Thermal inactivation profiles of the hMCADwt and p.K304E variant tetrameric fractions obtained in the absence of FAD in the lysis buffer during purification. The enzymatic activity was determined as described above using 33 μM of C8-CoA as substrate, 8 μg protein, the redox pair PMS/DCPIP, after incubation of the protein at increasing temperatures (25°C to 60°C) and following the variation of the absorbance of DCPIP at 600 nm for at least 60 seconds. The enzymatic activity determined after incubation at 25°C was considered 100% and the relative enzyme activity was determined for the rest of the tested temperature points. Graph A: Data represents the mean ± SD of two (hMCADwt) and three (p.K304E) independent reactions performed in duplicate; Graph B: Non-linear regression (using the software GraphPad Prism 6 for Windows, version 6.01) was applied to the experimental data of two (hMCADwt) and three (p.K304E) independent reactions performed in duplicate. This allowed the determination of the T1/2 for each protein.
6. Computational results

6.1 Characterization of the pig MCAD wild-type enzyme (pMCADwt)

The first goal of the MD studies consisted in the structural characterization of the pMCADwt enzyme as well as to provide insights in its dynamic behavior under different conditions. Thus, three systems were built for the pMCADwt enzyme: APO (no substrate, no FAD), FAD (no substrate, with FAD), and LIPID (with substrate and FAD). The MD simulations were performed during 100 ns using the crystallographic structure of the Sus scrofa (pig) MCAD enzyme (PDB ID: 1UDY), in complex with the FAD cofactor and the SCOA substrate as the initial coordinates. The analysis of the MD trajectories allowed inferring about the MCAD structural stability throughout the simulations as well as the importance of the presence of the FAD and of both FAD and the substrate on MCAD stability and dynamics.

6.1.1 Analysis of protein stability

Root mean square deviation (RMSD) and radius of gyration (Rg). The RMSD and Rg values are important tools to analyze protein stability during a molecular dynamics simulation by comparing the deviations of the atomic positions with the initial structure – in this case, the crystallographic structure of pMCADwt (PDB ID: 1UDY). The RMSD and Rg (calculated using the protein’s Cα atoms) of the protein’s tetrameric (Fig. 25A and 25B) and monomeric forms (Annex, Fig. S2) are plotted for each system developed (APO, FAD and LIPID).

Regarding the root mean square deviation (RMSD), a similar behavior was observed for all simulations, the values increased in the first 15 ns until a plateau was reached. In all simulations, after 60 ns the protein was considered to be stabilized and the following calculations were carried out from that point forward. The mean RMSD values obtained from the different simulations (APO – 0.35 nm, FAD – 0.34 nm and LIPID – 0.31 nm) revealed that the protein is stable during the simulation time (Fig. 26A), when compared with the crystallographic structure. Interestingly, these results point out that the presence of both FAD/SCOA (LIPID system) seems to confer a slight stabilization effect in the protein structure. However, while the RMSD of individual monomers revealed a distinct behavior among those in the APO system (0.22 nm to 0.35 nm), the presence of FAD (FAD system) and FAD/SCOA (LIPID system) led to a more homogeneous behavior (0.25 nm to 0.27 nm) between the four monomers (Annex, Figs. S2-S4).

The radius of gyration gives a rough measurement of the compactness of a protein as a function of time by calculating the position of an atom i against the center of mass of a molecule during the simulation. Again, no relevant differences between the initial (crystallographic structure) and the MD structure were observed for each system (Fig. 26B), thus indicating that
the protein’s spatial organization was maintained in all systems throughout the simulation (APO – 3.49 nm; FAD – 3.50 nm and LIPID – 3.52 nm).

Figure 25: (A) Root mean square deviation (RMSD) and (B) radius of gyration (Rg) of pMCADwt tetramer MD simulations against the crystallographic coordinates (PDB ID: 1UDY): APO (black), FAD (red) and LIPID (green) systems.

Protein secondary structure assessment. In order to validate the above mentioned results, and due to potential stability issues related with incorrect dihedral parameters of the backbone in GROMOS 53a6 parameter set, which may lead to equivocal conclusions, the evolution of the secondary structure in function of time was also calculated applying the DSSP software.80,81 The results show that the secondary structure of the pMCADwt protein remained stable in all simulations and is similar to the crystallographic structure (50% vs. 54% helices and 14% vs. 15% β-sheets, respectively).

Root mean square fluctuation (RMSF) and analysis of β-factors. The root mean square fluctuation (RMSF) represents the average atomic fluctuations of the protein’s residues and provides information about the secondary structure flexibility.41,43–46 Similar residue fluctuation profiles were observed in all systems (Fig. 26). The majority of the residues fluctuate below 0.2 nm and the largest fluctuations (above 0.4 nm) were detected in residues located in coils, α-helices exposed to solvent (residues 352-366) and β-sheets (residues 180-200). The most stable residues (below 0.1 nm) are located in α-helices at the protein core, essential for the tetramer assembly and stability. The residues located in the active site are equally stable, especially those interacting with the flavin ring of FAD and with the terminal acyl group of the substrate.

In addition to RMSF, β-factors also give insights on the protein dynamics and, more importantly, allow a direct comparison between residue fluctuations in the crystallographic structure against MD simulations. As shown in Figure 27, the β-factors obtained from the MD simulations are in agreement with the ones retrieved from the crystallographic structure, both showing an increased stability of the internal α-helices when compared with the outer β-sheet domains.
Figure 26: Root mean square fluctuation (RMSF) of pMCADwt monomers’ MD simulations against the crystallographic coordinates (PDB ID: 1UDY): APO (black), FAD (red) and LIPID (green) systems.

Figure 27: Comparison between the β-factors retrieved from the crystallographic structure (left) (PDB ID: 1UDY) and from MD simulations (right) of the pMCADwt in the presence of FAD/SCOA (LIPID system). High β-factor values represent more flexible regions (red) while low values indicate more stable regions (blue). Figures were created with MOE v2013.08.

6.1.2 Characterization of the catalytic pocket

For the maximum catalytic efficiency of MCAD enzyme, the catalytic residue (E376) and the FAD cofactor must be accessible by the substrate. The crystallographic structure shows that the catalytic pocket of pMCAD where the E376 residue and the FAD are located is formed by the two monomers of one dimer. Alterations in the monomer’s folding or assembly may compromise the formation of the catalytic pocket and impair MCAD function. It is therefore
essential to thoroughly characterize the structure and dynamics of the catalytic pocket to further understand how mutations may affect its structure and ultimately the enzymatic function.

**Volume.** One of the properties of the catalytic pocket that can change directly by mutations or by impairment of the tetramer assembly is its internal volume. An estimation of the changes of these volumes over the course of the simulation was performed with the VOIDOO\textsuperscript{71} package. By specifying spatial coordinates of two opposite residues located inside the cavity (e.g. W166 and E376), as a seed point for a probe sphere, all the points in a specific 3D grid, accessible to the center of the probe, were used to calculate the volume of the cavity. In addition, the spatial coordinates of the lining cavity residues and waters inside the catalytic pockets were also estimated. The probability distribution of the mean volumes of the binding pockets, disregarding the FAD and SCOA ligands, was plotted for all systems (APO, FAD and LIPID) (Fig. 28).

![Figure 28: Probability distribution function of the average volumes of the catalytic pockets (A, B, C and D) of pMCADwt. Data was obtained with VOIDOO\textsuperscript{71} package.](image)

Figure 28 shows that, as expected, in the absence of ligands (FAD/SCOA; APO system) the catalytic pocket collapses and only residual volumes are observed (80% of volumes within 75-315 Å\textsuperscript{3}), when compared with those observed in the FAD system (80% of volumes within 235-700 Å\textsuperscript{3}) and in the LIPID system (80% of volumes within 425-700 Å\textsuperscript{3}). In the presence of FAD (FAD system) and of FAD/SCOA (LIPID system), the pocket volumes became increasingly larger indicating that both ligands have important roles in defining the pocket volumes.

The mean number of water molecules (calculated by VOIDOO\textsuperscript{71}) inside the catalytic pockets did not change significantly in the APO (12 water molecules) and FAD (13 water molecules) systems. Nonetheless, this number decreased to values similar to those reported for the crystallographic
structure in the presence of the substrate (LIPID system – 6 water molecules). Thus, the entry of the acyl-CoA substrate into the catalytic pocket is followed by a decrease in the total number of waters inside the cavity, in order to better accommodate it and provide a hydrophobic environment required for the α,β-dehydrogenation reaction.\textsuperscript{6,7,11}

**FAD and SCOA binding regions.** From the MD simulations and in agreement with the crystallographic data, each catalytic pocket is mainly formed by residues from both monomers of the dimer involving all three domains. The FAD and the SCOA binding sites are well defined inside the pocket with the FAD binding region being more hydrophilic and the SCOA binding region more hydrophobic. These features were assessed by EPOS\textsuperscript{66},\textsuperscript{73,82} which uses the PASS\textsuperscript{82} algorithm to detect protein’s cavities on protein surface as well as to predict the binding sites for ligands, mapping the pocket lining atoms.

As also described in the crystallographic structure, the FAD of one monomer has an extended conformation that enables the formation of hydrogen bonds (HBs) with residues from the neighboring monomer of the dimer [A/B; B/A; C/D; D/C - (Table 9 and Fig. 29)].

### Table 9: Hydrogen bonds between the FAD of one monomer and the neighboring monomer of the dimer unit in pMCADwt – FAD system (A) and LIPID system (B).

<table>
<thead>
<tr>
<th></th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG\textsubscript{HB}</td>
<td>&lt;N\textsubscript{HB}&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>FAD:R281</td>
<td>1579</td>
<td>-23.8</td>
<td>0.89</td>
<td>1678</td>
</tr>
<tr>
<td>FAD:V350</td>
<td>1497</td>
<td>-23.7</td>
<td>0.71</td>
<td>309</td>
</tr>
<tr>
<td>FAD:G353</td>
<td>2171</td>
<td>-24.6</td>
<td>0.79</td>
<td>1150</td>
</tr>
<tr>
<td>FAD:N354</td>
<td>3142</td>
<td>-25.7</td>
<td>0.93</td>
<td>---</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG\textsubscript{HB}</td>
<td>&lt;N\textsubscript{HB}&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>FAD:R281</td>
<td>1058</td>
<td>-22.8</td>
<td>1.14</td>
<td>321</td>
</tr>
<tr>
<td>FAD:V350</td>
<td>726</td>
<td>-21.8</td>
<td>0.99</td>
<td>3136</td>
</tr>
<tr>
<td>FAD:G353</td>
<td>3593</td>
<td>-25.9</td>
<td>0.96</td>
<td>109</td>
</tr>
<tr>
<td>FAD:N354</td>
<td>1731</td>
<td>-24.0</td>
<td>0.63</td>
<td>12502</td>
</tr>
</tbody>
</table>

τ, HB lifetime in ps; ΔG\textsubscript{HB}, free energy of hydrogen bonding in kJ.mol\textsuperscript{-1}; <N\textsubscript{HB}>, average number of HB per timeframe at 310K.

Although the total number of contacts (calculated by \textit{g_contacts}) remains similar in both FAD and LIPID systems (Annex, Tables S1 and S2), it was possible to verify that, in general upon SCOA binding (LIPID system), the average number of HBs per timeframe for each residue and in all monomers increased, suggesting that SCOA binding reinforces the HB network between the FAD
and the neighboring monomer and further increases the dimer cohesion. Moreover, in all monomers, HBs were observed between the FAD and a large number of residues lining the catalytic pocket, being the most representative ones (> 50% of contact frequencies) depicted in Figure 29 and Tables S2 and S4 (Annex).

