

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOQUÍMICA E IMUNOLOGIA

SAMUEL LEITE GUIMARÃES

**NahK/NahL: estrutura, função e mecanismo de ação de um complexo
macromolecular**

Belo Horizonte

2015

SAMUEL LEITE GUIMARÃES

**NahK/NahL: estrutura, função e mecanismo de ação de um complexo
macromolecular**

Tese apresentada ao Departamento de Bioquímica e Imunologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Doutor em Bioquímica.

Área de concentração: Estrutura e função de proteínas e peptídeos.

Orientador: Ronaldo Alves Pinto Nagem

Belo Horizonte

2015



Universidade Federal de Minas Gerais
 Curso de Pós-Graduação em Bioquímica e Imunologia ICB/UFMG
 Av. Antônio Carlos, 6627 – Pampulha
 31270-901 – Belo Horizonte – MG
 e-mail: pg-biq@icb.ufmg.br (31)3409-2615



ATA DA DEFESA DA TESE DE DOUTORADO DE SAMUEL LEITE GUIMARAES. Aos quinze dias do mês de dezembro de 2015 às 13:00 horas, reuniu-se no Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, a Comissão Examinadora da tese de Doutorado, indicada *ad referendum* do Colegiado do Curso, para julgar, em exame final, o trabalho intitulado ""NahK/NahL: Estrutura, função e mecanismo de ação de um complexo macromolecular"", requisito final para a obtenção do grau de Doutor em Ciências: Bioquímica. Abrindo a sessão, o Presidente da Comissão, Prof. Ronaldo Alves Pinto Nagem, da Universidade Federal de Minas Gerais, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa do candidato. Logo após a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações: Dr. Carlos Basílio Pinheiro (Universidade Federal de Minas Gerais), aprovado; Dr. Carlos Renato Machado (Universidade Federal de Minas Gerais), aprovado; Dr. Antônio José da Costa Filho (Universidade de São Paulo), aprovado; Dr. Adriano Andricopulo (Universidade de São Paulo), aprovado; Dr. Ronaldo Alves Pinto Nagem - Orientador (Universidade Federal de Minas Gerais), aprovado. Pelas indicações o candidato foi considerado APROVADO. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente da Comissão encerrou a reunião e lavrou a presente Ata que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 15 de dezembro de 2015.

Dr. Carlos Basílio Pinheiro (UFMG)

Dr. Carlos Renato Machado (UFMG)

Dr. Antônio José da Costa Filho (Universidade de São Paulo)

Dr. Adriano Andricopulo (Universidade de São Paulo)

Dr. Ronaldo Alves Pinto Nagem - Orientador (UFMG)

Prof^a Leda Quercia Vieira
 Coordenadora do Curso de Pós-Graduação
 em Bioquímica e Imunologia
 ICB - UFMG

Aos meus pais, Custódio e Marizete. Essa conquista também pertence a eles!

AGRADECIMENTOS

Agradeço imensamente aos meus pais por todo amor e apoio incondicional, e que mesmo distantes geograficamente, foram extremamente presentes e importantes nessa longa caminhada. Muito obrigado também pelo enorme incentivo e por jamais medirem esforços em lutar junto comigo para que eu conseguisse alcançar meus objetivos.

À minha irmã, sobrinha, demais familiares e grandes amigos do Triângulo mineiro, pelo carinho, apoio e compreensão pela minha ausência em inúmeros momentos importantes.

Aos meus padrinhos pelo apoio e por sempre torcerem e vibrarem junto comigo por minhas conquistas.

Ao meu orientador Ronaldo pelos inúmeros ensinamentos, oportunidades e por auxiliar tanto em questões profissionais como pessoais.

Aos meus amigos de Belo Horizonte pelo apoio, amizade e risadas sem as quais os dias aqui não seriam tão bons.

Aos colegas do Laboratório de Biologia Estrutural pela paciência e apoio em vários experimentos, bem como pelos bons momentos compartilhados. Todos me ensinaram inúmeras coisas e sou muito grato.

A todos os laboratórios, professores, alunos e técnicos do Departamento de Bioquímica e Imunologia que contribuíram de alguma forma para a realização deste trabalho.

Ao Laboratório Nacional de Luz Síncrotron pelo acesso as linhas de luz para os experimentos utilizando raios-X, assim como ao Laboratório Nacional de Biociências pelo acesso ao ROBOLAB.