This assumption was validated by calculating the free energies of FAD binding (calculated by $g_{mmpbsa}^{75}$) to the protein for all monomers. When no substrate is present inside the catalytic pocket the average binding energy for the FAD cofactor is $-886 \text{ kJ.mol}^{-1}$ (FAD system) while in the presence of SCOA (LIPID system) the binding energy is 13% lower ($-1023 \text{ kJ.mol}^{-1}$). Thus, together with the above data, it is clear that the FAD binding energy improved, as a result of the reinforcement of the FAD-protein HB network induced by the substrate binding.

As shown in Figures 29A and 29B, the FAD interaction pattern remains similar in both FAD and LIPID systems, in agreement with the crystallographic data. The HB network is distributed by the flavin ring of FAD as well as by the ribityl and phosphate groups of the cofactor in the FAD system (Figure 29A). Upon substrate binding (LIPID system), this HB networked is reinforced mainly in the phosphate region and adenine group (Figure 29B). Nevertheless, the aromatic residues Y133 and Y375, near the flavin ring, as well as the H291 residue (Figures 29A/29B), surrounding the adenine group, are able to establish π-π stacking interactions (between aromatic rings), thus contributing to stabilize these regions.

From the analysis of the Figure 29C, is possible to see that the SCOA is also mainly anchored to protein by HBs with the protein residues lining the catalytic pocket mainly in the phosphate and 3'-phosphoadenosine groups. However, as mentioned above for the FAD interactions, the Y133 and Y375 aromatic residues also contribute to stabilize the substrate inside the pocket by creating an aromatic cage that interacts with the terminal acyl group and the thioester region of the substrate, respectively. Interestingly, the F252 and F245 residues interact with polar groups of the SCOA (near the pantothenate moiety), suggesting an important role of these residues on the substrate’s stabilization within the catalytic pocket.

Summarizing, HBs seem to be the main features by which the FAD and the SCOA bind to pMCADwt. Moreover, the aromatic residues Y133 and Y375 also seem to play an important role in stabilizing both ligands inside the catalytic pocket.
Figure 29: Representation of the ligand interactions with the protein residues lining the catalytic pocket of pMCADwt (observed in all monomers). (A) Representation of the FAD interactions with the protein residues showing > 50% of contacts frequency along the simulation (calculated by g_contacts \(^{24}\) (Annex, Tables S1, S3 and S5) in the FAD system and (B) in the LIPID system; (C) representation of the SCOA interactions with the protein residues showing > 50% of contacts frequency along the simulation. Aromatic residues (orange), non-polar residues (green), polar residues (blue); HBs are represented by dotted lines. The arrows represent HBs between the ligands and the protein’s residues: HBs with the backbone’s residues (blue) and side-chain’s residues (green). Solvent accessible surface area is represented as a blue smudge. Figures were created with MOE v2013.08 from MD simulations.

6.1.3 Active site and molecular mechanism of the enzymatic reaction

The geometry of the active site of the pMCADwt and the correct alignment between the FAD, the substrate and the catalytic residue E376 are essential for the \(\alpha,\beta\)-dehydrogenation reaction to occur.\(^6\)

Other residues (Fig. 30), located at the enzyme’s active site are also described as being crucial on MCAD catalysis.\(^6,7\) In fact, we observed that, upon complex formation, some residues changed their side-chain conformations in order to better accommodate the substrate. The data
gathered from MD simulations show that in presence of FAD (FAD system), the carboxylate of E99 and the hydroxyl group of Y375 residues were able to interact by HBs (Tables 10A/10B/10C). Upon SCOA binding (LIPID system), the E99–Y375 HBs were broken allowing the fatty acyl moiety of the substrate to bind deeper inside the pocket. These findings are in agreement with the crystallographic structure, which emphasizes the alterations in the side-chain conformations of the E99 and Y375 residues as the most relevant to increase the pocket depth. As described in the crystallographic structure and in other experimental studies, in the presence of FAD/SCOA (LIPID system), HBs were observed between the carbonyl moiety of SCOA, the catalytic residue E376 and the 2′-OH of the ribityl group of FAD (Table 10C). According to the literature, these HBs are the ones responsible for a precise positioning and alignment of the substrate, and also participate in the substrate’s recognition and activation. The catalytic residue E376 may thus play an important role in the active site geometry.

Figure 30: Representation of the catalytic pocket of pMCADwt in two different perspectives, highlighting the location of the residues that are important for the MCAD catalysis (described below). The FAD binding site is mainly hydrophilic (purple) and the SCOA binding site is mostly hydrophobic (green). Figures were created with MOE v2013.08 from MD simulations.

Surprisingly, and unlike described in the crystallographic structures, the simulations showed that upon SCOA binding, the carboxylate group of the E376 residue switches dramatically towards the opposite side of the lipid’s Cα-Cβ bond. The mean distance between the carboxylate group of the E376 residue and the Cα atom of SCOA is 0.72 nm, which is too long to abstract the α-
proton, according with the crystallographic structure.\textsuperscript{6} Nevertheless, during the simulation (LIPID system) the carboxylate group of the E376 residue switches towards the water molecules found in the vicinity of its side-chain.

**Table 10:** Hydrogen bonds between the protein residues located at the active site of pMCADwt (APO system), that are described to be important on MCAD catalysis. (A) APO system; (B) FAD system and (C) LIPID system.

<table>
<thead>
<tr>
<th></th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Hbond</td>
<td>τ</td>
<td>ΔG&lt;sub&gt;HB&lt;/sub&gt;</td>
<td>N&lt;sub&gt;HB&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td>E99:Y375</td>
<td>701</td>
<td>-21.7</td>
<td>1.55</td>
</tr>
<tr>
<td>B</td>
<td>E99:Y375</td>
<td>1465</td>
<td>-23.6</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>R256:E376</td>
<td>2735</td>
<td>-25.2</td>
<td>0.50</td>
</tr>
<tr>
<td>C</td>
<td>E99:Y375</td>
<td>1865</td>
<td>-24.2</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>R256:E376</td>
<td>218</td>
<td>-18.7</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>SCOA:FAD</td>
<td>1412</td>
<td>-23.5</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>SCOA:E376</td>
<td>218</td>
<td>-18.7</td>
<td>0.07</td>
</tr>
</tbody>
</table>

\(τ\), HB lifetime in ps; \(ΔG_{HB}\), free energy of hydrogen bonding in kJ.mol\(^{-1}\); \(\langle N_{HB}\rangle\), average number of HB per timeframe at 310K.

These data seem to suggest that, instead of a direct abstraction of the \(\alpha\)-proton by the catalytic residue E376, the \(\alpha_β\)-dehydrogenation reaction may occur through a water-mediated mechanism.

Another important residue for the MCAD catalysis is the residue R256, which is conserved in the substrate binding domain of the acyl-CoA dehydrogenases family.\textsuperscript{27,86} The simulations showed that the mean distance between the carboxylate group of the E376 residue and the guanidine group of the R256 residue was similar in both APO and FAD systems (0.54 nm and 0.56 nm respectively), while in the LIPID system the mean distance between them is reduced to 0.4 nm, instead to increase according with the crystallographic data [native enzyme (0.43 nm); upon SCOA binding (0.48 nm)]. Although the results do not fully agree with the crystallographic structure, they can be explained by a distinct orientation of the carboxylate group of the E376 residue towards to the opposite side of the lipid’s Cα-Cβ bond, as mentioned above. Hydrogen Bonds were observed between the R256 and E376 residues in all systems (Table 10A/10B/10C),
being more frequent and stronger upon SCOA binding, due to side-chain rearrangements of both of these residues. The proximity of the negatively charged E376 residue and of the positively charged R256 residue may also strongly infer a contribution to the α,β-dehydrogenation reaction mechanism.

6.2 Characterization of human MCAD wild-type enzyme (hMCADwt)

In order to provide insights into the experimental data and also validate the MD method applied to MCAD proteins, the hMCADwt protein structure was also used after reversing the crystallographic structure of the E376G/T255E double mutant of hMCAD protein, in complex with FAD and the CCOA substrate into the hMCADwt. Different systems were simulated as for in pMCADwt: APO (no substrate, no FAD), FAD (no substrate, with FAD), and LIPID (with substrate and FAD). The MD simulations were performed during 100 ns.

6.2.1 Analysis of protein stability

Root mean square deviation (RMSD) and radius of gyration (Rg). After reversing the T255E/E376G mutations present in the crystallographic structure of the hMCAD protein retrieved from PDB ID: 1EGC, the hMCADwt protein stability was assessed by comparison with the reversed initial crystallographic coordinates. Similarly to pMCADwt, the protein’s Ca atoms were used to calculate the RMSD and Rg values of the hMCADwt tetramer (Figure 31) and of its individual monomers (Annex, Fig. S4-S6) for each system (APO, FAD and LIPID).

A similar RMSD profile (Figure 31A) was found in all simulations, increasing in the first 20 ns until stabilization at 60 ns of the simulation time. Mean RMSD values (calculated from the equilibrated structures) obtained from the different simulations (APO – 0.33 nm; FAD- 0.28 nm and LIPID – 0.31 nm), revealed that the protein was stable throughout the simulation time as observed in the pMCADwt (APO – 0.35 nm, FAD – 0.34 nm and LIPID – 0.31 nm).

![Figure 31: (A) Root mean square deviation (RMSD) and (B) radius of gyration (Rg) of hMCADwt tetramers’ MD simulations against the crystallographic coordinates (PDB ID:1EGC - reversed): APO (black), FAD (red), and LIPID (green) systems.](image-url)
Nevertheless, and unlike the pMCADwt, the hMCADwt seems to be more stable in the presence of FAD (FAD system) during the simulation. The RMSD values of individual monomers revealed a similar behavior between them, for all the considered systems (Annex, Fig. S4-S6).

In respect to Rg, no significant differences were found for each system between the crystallographic and the MD structures, indicating that the protein’s three-dimensional structure was maintained, in all systems, along the simulation time (Fig. 31B). However, the lower Rg values in the FAD system, together with the RMSD evolution described above, suggests that the presence of FAD had a positive effect in the whole tetramer stability.

**Protein secondary structure assessment.** To analyze the evolution of the secondary structure of the protein, DSSP \(^{80,81}\) calculations were performed. The results obtained revealed that the protein remained stable in all simulations with both α-helices (53%) and β-sheets (14%) achieving equal results when compared to the original crystallographic structure (PDB ID: 1EGC). These data also demonstrate that the GROMOS 53a6\(^{79}\) parameter set can be successfully applied in the hMCADwt MD simulations as already observed in the pMCADwt.

**Root mean square fluctuation (RMSF) and analysis of β-factors.** A similar behavior was found in all systems for the residues fluctuation (Fig. 32), with the majority of the residues fluctuating below 0.2 nm.

---

**Figure 32:** Root mean square fluctuation (RMSF) of hMCADwt monomers’ MD simulations against crystallographic coordinates (PDB ID: 1EGC – reversed): APO (black), FAD (red), and LIPID (green) systems.
The largest fluctuations (above a cut-off value of 0.4 nm) were detected in coils or in \( \alpha \)-helices exposed to the solvent (residues 352-366), and \( \beta \)-sheets (residues 180-200) as observed in pMCADwt. Noteworthy, and similarly to pMCADwt, the residues interacting with the flavin ring of FAD and with the terminal acyl group of CCOA, as well as the internal \( \alpha \)-helices located at the dimer-dimer interface, were found to be the most stable regions.

The residue dynamic mobilities were also analyzed through the calculation of the \( \beta \)-factors. The results obtained indicate an increasing stability from the surface (\( \beta \)-sheet domains) towards the protein core (\( \alpha \)-helical domains), in agreement with the \( \beta \)-factors calculated directly from the crystallographic data and with those obtained in the pMCADwt protein (Fig. 33).

![Figure 33](image)

**Figure 33:** Comparison between the \( \beta \)-factors retrieved from the crystallographic structure (PDB ID: 1EGC) (left) and from MD simulations (right) of the hMCADwt. High \( \beta \)-factor values represent more flexible regions (red) while low values indicate more stable regions (blue). Figures were created with MOE v2013.08.

In summary, the hMCADwt protein was stable throughout the simulation time, maintaining not only its secondary structure but also the stability of the whole tetramer (when compared with the crystallographic structure). Interestingly, the presence of the cofactor (FAD system) seems to confer additional stability to the tetramer, similarly to that observed in the pMCADwt MD simulations.

### 6.2.2 Characterization of the catalytic pocket

**Volume.** The volume and the number of water molecules inside the catalytic pockets were analyzed for the whole trajectory with the VOIDOO\(^{71}\) package, ignoring both FAD and CCOA ligands (as mentioned in pMCADwt), in order to correctly estimate the cavities volume.

The probability distribution of the mean volumes of the four catalytic pockets from all simulations is represented in Figure 34.
While in the absence of ligands (APO system) the pocket volumes were smaller (80% of the volumes within 90-290 Å³), the presence of the FAD cofactor alone (FAD system) or of FAD and the substrate (LIPID system) induced the formation of cavities with larger volumes (80% of the volumes within 235-740 Å³ and 290-650 Å³, respectively).

When analyzing the mean number of water molecules content inside the catalytic pockets of the hMCADwt protein, no considerably differences were observed between the APO (13 water molecules) and the FAD (15 water molecules) systems. However, these numbers decrease in the LIPID system (7 water molecules), being in agreement with the crystallographic data which point to a reduction of the number of water molecules inside the pocket upon substrate binding, as also observed in pMCADwt.⁶

**Figure 34:** Probability distribution function of the mean volumes of the catalytic pockets (A, B, C and D) of hMCADwt. Data obtained with VOIDOO⁷¹ package.

**FAD and CCOA binding regions.** From the MD simulations and in agreement with the pMCADwt structure, each catalytic pocket of hMCADwt also comprises residues from both monomers of the dimer and from the three domains. The FAD binding site is, generally, more hydrophilic when compared with the CCOA binding region (assessed by EPOSBP⁷³,8²). The extended conformation of FAD allowed the formation of stable HBs with residues from the neighboring monomer (Fig. 35, Table 11). Although the residues described in the tables below do not present > 50% of contacts frequency [calculated by g_contacts⁷⁴ (Annex, Tables S6, S8 and S10)] along the simulation in all monomers, and thus are not shown in Fig. 36, they contribute to attach the FAD to the neighboring monomer, as observed in pMCADwt.

The average number of HBs per timeframe tended to increase in the LIPID system suggesting that the presence of CCOA strengthens the HB network induced by the FAD, which may further
increase the cohesion between the cofactor and the protein and contribute to the quaternary structure stability (as also seen in pMCADwt). As shown in Figures 35A/35B and in Table 11, the FAD cofactor established HBs with a large number of residues inside the catalytic pocket in both FAD and LIPID systems of the hMCADwt. As for the pMCADwt, only the most representative (> 50% of contacts frequency) residues are described (Annex, Table S7 and S9). Both Figures 35A and 35B, as in pMCADwt, show similar interaction patterns in both FAD and LIPID systems, mainly located in the phosphates and flavin moieties.

Figure 35: Representation of the ligand interactions with protein residues lining the catalytic pocket of hMCADwt (observed in all monomers), (A) Representation of the FAD interactions with residues showing > 50% of contacts frequency along the simulation (calculated by g_contacts (Annex, Tables S6, S8 and S10)) in the FAD system and (B) in the LIPID system; (C) representation of the CCOA interactions with protein residues showing > 50% of contacts frequency along the simulation. Aromatic residues (orange), non-polar residues (green), and polar residues (blue); HBs are represented by dotted lines. The arrows represent HBs between the ligands and the protein residues: HBs with the backbone’s residues (blue) and side-chain’s residues (green). Solvent accessible surface area is represented as a blue smudge. Figures were created with MOE v2013.08 from MD simulations.
Table 11: Hydrogen bonds between the FAD of one monomer and the neighboring monomer of the dimer unit in hMCADwt – FAD system (A) and LIPID system (B).

<table>
<thead>
<tr>
<th></th>
<th>A Monomer A</th>
<th></th>
<th>A Monomer B</th>
<th></th>
<th>A Monomer C</th>
<th></th>
<th>A Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG&lt;sub&gt;HB&lt;/sub&gt;</td>
<td>&lt;N&lt;sub&gt;HB&lt;/sub&gt;&gt;</td>
<td>τ</td>
<td>ΔG&lt;sub&gt;HB&lt;/sub&gt;</td>
<td>&lt;N&lt;sub&gt;HB&lt;/sub&gt;&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>FAD:R281</td>
<td>961</td>
<td>-22.5</td>
<td>0.40</td>
<td>823</td>
<td>-22.1</td>
<td>0.15</td>
<td>3170</td>
</tr>
<tr>
<td>FAD:I350</td>
<td>4759</td>
<td>-26.6</td>
<td>0.25</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>1248</td>
</tr>
<tr>
<td>FAD:G353</td>
<td>1608</td>
<td>-23.8</td>
<td>0.6</td>
<td>5760</td>
<td>-27.1</td>
<td>0.19</td>
<td>9999</td>
</tr>
<tr>
<td>FAD:N354</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>670</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B Monomer A</th>
<th></th>
<th>B Monomer B</th>
<th></th>
<th>B Monomer C</th>
<th></th>
<th>B Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG&lt;sub&gt;HB&lt;/sub&gt;</td>
<td>&lt;N&lt;sub&gt;HB&lt;/sub&gt;&gt;</td>
<td>τ</td>
<td>ΔG&lt;sub&gt;HB&lt;/sub&gt;</td>
<td>&lt;N&lt;sub&gt;HB&lt;/sub&gt;&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>FAD:R281</td>
<td>809</td>
<td>-22.1</td>
<td>1.01</td>
<td>471</td>
<td>-20.7</td>
<td>1.22</td>
<td>1348</td>
</tr>
<tr>
<td>FAD:I350</td>
<td>1318</td>
<td>-23.3</td>
<td>0.93</td>
<td>270</td>
<td>-19.2</td>
<td>0.68</td>
<td>1571</td>
</tr>
<tr>
<td>FAD:G353</td>
<td>1838</td>
<td>-24.2</td>
<td>0.85</td>
<td>13143</td>
<td>-29.3</td>
<td>0.97</td>
<td>853</td>
</tr>
<tr>
<td>FAD:N354</td>
<td>4591</td>
<td>-26.5</td>
<td>1.63</td>
<td>11288</td>
<td>-28.9</td>
<td>1.22</td>
<td>3389</td>
</tr>
</tbody>
</table>

τ, HB lifetime in ps; ΔG<sub>HB</sub>, free energy of hydrogen bonding in kJ.mol<sup>-1</sup>; <N<sub>HB</sub>>, average number of HB per timeframe at 310K.

By calculating the free energy of binding for the FAD (calculated by g_mmpbsa<sup>75</sup>), it was possible to verify that in the presence of the CCOA (LIPID system) the mean binding energy for the cofactor of all monomers was lower when compared with the FAD system (-1016. vs -845 kJ.mol<sup>-1</sup> respectively), meaning that the cofactor stability increased upon substrate binding, mainly due to a reinforcement of the HBs network with the protein. Nevertheless, and similarly to that observed in pMCADwt, the presence of the aromatic residue Y375 (flavin ring), as well as of the H291 residue (adenine group), also played an important role on FAD stability inside the catalytic pocket.

Regarding the CCOA substrate, it also established HBs with residues inside the catalytic pocket (Fig. 35C) around the phosphate region, as observed in pMCADwt. Interestingly, in all monomers, the hydrophobic residues F252 and F284 interacted with polar groups of CCOA (near the phosphate moiety), suggesting an important role of these residues on the substrate stability, in particular the F252 residue which contributed to the substrate stability in both MCAD wt proteins.

6.2.3 Active site and molecular mechanism of the enzymatic reaction. As previously mentioned, the correct alignment of the flavin ring of the cofactor, the fatty acyl moiety of the substrate and the catalytic residue E376 is crucial for the α,β-dehydrogenation reaction to occur (Fig. 36).<sup>6</sup>
Figure 36: Representation of the catalytic pocket of hMCADwt in two different perspectives, highlighting the location of residues that are important for MCAD catalysis (described below). The FAD binding site is mainly hydrophilic (purple) and the SCOA binding site is mostly hydrophobic (green). Figures were created with MOE v2013.08 from MD simulations.

From MD simulations (LIPID system), it was observed a severe misalignment between the fatty acyl moiety of CCOA and the flavin ring of FAD for the monomers A and C. The trajectory analysis (by visual inspection) showed that CCOA did not bind correctly in the active site (Fig. 37A), being located almost entirely outside the binding cavity in monomer A (Fig. 37B) and not binding deeper inside the pocket in case of monomer C (Fig. 37C). However, a careful analysis of the information provided by the PDB file (ID: 1EGC) revealed that the binding site was better characterized, regarding CCOA, for both monomers B and D than for monomers A and C (based on the site description available in the correspondent PDB header).

Moreover, a direct inspection of the ligand interactions (using MOE) in the original crystallographic structure also showed that, for both monomers A and C, the number of CCOA-protein interactions is fewer when compared with the other subunits. Therefore, monomers B and D were chosen to analyze the structure and dynamics of the active site of the hMCADwt protein.

Although fewer than in the pMCADwt MD simulation (LIPID system), HBs were also found between the carbonyl oxygen of CCOA, the catalytic residue E376, and the 2'-OH of the ribityl
group of FAD (Table 25). This observation emphasizes the importance of these HBs to substrate recognition by hMCAD and mainly to substrate activation.\textsuperscript{6}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure37.png}
\caption{Representation of the misalignment observed in the hMCADwt (LIPID system): (A) Correct substrate position with the terminal acyl group between the (re)-face of the flavin ring of FAD (as described) and the catalytic residue (E376); (B) representation of the CCOA exit from the catalytic pocket in monomer A; and (C) representation of the substrate that not bind deep inside the pocket. Figures were created with MOE v2013.08 from MD simulations.}
\end{figure}

As already highlighted for the pMCADwt protein, also for the hMCADwt protein the data obtained from MD simulations showed that the E99 and Y375 residues had the most pronounced side-chain rearrangement upon CCOA binding. Moreover, while these residues were able to establish HBs between them in the absence of substrate (FAD system), when the substrate was located inside the cavity the HBs were broken, allowing the terminal acyl group
of the substrate to better fit the pocket and to bind deeper (Tables 12A/12B/12C). The exception to these findings lies in the monomer C where the substrate did not bind deep inside the pocket.

The MD simulations of hMCADwt also showed, as in pMCADwt, that in the LIPID system, the carboxylate group of E376 residue switched dramatically towards the opposite side of the lipid’s Ca-Cβ bond. The mean distance between the carboxylate group of the E376 residue and the Ca atom of CCOA was 0.70 nm (similarly to pMCADwt), too long to be able to abstract the α-proton. Nevertheless, in one of the MD simulations (LIPID system), the carboxylate group of E376 residue switched toward the water molecules found in the vicinity of its side-chain. Thus, we hypothesize that an indirect approach for the catalytic residue to abstract the α-proton of the substrate may occur in hMCADwt as in pMCADwt.

As mentioned above (and also in pMCADwt), the orientation of the catalytic residue E376 towards the opposite side of the lipid’s Ca-Cβ bond and the rearrangement of the guanidine group of the R256 residue upon CCOA binding, led to a distance reduction between their side-chains (APO – 0.73 nm; FAD – 0.59 nm and LIPID – 0.43 nm). These results, although not in agreement with the crystallographic structure, can still be explained by the orientation of the side-chain of the E376 residue.

**Table 12:** Hydrogen bonds between the residues located in the active site of hMCADwt (APO system), that are described to be important on MCAD catalysis. (A), APO system; (B), FAD system and (C), LIPID system.

<table>
<thead>
<tr>
<th>A</th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG_Hb</td>
<td>&lt;N_Hb&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>E99:Y375</td>
<td>3886</td>
<td>-26.1</td>
<td>0.22</td>
<td>---</td>
</tr>
<tr>
<td>R256:E376</td>
<td>363</td>
<td>-20.0</td>
<td>0.03</td>
<td>345</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG_Hb</td>
<td>&lt;N_Hb&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>E99:Y375</td>
<td>5913</td>
<td>-27.2</td>
<td>0.57</td>
<td>6230</td>
</tr>
<tr>
<td>R256:E376</td>
<td>269</td>
<td>-19.2</td>
<td>0.07</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG_Hb</td>
<td>&lt;N_Hb&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>E99:Y375</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>R256:E376</td>
<td>1521</td>
<td>-23.7</td>
<td>1.89</td>
<td>1019</td>
</tr>
<tr>
<td>CCOA:FAD</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>414</td>
</tr>
<tr>
<td>CCOA:E376</td>
<td>---</td>
<td>---</td>
<td>1227</td>
<td>-24.1</td>
</tr>
</tbody>
</table>

τ, HB lifetime in ps; ΔG_Hb, free energy of hydrogen bonding in kJ.mol⁻¹; <N_Hb>, average number of HB per timeframe at 310K.
Strong HBs were also observed between R256 and E376 residues in all systems for hMCADwt, increasing the average number of HBs per timeframe upon substrate binding (Table 12A/12B/12C). In addition, the proximity of the negatively charged E376 residue and the positively charged R256 residue also allowed us to infer the existence of strong electrostatic interactions between them in the human, as in the pig form of MCAD protein.


The most common mutation found in MCADD patients (c.985A>G) causes the exchange of a lysine by a glutamic acid in the position 304 (p.K304E) of the mature MCAD sequence. This mutation is associated with a severe phenotype characterized by metabolic crisis, triggered by intense exercise or intercurrent illness with fever episodes. The change of a positively charged residue (lysine) by a negatively charged residue (glutamic acid) in the vicinity of two other negatively charged amino acids (D300 and D346), results in a negative charge concentration at the dimer-dimer interface of the tetramer. Although some authors suggest that an excess of negative charges may be the responsible factor for the impairment of the tetramer assembly due an increased charge-repulsion, the molecular mechanisms by which this occurs are not fully characterized yet. Thus, further insights on the effect of the p.K304E mutation in the MCAD structure were obtained from pMCAD MD simulations in order to better understand the molecular basis of MCADD associated with the p.K304E variant. The MD simulations were performed during 100 ns after mutation of the crystallographic structure of Sus scrofa (pig) MCAD protein (PDB ID: 1UDY), in complex with FAD and the SCOA substrate. As mentioned in the chapter methods, three systems were simulated: APO (no substrate, no FAD), FAD (no substrate, with FAD), and LIPID (with substrate and FAD).

6.3.1 Analysis of protein stability

Root mean square deviation (RMSD) and radius of gyration (Rg). Unlike the MD simulations of the pMCADwt, in the study of the p.K304E variant of this protein, the RMSD values continuously increased for about half of the simulation time (Fig. 38A), only reaching more stable values around 60 ns. Although the RMSD mean values obtained at this plateau were higher (APO – 0.52 nm; FAD – 0.42 nm; LIPID – 0.45 nm) than those observed for pMCADwt (APO – 0.35 nm, FAD – 0.34 nm and LIPID – 0.31 nm) the tetrameric form of the protein was maintained and the presence of the FAD cofactor seemed to increase the stability of the protein’s structure. This is in agreement with the data obtained from the pMCADwt simulations, reinforcing the hypothesis that the presence of the cofactor is mandatory to improve the tetramer’s stability.
However, and unlike the simulations obtained for the pMCADwt, when the RMSD values of individual monomers of the p.K304E protein are plotted (Annex, Figs. S7 to S9), it is clear that in all systems there was not a homogeneous behavior among the monomers. The monomer A was more affected by the p.K304E mutation in the APO system but stabilized in the presence of the cofactor (FAD system), however in the presence of the substrate (LIPID system), the monomers A and B (dimer) were the ones mostly affected by the mutation.

Regarding the radius of gyration (Rg), the values obtained were found to increase during all simulations (APO – 3.58 nm; FAD – 3.58 nm; LIPID – 3.57 nm) and also in respect to the pMCADwt (APO – 3.49 nm; FAD – 3.50 nm and LIPID – 3.52 nm) indicating that although the spatial organization was generally maintained, the p.K304E tetramer structure becomes less compact (higher values) than the pMCADwt (Fig. 38B). This also seems to be in agreement with the proposed charge-repulsion model, in which the presence of excessive negative charges at the dimer-dimer interface would cause structural changes to minimize repulsion between the monomers.

**Protein secondary structure assessment.** Having previously discarded any stability issues due to the utilization of the GROMOS 53a6 parameter set, the impact of p.K304E mutation on the evolution of the secondary structure of pMCAD in function of time was also assessed by DSSP. The secondary structure remained stable throughout all simulations and comparable to the ones found for the crystallographic structure of pMCADwt (49% vs 54% for α-helices and 16% vs 15% β-sheets, respectively).

**Root mean square fluctuation (RMSF) and analysis of β-factors.** As no significant changes in the secondary structure were observed in DSSP calculations, the flexibility of the secondary
structure or the presence of abnormal motions in several structural motifs was characterized by analyzing the average atomic fluctuations of the protein residues (Fig. 39).

![Figure 39: Root mean square fluctuation (RMSF) values obtained for the p.K304E/pMCAD variant for APO (black), FAD (red) and LIPID (green) systems.](image)

The large majority of the residues’ fluctuations were found with mean values around 0.2 nm in all monomers and for all systems, being the largest fluctuations located in coils and in regions exposed to the solvent, as observed in pMCADwt.

However, it is worth noticing the increased of RMSF values around 180-200 residue’s number in monomers B, C and D, when compared with the ones obtained for these residues in the pMCADwt. The visual inspection of both pMCAD structures revealed that they were located in a β-sheet (in the vicinity of the 3’-phosphatoadenosine group of SCOA) in pMCADwt that loses its secondary structure upon introduction of the p.K304E mutation.

When comparing the β-factors obtained from MD simulations (Fig. 40) with the ones reported in the crystallographic structure, a similar pattern to the one described for the pMCADwt was observed with the internal α-helices at the dimer-dimer interface being much more stable when compared with the solvent-exposed domains.

Summarizing, the RMSD values described above indicate that p.K304E/pMCAD variant starts to stabilize later (60 ns) when compared with the pMCADwt (20 ns) demonstrating that the p.K304E/pMCAD variant is more unstable during the simulations. Nonetheless, the stabilizing
role of FAD cofactor in the structure is noteworthy and much clearer when comparing the RMSD values from both FAD and LIPID systems with the APO system. In addition, from the RMSF values and β-factors it is also possible to conclude that large residue fluctuations are located within the β-domain near of the 3’-phosphatoadenosine group of the substrate.

Figure 40: Comparison between the β-factors retrieved from the crystallographic structure of p.K304E/pMCAD variant (PDB ID: 1UDY-K304E) (left) and from MD simulations (right) of the p.K304E/pMCAD. High β-factor values represent more flexible regions (red) while low values indicate more stable regions (blue). Figures were created with MOE v2013.08.

6.3.2 Characterization of the catalytic pocket

As already stated for the pMCADwt enzyme, structural changes in monomer’s folding and dimer and tetramer assembly may impair a correct binding of the cofactor and/or the substrate to the binding cavity or even disrupt the formation of the catalytic pocket. Thus, it is imperative to assess the structural and dynamic functions of the catalytic pocket in order to better understand how the p.K304E mutation can specifically affect the MCAD enzymatic function.

Volume. The volume of the catalytic pocket was calculated through the VOIDOO\textsuperscript{71} software, as previously described for the wild-type enzyme. Also, as observed for the pMCADwt, a smaller volume distribution was observed for the p.K304E variant in the APO system (Fig. 41) with the 80% of the volumes within 95-265 Å\textsuperscript{3}. Interestingly, while the presence of FAD significantly increased the mean pocket volumes (80% of the volumes within 235-740 Å\textsuperscript{3}), the SCOA binding did not seem to have a significant effect on pocket volume distributions (80% of volumes within 290-650 Å\textsuperscript{3}).

- 61 -
The mean number of water molecules registered inside the catalytic pockets of the p.K304E variant was similar to the one reported above for the pMCADwt simulations, with both APO and FAD systems presenting higher water content (12) when compared with the LIPID system (6). Again, as previously observed, in order to enable the substrate’s entry into the catalytic pocket, water molecules must be displaced, thus creating an adequate hydrophobic environment where the enzymatic reaction occurs.

**FAD and SCOA binding regions.** From all the data gathered from the several MD simulations herein reported, the interaction between the FAD cofactor and the neighboring subunit is being considered as determinant to increase the cohesion between the monomers. Thus, in the p.K304E simulations, the hydrogen bonding between FAD and the neighboring monomer’s residues (Tables 13A/13B) were calculated to assess potential differences that could be induced by the mutation.

By the analysis of the tables above, it possible to see that the HB network between the FAD of one monomer and the adjacent monomer did not seem to be affected in a large extent in the p.K304E variant, when compared with the wild-type pMCAD enzyme.

**Table 13:** Hydrogen bonds between the FAD of one monomer and the neighboring monomer of the dimer unit in p.K304E/pMCAD – FAD system (A) and LIPID system (B).

<table>
<thead>
<tr>
<th>A</th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG_{HB}</td>
<td>&lt;N_{HB}&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>FAD:R281</td>
<td>1602</td>
<td>-23.8</td>
<td>0.93</td>
<td>1066</td>
</tr>
<tr>
<td>FAD:V350</td>
<td>1336</td>
<td>-23.4</td>
<td>0.52</td>
<td>515</td>
</tr>
<tr>
<td>FAD:G353</td>
<td>2716</td>
<td>-25.2</td>
<td>0.89</td>
<td>8046</td>
</tr>
<tr>
<td>FAD:N354</td>
<td>3142</td>
<td>-25.7</td>
<td>0.93</td>
<td>---</td>
</tr>
</tbody>
</table>
Regarding HBs between the FAD and the residues inside of the catalytic pocket (Annex, Tables S13 and S14), the cofactor interacted with fewer residues in the p.K304E variant than in the pMCADwt. In fact, when calculating the mean free energies of binding for FAD (calculated by g_mmpbsa) in the absence of substrate (FAD system) for the p.K304E variant, it had a ΔGbind of only -718 kJ.mol⁻¹, meaning a 19% decrease in affinity when compared with the FAD system of the pMCADwt structure (-886 kJ.mol⁻¹). Concomitantly, the free energy of binding for the FAD increased to -814 kJ.mol⁻¹ in the p.K304E variant in the presence of substrate (LIPID system), although representing a 20% decrease in the cofactor affinity when compared with the LIPID system of the pMCADwt (-1023 kJ.mol⁻¹).

Despite the loss of affinity of the FAD in respect to the wild-type enzyme in both FAD and LIPID systems, the SCOA binding still reinforces the FAD binding to the protein in p.K304E variant [-814 (LIPID system) vs -718 kJ.mol⁻¹ (FAD system)], as observed in the pMCADwt enzyme. Regarding the ΔGbind for the cofactor in each monomer (Table 13 and Fig. 42A), the loss of affinity is more pronounced in the monomers C and D when compared with the other subunits.

### Table 14: Free energy of binding (ΔGbind) in kJ.mol⁻¹ for the FAD and SCOA ligands in each monomer (A, B, C, D) of the p.K304E/pMCAD variant, in the LIPID system.

<table>
<thead>
<tr>
<th>Monomer</th>
<th>pMCADwt (kJ.mol⁻¹)</th>
<th>p.K304E/pMCAD (kJ.mol⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1103.7 ± 76.5</td>
<td>-874.2 ± 63.6</td>
</tr>
<tr>
<td>B</td>
<td>-833.1 ± 85.8</td>
<td>-893.3 ± 65.3</td>
</tr>
<tr>
<td>C</td>
<td>-1092.1 ± 65.7</td>
<td>-612.6 ± 117.1</td>
</tr>
<tr>
<td>D</td>
<td>-1063.4 ± 67.8</td>
<td>-877.7 ± 79.1</td>
</tr>
<tr>
<td>A</td>
<td>-1402.2 ± 178.7</td>
<td>-1674.2 ± 188.1</td>
</tr>
<tr>
<td>B</td>
<td>-793.1 ± 168.2</td>
<td>-1739.5 ± 228.9</td>
</tr>
<tr>
<td>C</td>
<td>-1519.6 ± 149.7</td>
<td>-718.4 ± 165.5</td>
</tr>
<tr>
<td>D</td>
<td>-1115.9 ± 190.9</td>
<td>-999.4 ± 121.3</td>
</tr>
</tbody>
</table>

The data represents the mean ± SD values obtained from the all MD simulation (LIPID system).

Surprisingly, the values of SCOA binding affinities were found to be higher in the p.K304E variant for the monomers A [p.K304E/pMCAD (-1674 kJ.mol⁻¹) vs pMCADwt (-1402 kJ.mol⁻¹)] and B [p.K304E/pMCAD (-1739 kJ.mol⁻¹) vs pMCADwt (-793 kJ.mol⁻¹)] than the ones registered in the LIPID system of pMCADwt (Table 13).
Figure 42: Free energy of binding ($\Delta G_{\text{bind}}$) in KJ.mol$^{-1}$ for the FAD and SCOA ligands in each monomer (A, B, C, D) of the p.K304E/pMCAD variant, in the LIPID system: Graph A, represent the $\Delta G_{\text{bind}}$ of FAD cofactor and graph B, represent the $\Delta G_{\text{bind}}$ of SCOA substrate. However, a closer inspection of the MD simulations revealed that the substrate in the monomers A and B left the binding pocket, becoming adsorbed in a region next to the pocket entrance, which contributed to increase the values of the binding energies for the substrate in these monomers. In respect to the other monomers, a decrease in the SCOA binding affinities was observed for monomers C [p.K304E/pMCAD (-718 KJ.mol$^{-1}$) vs pMCADwt (-1519 KJ.mol$^{-1}$)] and D [p.K304E/pMCAD (-999 KJ.mol$^{-1}$) vs pMCADwt (-1116 KJ.mol$^{-1}$)] when compared with the wild-type structure.

These affinities values were further corroborated by the results obtained from the contact frequencies per time frame (calculated by $g_{\text{contacts}}^{34}$; Annex, Tables S11 and S12) showing
that the contact and the number of residues that interacted with FAD (>50%) were lower in the p.K304E structure than in pMCADwt. This reduction was mostly located in the *sinister* (si) face of FAD (Fig43A and 43B) namely for the T168, G139, A140 and D143 residues, which may compromise the FAD binding to the protein.

Moreover, by comparing the FAD-protein interactions in the absence of substrate (FAD system) in both the pMCADwt and its p.K304E variant, a reduction of the number of HBs between the flavin ring and the protein was also observed in the p.K304E variant, which again may indicate a compromise of the cofactor affinity towards the protein (Fig. 29A vs Fig. 43A and Tables 12 vs 28). Nonetheless, the FAD was still anchored to the protein by an extensive HB network involving the phosphate and riboflavin moieties with the G141, R281, S142 and G353 residues of the protein. In agreement with the pMCADwt structure, the Y133 and Y375 aromatic residues in the p.K304E variant seemed to be in close contact with the flavin ring as well as with the H291 residue that seemed to maintain a relatively stable position next to the adenine ring of FAD.

However, and unlike the pMCADwt in the LIPID system, the FAD-protein HB network appeared to be even more reduced in the p.K304E variant, with the FAD cofactor being only anchored to the protein by its phosphate moiety through HBs with the G141, S142, R281 and G353 residues of the protein (Fig. 43B vs 29B and Table 13 vs 29). As the cofactor was never found to leave the binding pocket in all systems, it is assumed that HBs between the phosphates and the aromatic residues mentioned above (Y133, Y375 and H291) are the most important ones for anchoring FAD to the binding cavity in both pMCADwt and its p.K304E variant.

In respect to the CCOA substrate, changes in the contacts and in the HBs network patterns related to the wild-type pMCAD enzyme were observed and can be visualized in Figure 43C. While a HB network was formed with the 3’-AMP and the phosphate moieties of the substrate involving three different monomers (A, B and C), the terminal acyl group was located in an upper position within the pocket in monomers A and B (Fig. 43C) when compared with the monomers C and D (Fig. 43D), next to phenylalanine residues 254 and 252. In the pMCADwt, these residues were located next to the pantothenate moiety.

Although the substrate was still located correctly within the binding cavity in monomers C and D of the p.K304E/pMCAD variant, a large deformation of both catalytic pockets was observed, exposing the FAD and the SCOA to the polar water environment (Fig. 44, right). In monomers A and B, and despite the exit of the substrate from the catalytic pocket, its structure was similar to the one found for the wild-type pMCAD enzyme (Fig. 44, left).
Figure 43: Representation of the ligand interactions with the residues lining the catalytic pocket of p.K304E/pMCAD variant: (A) representation of the FAD interactions with the protein residues showing > 50% of contact frequency along the simulation [calculated by g_contacts (Table XX-SI)] in the FAD system and (B) in the LIPID system; (C) representation of SCOA interactions with the protein residues lining the catalytic pocket (>50%) of monomers A and B and finally, (D) in monomers C and D of the p.K304E/pMCAD variant. Aromatic residues (orange), non-polar residues (green), polar residues (blue); HBs are represented by dotted lines. The arrows represent HBs between the ligands and protein residues: HBs with the backbone’s residues (blue) and side-chain’s residues (green). Solvent accessible surface area is represented as a blue smudge. Figures were created with MOE v2013.08 from MD simulations.
6.4 Characterization of p.K304E variant of hMCAD (p.K304E/hMCAD)

Similarly to the workflow previously applied to the p.K304E/pMCAD variant, the p.K304E/hMCAD variant was also assessed through MD simulations during 100 ns, in order to provide insights that could explain and complement the experimental data obtained and also to additionally validate the results obtained in the p.K304E/pMCAD variant simulations. Thus, based on the initially reverted crystallographic structure from which hMCADwt was obtained and further mutagenized to introduce the p.K304E substitution, three simulations were performed: without any ligands (APO), with the FAD cofactor (FAD) and with both the FAD and octanoyl-CoA (CCOA).

6.4.1 Analysis of protein stability.

Root mean square deviation (RMSD) and radius of gyration (Rg). Unlike the MD simulations of the hMCADwt it was possible to observe that in its p.K304E variant the MD simulation of the LIPID system only seems to stabilize around 90 ns to a value close to 0.40 nm, while the APO and FAD simulations achieved stable RMSD values around 60 ns (0.49 and 0.41 nm respectively – Fig. 45A). However, from the Rg values (Fig. 45B) obtained for all systems it was possible to observe that while the Rg started to stabilize after 20 ns of the simulation time in the APO and FAD systems (3.58 nm for both systems), in the system containing both the FAD and CCOA (LIPID system) the Rg values were still increasing to higher values at the end of the simulation time (3.61 nm).
Figure 45: (A) Root mean square deviation (RMSD) and (B) radius of gyration (Rg) of p.K304E/hMCAD variant MD simulations against the reversed crystallographic coordinates (PDB 1EGC- K304): APO (black), FAD (red), and LIPID (green) systems.

Again, and as it was observed for the p.K304E/pMCAD variant, from the RMSD and Rg values for the individual monomers (Annex, S10-S12) of p.K304E/hMCAD variant, it was possible to observe distinct behaviors between them. In respect to RMSD, the monomer A seems to be the most affected by the p.K304E mutation in the APO and LIPID systems. However, the RMSD values in the FAD systems, as well as the Rg in all simulations, revealed that monomers B and D that are not part of the same dimer displayed the largest alterations when compared with the crystallographic coordinates. Thus, it seems that in these simulations at least one of the monomers in each dimer was more affected by the presence of the mutation than its partner.

**Protein secondary structure assessment.** The secondary structure, assessed by DSSP program in function of time, remained stable in all simulations with 50% of α-helices and 14% of β-sheets [against 53% and 14% in the original crystallographic structure of hMCAD (PDB ID: 1EGC), respectively]. Similarly to the results described for the p.K304E/pMCAD variant, the p.K304E mutation does not seem to have a significant effect on the secondary structure of the both p.K304E variant proteins.

**Root mean square fluctuation (RMSF) and analysis of β-factors.** In the p.K304E/hMCAD variant simulations, the same fluctuation pattern of the β-domain loop located between residues 180-200 was observed, however, achieving higher values than for the p.K304E/pMCAD variant (Figure 46). Interestingly, the baseline of the C-terminal domain (from the residue 240 to 396) for monomers C and D also seems to increase in the FAD system, which suggests that the dimer comprising these two monomers may be the most affected one by the presence of the p.K304E mutation in the presence of the cofactor (FAD system) in hMCAD protein.
Figure 46: Root mean square fluctuation (RMSF) values obtained for the p.K304E/pMCAD variant for APO (black), FAD (red) and LIPID (green) systems.

Regarding the β-factors (Fig. 47), the pattern obtained from the p.K304E/hMCAD variant simulations is similar to the ones previously obtained from the reversed crystallographic structure (PDB ID: 1EGC reversed – K304E) or from the MD simulations of p.K304E/pMCAD, with the internal α-helices at the dimer-dimer interface being the most stable motifs and the solvent-exposed β-domains showing the highest fluctuations.

Figure 47: Comparison between the β-factors retrieved from the reversed crystallographic structure of p.K304E/hMCAD variant (left) (PDB ID: 1EGC – K304E) (left) and from the structure obtained from MD simulations (right) of p.K304E/hMCAD. High β-factor values represent more flexible regions (red) while low values indicate more stable regions (blue). Figures were created with MOE v2013.08.
Altogether the results obtained indicate that, as observed for the p.K304E/pMCAD variant, the hMCAD tetramer bearing the p.K304E mutation show more instability when compared with the respective wild-type protein. Regarding the RMSD and Rg values, and likewise the hMCADwt protein, as well as in both pMCADwt and its p.K304E variant, the presence of the FAD cofactor seems to be determinant for an increased stability of whole hMCAD tetramer. Furthermore, the data obtained from RMSF and β-factors shows that in the p.K304E/hMCAD variant the largest fluctuations were located at the β-domain, the most solvent-exposed motifs in the tetramer. Unlike the p.K304E/pMCAD variant, in which one of the dimers seems to be more affected, the results herein reported for the p.K304E/hMCAD do not allow a direct confirmation of this observation. Regarding the RMSF values, and as mentioned above, the monomers C and D seem to be the mostly affected by the mutation however, from RMSD and Rg values obtained, it seems clear that, at least, one of the monomers in both dimers is affected by the presence of the mutation, namely monomers B and D. Thus, these results require further investigation.

6.4.2 Characterization of the catalytic pocket.

As observed in the p.K304E/pMCAD variant, the p.K304E mutation may not only affect the monomers and the whole dimer stability but, because it is present at the dimer-dimer interface, may compromise the correct tetramer assembly. In the case of the p.K304E/hMCAD variant, the above data obtained is not conclusive but suggests that at least one of the monomers in each dimer was affected by the presence of the mutation. Nonetheless, it is imperative to clarify the impact of such alterations in the structure and binding of the cofactor and substrate molecules. We achieved this by thoroughly analyzing the catalytic pocket.

**Volume.** By means of VOIDOO software, the volume distribution of the catalytic pockets of p.K304E/hMCAD variant was assessed by sampling the whole MD trajectory in intervals of 100 ps each. Interestingly, the results herein obtained show that, in this case, no significant difference existed between the volumes registered in the absence (80% of volumes within 110-310 Å³) or presence of the FAD cofactor (80% of volumes within 195-530 Å³ – Fig. 48) unlike it was observed for the p.K304E/pMCAD variant. Thus, the presence of the FAD cofactor does not seem to revert the structural modifications caused by the p.K304E mutation in the hMCAD protein. Nonetheless, the volumes of the catalytic pockets, in the presence of CCOA were shifted to higher values (80% of volumes within 390-830 Å³), which suggest that the binding of the substrate induced strong changes in the pocket structure (already seen in p.K304E/pMCAD variant simulations).
The mean number of water molecules inside the catalytic pockets revealed also to be different from the hMCADwt, as well as from the pMCADwt and its p.K304E variant. A higher content of water molecules inside of each catalytic pocket was found in the APO system (17 water molecules) when compared with the ones observed in the presence of FAD (FAD system; 7 water molecules). This means that, in the p.K304E/hMCAD variant, the pocket structure was rigid enough to allow higher water content in the APO system occupying the FAD binding site. Interestingly, in the LIPID system the presence of the substrate did not induce a desolvation of the binding pocket, being even registered a slight increase of the water content inside the pocket (9 water molecules) in respect to the FAD system. Thus, in the p.K304E/hMCAD variant the presence of CCOA seems not only to increase the volumes of the catalytic pockets but also to affect the ability of the substrate to bind and displace water molecules from the inside of the pocket, which may compromise the dehydrogenation reaction in which a hydrophobic environment is required inside of the catalytic pocket.6,7,11

**FAD and CCOA binding regions.** The FAD interactions with the protein’s residues as well as the cofactor binding affinities, already proved to be important for the structural stability of the dimer and the tetramer of p.K304E/pMCAD variant, were also assessed by calculating the contact frequencies per timeframe (assessed by g_contactsN74 – Annex S16, S17 and S20) and the HBs between the FAD and the neighboring monomer residues (Table 14). By these data, it was possible to observe the dramatically reduction of the HB interactions between the FAD and the neighboring monomer in respect to the hMCADwt and even in relation p.K304E/pMCAD variant, which seriously compromise the cohesion between the monomers.
Table 15: Hydrogen bonds between the FAD of one monomer and the neighboring monomer of the dimer unit in p.K304E/hMCAD – FAD system (A) and LIPID system (B).

<table>
<thead>
<tr>
<th>A</th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG_{HB}</td>
<td>&lt;N_{HB}&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>FAD:R281</td>
<td>1139</td>
<td>-22.9</td>
<td>1.26</td>
<td>3219</td>
</tr>
<tr>
<td>FAD:I350</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>FAD:G353</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>FAD:N354</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Monomer C</th>
<th>Monomer D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hbond</td>
<td>τ</td>
<td>ΔG_{HB}</td>
<td>&lt;N_{HB}&gt;</td>
<td>τ</td>
</tr>
<tr>
<td>FAD:R281</td>
<td>2029</td>
<td>-24.4</td>
<td>1.19</td>
<td>2164</td>
</tr>
<tr>
<td>FAD:I350</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>FAD:G353</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>9082</td>
</tr>
<tr>
<td>FAD:N354</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

τ, HB lifetime in ps; ΔG_{HB}, free energy of hydrogen bonding in kJ.mol⁻¹; <N_{HB}>, average number of HB per timeframe at 310K.

Likewise in the p.K304E/pMCAD variant, the contact frequencies between the residues lining the catalytic pocket and the cofactor are lower in the mutated human tetramer than in hMCADwt (Annex, Tables S6-S7). However, it is worth noticing that in the FAD system, despite a reduction of the contact frequencies per timeframe, the number of residues that interact with FAD is similar to the one registered for the hMCADwt. In the LIPID system, the reduction in contact frequencies was observed for residues that mainly interact with the flavin ring of FAD (especially in monomers A and C).

The free energy of binding ΔG_{bind} for the cofactor (calculated by g_mmpbsa75) in the FAD system increased with mean values of -663 kJ.mol⁻¹ (-844 kJ.mol⁻¹ for the hMCADwt) and -852 kJ.mol⁻¹ when in the presence of the substrate (-1016 kJ.mol⁻¹ for the hMCADwt) which is in translated in the decrease of the FAD binding affinity, similar to that observed for the pK304E/pMCAD variant.

Table 16: Free energy of binding (ΔG_{bind}) in kJ.mol⁻¹ for the FAD and CCOA ligands in each monomer (A, B, C, D) of the p.K304E/hMCAD variant, in the LIPID system.

<table>
<thead>
<tr>
<th></th>
<th>Monomer</th>
<th>hMCADwt (kJ.mol⁻¹)</th>
<th>p.K304E/hMCAD (kJ.mol⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FAD cofactor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-1103.7± 76.5</td>
<td>-1046.8 ± 69.6</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-833.1 ± 85.8</td>
<td>-1037.6 ± 78.5</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>-1092.1 ± 65.7</td>
<td>-900.8 ± 92.8</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>-1063.4 ± 67.8</td>
<td>-1079.7 ± 52.2</td>
<td></td>
</tr>
<tr>
<td><strong>CCOA substrate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-1402.2 ± 178.7</td>
<td>-730.9 ± 118.5</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-793.1 ± 168.2</td>
<td>-584.6 ± 133.4</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>-1519.6 ± 149.7</td>
<td>-978.4 ± 193.5</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>-1115.9 ± 190.9</td>
<td>-1142.8 ± 309.3</td>
<td></td>
</tr>
</tbody>
</table>

The data represents the mean ± SD values obtained from the all MD simulation (LIPID system).
Concerning the $\Delta G_{\text{bind}}$ for the cofactor in each monomer (Table 15 and Fig. 43A), the loss of affinity is more evident in the monomers A and B when compared with the other subunits. When compared the CCOA binding affinities of each monomer (Fig. 49) in the p.K304E/hMCAD variant, it is possible to verify a decrease of the binding energy values, in respect to the hMCADwt, in the monomers A and (Table 23) with a slightly increase in the monomer D (Table 23). A little decrease in the CCOA binding energy was also observed in monomer C, but not significantly.

**Figure 49:** Free energy of binding ($\Delta G_{\text{bind}}$) in KJ.mol$^{-1}$ for the FAD and CCOA ligands in each monomer (A, B, C, D) of the p.K304E/hMCAD variant, in the LIPID system: Graph A, represent the $\Delta G_{\text{bind}}$ of FAD cofactor and graph B, represent the $\Delta G_{\text{bind}}$ of SCOA substrate.

The above results together with the FAD binding affinities calculated per monomer (Fig. 49 and Table 23) led us to conclude, and despite the structural data that pointed out to be the
monomers B and D as the most affected by the mutation, the monomers A and B are the ones presenting the lower FAD and substrate affinities when compared with the monomers C and D. Additionally, the number of contacts between the substrate and the p.K304E/hMCAD protein as well as their frequencies are distinct between dimers, being the monomers A and B the most affected by the mutation (Annex, Table S20).

A visual inspection of the p.K304E/hMCAD structures resultant from the MD simulations show that the FAD is still able to maintain the HB network with the protein (above described for both MCAD wt proteins and for p.K304E/pMCAD variant) in the FAD system (Fig. 50A), involving the flavin ring and the ribityl group (riboflavin moiety) as well as the phosphate region and the ribose moiety of the adenine group (Annex, Tables S18-S19).

---

**Figure 50A:** Representation of the ligand interactions with the residues lining the catalytic pocket of p.K304E/hMCAD variant: (A) representation of the FAD interactions with residues showing > 50% of contact frequency along the simulation [calculated by \textit{g\_contacts}] (Table XX-SI) in the FAD system and (B) in the LIPID system. Aromatic residues (orange), non-polar residues (green), polar residues (blue); HBs are represented by dotted lines. The arrows represent HBs between the ligands and protein’s residues: HBs with the backbone’s residues (blue) and side-chain’s residues (green). Solvent accessible surface area is represented as a blue smudge. Figures were created with MOE v2013.08 from MD simulations.
Figure 50B: Representation of the ligand interactions with the residues lining the catalytic pocket of p.K304E/hMCAD variant: (C) representation of CCOA interactions with residues lining the catalytic pocket (>50%) of monomer C and finally, (D) representation of CCOA interactions with residues lining the catalytic pocket (>50%) of monomer D of the p.K304E/hMCAD variant. Aromatic residues (orange), non-polar residues (green), polar residues (blue); HBs are represented by dotted lines. The arrows represent HBs between the ligands and protein's residues: HBs with the backbone's residues (blue) and side-chain's residues (green). Solvent accessible surface area is represented as a blue smudge. Figures were created with MOE v2013.08 from MD simulations.

Moreover, the catalytic residue also interacts with the cofactor by means of an HB between the E376 of the main-chain and the ribityl moiety of FAD (Fig. 50A). However, in the LIPID system the FAD cofactor mainly establishes HBs with the protein through its phosphate moiety (Fig. 50B).

As the cofactor was never found to leave the binding pocket in all systems, it is assumed that HBs involving the phosphate moieties are the most important ones for anchoring FAD to the catalytic pocket. The interactions between the phosphate moiety of the substrate and the positively charged residues of the protein were found to be a common denominator in both p.K304E variants of pMCAD and hMCAD. Moreover, when the CCOA is inside of the catalytic pocket (monomer D, Fig. 50B - D) in the p.K304E/hMCAD variant is also able to establish HBs and electrostatic interactions which contribute to anchor the substrate to the protein.

However, a displacement of the lipid into an upper position within the pocket C (Fig. 50B - C) induced by the mutation turns the main interactions towards the 3'-phosphadenosine group of the CCOA substrate. Moreover, this displacement may induce misalignments between the catalytic residue E376 and the terminal acyl group of the substrate. Thus, the loss of binding affinities due to the p.K304E mutation in both the cofactor and substrate (Table 23) seems to be
the major cause for the catalytic impairment in the p.K304E variant, inducing changes in their interaction patterns and native HBs network that ultimately impairs the reaction.

Surprisingly, a direct observation of the four catalytic pockets revealed that three distinct events occurred within the 100 ns of simulation time in the LIPID system (not observed in the FAD system). While in the monomer D (Fig. 50B - C and Fig. 51 - left), the catalytic pocket remained relatively stable and the relative positions of both the FAD and CCOA were maintained, in the monomer C (Fig. 50B - D and Fig. 51 - middle), the substrate was found to exit the pocket (as mentioned for the monomers C and D of the p.K304E/pMCAD variant), occupying an upper position and forcing the cofactor to occupy a deeper position and a more packed conformation. Finally, in the case of the monomers A (not shown) and B (Fig. 51 - right), a severe distortion of the pocket was observed in which a rupture of the pocket wall between both molecules, exposing the flavin ring of the cofactor and the terminal acyl-chain of the substrate to a more polar environment.

Figure 51: Depiction of the catalytic pockets found in the p.K304E/hMCAD simulation. Left, catalytic pocket from monomer D, similar to the wild-type structure; middle, catalytic pocket from monomer C with the lipid bound outside the pocket; Right, catalytic pocket of monomer B (equal for B), presenting a large distortion that exposes both FAD and CCOA to the water environment while maintaining the relative positions of both molecules.

Discussion and Conclusion

The main aim of this thesis was to contribute to a better understanding of the molecular mechanisms underlying the MCADD associated with the most common variant of MCAD. With this goal in vitro and in silico approaches were carried out.

The in vitro studies aimed the structural and functional characterization of the p.K304E variant of the hMCAD enzyme as well as the study of the effect of a selected group of small molecules in its structure and function.
The *in silico* studies were performed namely by the structural characterization and dynamic behavior assessments of the p.K304E variant vs. the wt form of the enzyme and aimed to complement the experimental results. For this purpose, crystallographic data from the protein of interest is required. Five structures of the hMCAD enzyme are available in the Protein Data Bank (PDB) website, however, these crystallographic structures were obtained either with the protein modified with point mutations, namely as a E376G/T255E double mutant (PDB: 1EGC), or in complex with the ETF flavoprotein. Probably due to this fact, in many of the reports published on MCAD, the potential pathogenicity of the different variants found in MCADD patients is discussed based upon the crystallographic structure available for the porcine enzyme. In order to better select the crystallographic data to be used in the *in silico* studies, the first approach of the computational studies consisted in the comparison of the structure and the dynamic behavior of both pMCAD and hMCAD wt enzymes. This would also allow the validation of the MD method developed.

**Comparison between the pMCAD and hMCAD wt proteins.**

The crystallographic structures of pMCADwt (PDB ID: 1UDY) and of the hMCADwt, obtained by reversing the structure of the E376G/T255E variant of the hMCAD enzyme (PDB: 1EGC), were used and three systems built for both proteins: APO (no substrate, no FAD), FAD (no substrate, with FAD), and LIPID (with substrate and FAD). Unrestraint MD simulations were performed during 100 ns. The stability of both MCADwt proteins was firstly analyzed, followed by a description of the structure and dynamics of their catalytic pockets. The volumes of the catalytic pockets were characterized as well as the contact frequencies and the interactions between the ligands and the protein’s residues in both MCADwt proteins. Concomitantly, the free energies of binding for both FAD and substrates to these proteins were also assessed. Finally, a structural characterization of the active site in both pMCAD and hMCAD wt proteins was performed, focused in the involvement of some residues previously described in the literature to be important for MCAD catalytic properties.

The low RMSD values obtained from all simulations revealed that both pMCAD and hMCAD wt proteins were stable during the simulation, in respect to the corresponding crystallographic structures. The protein’s spatial organization and secondary structures of both proteins also remained stable throughout the simulation time and similar to the respective crystallographic structures. Thus, it is possible to conclude that both MCADwt proteins are suitable to be characterized by MD simulations and the GROMOS 53a6 parameter set can be successfully applied.
Regarding the different simulations performed for both pMCAD and hMCAD wt proteins, it is also possible to conclude that the presence of the FAD cofactor seems to contribute to the monomers’ and tetramer’s stability. The most flexible protein regions lied in coils, α–helices exposed to the solvent and β-sheets, while the α-helices of protein core, essential for the tetramer assembly and the residues located at enzyme’s active site were the most stable regions in both pMCAD and hMCAD wt enzymes.

Each catalytic pocket is formed by the two monomers of the same dimer, comprising the N- and C-terminal α-domains as well as the β-sheet domain. The pockets are very dynamic and flexible, being smaller in the absence of ligands in both MCADwt enzymes. The presence of the FAD and substrate lead to a pocket adjustment to larger volumes in both enzymes indicating that both ligands are important to define the volumes of the catalytic pocket. The considerable reduction of the water molecules inside the pocket in both proteins, upon substrate binding, is in agreement with the crystallographic structures and the reaction mechanism, in which a hydrophobic environment is required for the α,β–dehydrogenation reaction to occur.6,11

In both MCADwt proteins, the FAD is seen to be tightly bound inside the catalytic pocket by strong HB interactions with the protein’s residues, namely in the phosphate and adenine regions of the cofactor. Nevertheless, some aromatic residues in the vicinity of the flavin ring and the adenine group of FAD also seem to play an important role in the stability of the cofactor within the pocket. Moreover, the extended conformation of FAD allows it to establish HBs with residues of the neighboring monomer and thus contribute to monomer’s cohesion.

Upon the substrate binding (SCOA/CCOA), the HB network between the cofactor and the protein was reinforced, being translated in an increased FAD binding affinity towards the protein. Furthermore, the presence of the substrate also contributed to increase the cohesion between the monomers through the strengthening of the HB interactions between the FAD and the neighboring monomer. Similarly to FAD, the substrate (SCOA/CCOA) was mainly anchored to the protein by HB interactions, nonetheless, some aromatic residues near the terminal acyl group and the pantothenate moiety of the substrate proved to be important in its stability inside the catalytic pocket, in both pMCAD and hMCAD wt proteins.

As mentioned above, the geometry of the active site is determinant for the α,β–dehydrogenation reaction to occur (sensitive). The MD simulations of both pMCAD and hMCAD wt proteins revealed the existence of HBs between the carbonyl oxygen of the substrate with the 2′-OH of the ribityl group of FAD and the main-chain amide nitrogen of the catalytic residue (E376), in line with the crystallographic structure of pMCADwt (PDB: 1UDY). These HBs created a partial
positive charge in the Cα atom of the substrate, which makes the α–proton more susceptible to be removed, as described in various experimental studies.\textsuperscript{6,7,83,84}

Moreover, the side-chain conformations of some residues of the active site suffered alterations to better accommodate the substrate, being the most representative the changes in the side-chain conformations of the E99 and Y375 residues. The HBs between these two residues, observed in the presence of FAD, were immediately broken upon substrate binding. These observations were found in both proteins and contributed for the deeper binding of the substrate inside the pocket.

Unlike the crystallographic structure, but similar in both proteins, a different behavior was found in the catalytic residue E376 upon substrate binding. The carboxylate group (side-chain) of the E376 residue presented high mobility in the absence and presence of FAD, nevertheless, upon substrate binding its side-chain switches to the opposite side of the lipid’s Cα-Cβ bond, towards the water molecules found in the vicinity. In theory, this rearrangement does not favor the direct attack to the α-proton of the substrate by the E376 residue suggesting instead an indirect approach for the E376 residue to remove the α– proton of the substrate, as occurs in the Enoyl-CoA hydratase (ECH). This enzyme catalyzes the second step of mFAO and its two catalytic residues (E144 and E164) facilitate the protonation/deprotonation of the α-carbon atom of the substrate, mediated by water molecules.\textsuperscript{10} As occurs in ECH, we propose that the attack of the α-proton of the substrate may be mediated by water molecules, in which the E376 residue removes a proton from a water molecule forming a hydroxide ion (OH\textsuperscript{-}), which is a good nucleophile to attack the α-proton of the substrate. Moreover, the distances between the carboxylate group of the E376 residue and the Cα atom of the substrate (around 7.0 Å), found in both MCADwt enzymes, are compatible with our hypothesis.

Another finding which may support this theory was the interactions between the E376 and the R256 residues. The orientation of the E376 residue’s side-chain, observed upon substrate binding, shortened the distance between the carboxylate group of the catalytic residue and the guanidine group of the R256 residue and strong interactions between them were observed in both MCADwt proteins. Thus, the ability of the E376 residue to direct its side-chain to a favorable position for the removal of the α-proton of the substrate is reduced. Experimental and \textit{in silico} studies demonstrated that the replacement of the R256 residue by another amino acid, causes a complete enzyme inactivation, severely affecting the interactions with the substrate and the hydride transfer to the flavin ring of FAD.\textsuperscript{27,86} These results, support the idea that the R256 residue plays a crucial role on MCAD catalysis namely in the interactions between the enzyme
and substrate.\textsuperscript{27,86} Moreover, Zeng and Li emphasized the importance of the positively charged guanidine group of the R256 residue on MCAD catalysis. When the R256 residue was replaced by a lysine, a residue also positively charged, it was observed that the MCAD activity was unaffected by the mutation. Altogether, these findings suggest that the R256 residue is also important to stabilize, through electrostatic interactions, the unstable enolate transient species (negatively charged) formed after the $\alpha$-proton removal by the catalytic residue.

A comparison between the two models of the pig and human MCADwt enzymes revealed that both enzymes were stable in simulation and can thus be used for further studies on mutants and eventually studies of drug design. The quaternary structure stability and the subsequent analysis suggest that both proteins are essentially similar and that the reversion of the double mutant E376G/T255E of the hMCAD enzyme does not affect the structure of the protein neither its behavior in simulation. As such the crystallographic structures available for the pMCAD and the hMCAD proteins may both be used for the structural analysis of MCAD.

Nevertheless an important aspect must be taken always in consideration which is the fact that the available crystallographic structures, which are in complex with the substrate, consist in reality in a mixture of protein in complex with substrate, product or in an equilibrium state, which may introduce some artifacts. Separated simulations with the substrate, enolate transient species and the enoyl-CoA product with a good sampling, may be suggested to provide more insights in the active site structure and in the catalytic mechanism of MCAD protein.

\textit{In vitro and in silico studies of the p.K304E variant of MCAD enzyme}

The experimental studies performed comprised the heterologous expression of the hMCADwt and p.K304E variant proteins, followed by the purification and functional/structural characterization of the recombinant proteins. Furthermore, as performed for the pMCAD and hMCAD wt proteins, unrestraint MD simulations were performed during 100 ns using the p.K304E/pMCAD and p.K304E/hMCAD proteins, generated from the respective pMCADwt and reverted-hMCADwt structures mentioned above. The variant proteins were also simulated in the absence (APO system) and presence of FAD (FAD system) and in presence of both FAD and substrate (LIPID system).

As mentioned above, the most common variant found in MCADD patients results from the exchange of a lysine by a glutamic acid in the position 304 (p.K304E) of the mature MCAD. This alteration in the amino acid sequence leads to an excess of negative charges in the dimer-dimer interface of the protein, being likely related with the impairment of the tetramer assembly.\textsuperscript{19}
While integrating the computational data gathered from the p.K304E variants of the pMCAD and the hMCAD proteins, with the experimental results obtained, it becomes possible to interpret a probable cause for the reduction of the enzymatic activity and catalytic efficiency as well as the tetramer instability observed in vitro for the p.K304E/hMCAD variant in comparison with the wt enzyme. The mutation seems to affect the stability of the whole tetramer probably due to charge-repulsion forces at the dimer-dimer interface, however, without a significant loss of the secondary structure in both variants. The tetramer instability is corroborated by the experimental data wherein an altered oligomeric profile was observed for the p.K304E variant with higher content of aggregates, high molecular weight forms, dimers and monomers than the wt enzyme. Likewise, a lower resistance to proteolysis by trypsin of the p.K304E variant, in respect to the hMCADwt protein, was observed, suggesting a higher conformational flexibility of the variant protein. In addition, thermal instability of the p.K304E variant was also observed in the DSF, DLS and thermal enzymatic inactivation assays in which the p.K304E variant starts to denature and aggregate sooner and at lower temperatures than the hMCADwt leading to an earlier impairment of the enzymatic activity (see experimental results section).

From the analysis of MD simulations it was possible to conclude that the tetramer stability is improved in the presence of the FAD cofactor. These results are in line with other reports in literature\(^3,^6\) and also in agreement with the experimental data gathered herein. The supplementation with FAD of the lysis buffer during the purification of recombinant proteins (see experimental methods section) improved the oligomeric profile of the p.K304E variant (by increasing the tetrameric fraction and reducing the high molecular weight forms, dimers and monomers’ content), and also apparently lead to the rescue of the enzyme activity to values found for the wt form.

The thermal denaturation profile of both hMCAD and its p.K304E variant also improved dramatically in the presence of the FAD cofactor (DSF studies), being more pronounced for the p.K304E variant. Being the MCAD’s cofactor, FAD is expected to act as pharmacological chaperone and lead to protein stabilization. Nevertheless, in the case of the variant protein, this effect seems to be concentration-dependent as lower concentrations of FAD resulted in higher thermal stabilization. More studies are therefore required to better assess the stability role of FAD, namely in vitro, wherein the recombinant proteins can be expressed in a FAD or riboflavin supplemented culture medium followed by their functional and structural characterization.
Despite the different results obtained \textit{in silico} for the p.K304E/hMCAD and the p.K304E/pMCAD variant, a common denominator was found for both structures related to the stability of the whole tetramer (see analysis of protein stability – results chapter). It was observed that one of the dimer units is more affected than the other dimer – the dimer A/B in pMCAD and dimer C/D in the hMCAD. This was especially evident upon substrate binding.

In general, the decreased stability observed in one of the dimer units has repercussions in the binding affinity of both the FAD cofactor and the substrate in all monomers (except in monomers A and B of p.K304E/pMCAD) as well as in the structure of the catalytic pockets of the neighboring dimer unit [i.e. pMCAD (C/D) and hMCAD (A/B)]. These findings may be a likely explanation for the two transitions observed in the thermal denaturation (DSF) and aggregation profiles (DLS) of both the wt and the p.K304E variant of the hMCAD, being more pronounced in the case of the variant protein.

Moreover, and in particularly regarding the DSF results, when comparing the wt and p.K304E variant of hMCAD, it was possible to conclude that the presence of the mutation only affects the first thermal transition, meaning that the second stage may be related with a protein region that is unaffected by the mutation. Interestingly, in the presence of FAD the second transition is abolished supporting the above hypothesis. Altogether, these data suggest that the dimer observed \textit{in silico} to be more affected by the mutation, may correspond to the existence of the first thermal transition, while the most stable dimer, may represent the second transition, which is not significantly destabilized by the mutation.

The free energies of binding for the FAD increased in all simulations in both p.K304E/pMCAD and p.K304E/hMCAD proteins, which corresponds to a loss of affinity of the cofactor towards the protein when compared with the respective wt proteins. Furthermore, the loss of binding affinity by the substrate seems also to occur in the pMCAD and hMCAD variants. In both proteins, while the molecules of the cofactor remain bound inside the catalytic pockets, in some monomers the substrates’ molecules were found outside the pocket, bound to the protein surface. This resulted in lower binding energies [pMCAD (A/B) and hMCAD (C)] that prevented the correct assessment of the substrates’ binding affinities and may affect the enzymatic reaction. In the other monomers, the substrates’ molecules were found inside of the catalytic pockets [pMCAD (B/C) and hMCAD (A/B)] but the structure of those pockets was affected, distorted beyond limits occurring a disruption of the lateral pocket walls (Fig. 44 and 51), exposing both the cofactor and the substrate to the polar water environment.
In the monomers where the catalytic pockets maintain their structure, and despite the loss of binding affinity described for both ligands, it would still be possible for the lipid to acquire a favorable position inside the pocket facilitating the catalysis of the α,β-dehydrogenation. However, in the case of a distortion or disruption of the catalytic pockets, also observed in both MCAD p.K304E variants, the enzymatic reaction may be affected due: i) to the exposing of a large portion of the cofactor and/or substrate to the polar water environment, reducing the catalytic residue E376 desolvation, thus decreasing its pKa to a value that precluded proton removal from the substrate (ref), ii) to the impairment of the correct binding of the substrate and increment of the distances between the flavin ring, the Ca-Cβ bond and the catalytic residue to non-optimal values; and/or iii) the incorrect coupling of the FAD cofactor, impairing the electron transfer to the ETF protein. Overall, the changes in the FAD and substrate binding affinities and the instability found in the structure of the catalytic pockets may explain the decrease of the specific enzymatic activity and catalytic efficiency of the variant protein observed in our experimental results. It may also elucidate why, when compared to wt form of the protein, the p.K304E variants still retains about 46% of its residual activity which may correspond to the catalytic pockets that do not disrupt and thus keep their functionality.

One of our aims was also the search for small molecules which could be used as chaperones to restore the structure and consequently the function of MCAD. Besides the improving of the thermal stability mentioned above for the FAD cofactor, which supports its use as a pharmacological chaperone, promising results were also obtained with glycerol and TMAO tested, among other compounds, as chemical chaperones for the p.K304E variant.

Altogether, the data obtained from both experimental and computational studies allowed the characterization and validation of in vitro and in silico models of the p.K304E variant, permitting to go forward for drug design studies which will contribute to the discovery and development of compounds that may constitute potential pharmacological strategies for the treatment of MCAD deficiency.

**Future Perspectives**

As mentioned, MCAD deficiency is the most common inborn metabolic error of the mitochondrial fatty acid β-oxidation pathway. In the absence of catabolic stress, individuals can remain asymptomatic until adulthood but fever episodes, intense exercise or prolonged fasting may trigger metabolic crisis with hypoketotic hypoglycemia and lethargy. The inclusion of MCAD deficiency in the group of inherited metabolic disorders identified by newborn screening reduced considerably the high rate of morbidity and mortality associated with this disorder.
especially in childhood since it allows the early identification and implementation of preventive measures. Yet, cases of sudden death and coma have been reported in controlled MCAD patients. Dietary control of the patients is the only available therapeutic approach in MCADD. These evidences highlight the urgency for the search for pharmacological strategies for MCADD treatment.

Most of MCADD patients are homozygous for the most common mutation (c.985A>G; p.K304E), nevertheless, an increasing number of compound heterozygous associated with p.K304E variant have been observed. These individuals carry the most common mutation in one allele and a rare mutation in the other allele. It is still unknown the mechanism by which these two allelic hMCAD variants assemble and how structurally and functionally different they are from the original individual proteins. Interallelic complementation (IC) has been reported in other genetic metabolic disorders but not in MCADD, being its elucidation essential to provide insights into the molecular mechanism of the disease and to the design of a drug that could be used for treating both homozygous and compound heterozygous individuals. Thus, the search of compounds with pharmacological interest is a challenge in MCADD.

In silico methods represent an innovation in the study of MCADD. Molecular dynamics (MD) simulations represent a powerful tool to provide insights in the molecular mechanism of the disease, complementing of the in vitro characterization of the behavior of novel mutants namely the ones found in Portugal (p. Y48C, p.D143V, p.G377V and) and of hybrid MCAD variant proteins.

Most of the variants causing MCADD, are the result of missense mutations and, like the p.K304E variant, are characterized by protein misfolding and/or misassembling with enzyme’s loss-of-function. Thus, this disease is considered a conformational disorder and a possible candidate for the treatment with small molecules (<500 Da). These compounds may act as chaperones promoting the stabilization and rescue of misfolded proteins preventing its early degradation.

We have gathered some promising preliminary data on the stabilization of hMCAD variants by the natural cofactor (FAD) and some osmolytes as glycerol and TMAO which will be helpful in the understanding of how MCAD variants may be stabilized and rescued.

The data collected in the MD studies herein reported, and in particular the characterization of the hMCADwt and p.K304E variant in silico models, will be highly useful to the implement a structure-based drug discovery (SBDD) strategy based on molecular docking. By screening large libraries of commercially available compounds (NCI and Zinc) including drug repurposing (Drug Bank) libraries, a series of hit compounds (potential stabilizers of MCAD) will be generated wherein the best will be further tested in vitro using the developed recombinant proteins.
Finally, and in an attempt to filter these compound databases, a pharmacophoric model will be generated to obtain molecules more suitable for the target variant of MCAD (homotetrameric or hybrid).

Hence, the future work will be continued following *in silico* and *in vitro* approaches aiming to: 1) identify small molecules as lead compounds for the development of a pharmacological therapy for MCADD; and 2) elucidate whether the IC is a mechanism underlying MCADD and thus essential in the discovery of common therapeutic strategies for homozygous and compounds heterozygous.
The work developed in this master thesis originated:

**Papers in conference proceedings without scientific refereeing**


**Posters Communications**


Bonito CA, Ventura FV, Guedes RC. *Structural and dynamic behavior of medium-chain Acyl-CoA dehydrogenase by molecular dynamics simulations*. 10th SPDM International Symposium, Cascais, Portugal, 20-21 March 2014;


**Oral Communications**

Bonito CA, Ventura FV, Guedes RC. Selected Short Oral E-poster Presentation of *Structural and dynamic behavior of Medium-Chain acyl-CoA dehydrogenase by molecular dynamics simulations*. 10th SPDM International Symposium, Cascais, Portugal, 20-21 March 2014;

Bonito CA, Leandro P, Ventura FV, Guedes RC. *A new approach to elucidate the molecular mechanisms of Medium-Chain Acyl-CoA dehydrogenase deficiency (MCADD)*. EJBCE II, Segundo
Encontro de Jovens Investigadores de Biologia Computacional Estrutural, Lisboa, Portugal, 18-19 December 2014.


Bonito CA, Leandro P, Ventura FV, Guedes RC. *Insights in the molecular mechanisms underlying the pathogenicity of the most common variant (p.K304E) causing medium-chain acyl-CoA dehydrogenase deficiency.* TheoBio2015, 7th International Theoretical Biophysics Simposyum, Cagliary, Italy, 7-12 June 2015 (accepted).

**Awards**


**Publications**

Bonito CA, Leandro P, Ventura FV, Guedes RC. Insights on Medium-chain acyl-CoA dehydrogenase structure by Molecular Dynamics simulations. (submitted);

Bonito CA, Leandro P, Ventura FV, Guedes RC. Characterization and Structural Dynamics of p.K304E mutation in Medium-chain acyl-CoA dehydrogenase deficiency by in silico methods. (in preparation);
References


*Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, *2015*.