Ao Cristian P. Whitman, William H. Johnson Jr. e demais alunos de seus laboratórios na Universidade do Texas em Austin – Estados Unidos, pelo apoio durante meu Doutorado Sanduíche.

Aos membros da banca examinadora, titulares e suplentes.

Às agências de fomento e instituições pelo apoio financeiro: FAPEMIG e VALE S/A.

“Bankai...” disse Kurosaki Ichigo.

“Tatakae!” gritou Eren Yeager.

“O mundo é cruel!” afirmou Mikasa Ackerman.

A verdade é que para viver há uma luta constante. Se você perder, você morre. Se você ganhar, você vive. Se você não lutar, não há como ganhar.

RESUMO

GUIMARÃES, S. L. **NahK/NahL: estrutura, função e mecanismo de ação de um complexo macromolecular**. Tese (Doutorado) – Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

Naftaleno é um composto tóxico bastante estudado dentre os hidrocarbonetos aromáticos policíclicos e sua degradação pela bactéria *Pseudomonas putida* G7 tem sido o foco de diversas pesquisas. Nesse organismo, as enzimas NahK (4-oxalocrotonato descarboxilase, 28,4 kDa) e NahL (vinilpiruvato hidratase, 27,9 kDa) pertencem à via inferior de degradação do naftaleno e ocorrem *in vivo* como um complexo que catalisa uma interessante sequência de reações utilizando um intermediário instável, levantando uma série de questões mecânicas e estruturais. Nosso objetivo foi caracterizar cinética e estruturalmente as enzimas NahK, NahL e seu complexo. Os genes *nahK* e *nahL* foram subclonados separadamente no vetor pET28a-TEV, e também subclonados juntos no vetor pETDuet-1 para coexpressão em *Escherichia coli* BL21(DE3). NahK e seu complexo com NahL foram purificados por cromatografia de afinidade ao níquel e por gel filtração. As amostras proteicas foram analisadas por Espalhamento de Luz Dinâmico, Dicroísmo Circular, e seus parâmetros cinéticos foram obtidos. O complexo também foi analisado por Espalhamento de Raios-X a Baixo Ângulo. Amostras de NahK e do complexo, em sua forma apo ou ligada a análogos de substrato, foram cristalizadas e dados de Difração de Raios-X foram coletados. NahK, tanto em forma isolada quanto complexada, e NahL exibem valores de k_{cat}/K_M próximos de $10^7 \text{ M}^{-1}\text{s}^{-1}$. As estruturas cristalinas de NahK em sua forma apo, ligada a Mg^{2+} e a análogos de substrato permitiram propor um mecanismo de reação envolvendo um domínio tampa. Resultados de SAXS mostram que o complexo se comporta como uma partícula de 275 kDa, e sua estrutura cristalina sugeriu uma montagem com 10 subunidades. NahK e NahL exibem as cavidades de seus respectivos sítios-ativos uma para a outra em uma estrutura quaternária grande, elegante e funcional. Juntas, essas enzimas realizam importantes passos na degradação do naftaleno em *P. putida* G7. Este trabalho provê um profundo conhecimento estrutural e funcional desse interessante complexo proteico.

Palavras-chave: *Pseudomonas putida*, naftaleno, cristalização, mecanismo de reação.

ABSTRACT

GUIMARÃES, S. L. **NahK/NahL: structure, function and mechanism of action of a macromolecular complex.** Thesis (doctorate degree) – Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

Naphthalene is a toxic and well-studied polycyclic aromatic hydrocarbon. Its degradation by the bacterium *Pseudomonas putida* G7 is the focus of numerous studies. In this organism, the enzymes NahK (4-oxalocrotonate decarboxylase, 28.4 kDa) and NahL (vinylpyruvate hydratase, 27.9 kDa) belong to the lower pathway of naphthalene degradation. They exist *in vivo* as a protein complex, catalyzing an interesting sequence of reactions with an unstable intermediate that raises a number of mechanistic and structural questions. Our aim was to characterize kinetically and structurally the enzymes NahK, NahL and their complex. The *nahK* and *nahL* genes were subcloned separately into pET28a-TEV vector for expression, and also subcloned together into pETDuet-1 vector for co-expression in *Escherichia coli* BL21(DE3) cells. NahK and its complex with NahL were purified by Immobilized Metal Ion Affinity and Size-Exclusion Chromatographies. The protein samples were analyzed by Dynamic Light Scattering, Circular Dichroism, and the steady-state kinetic parameters for their natural substrates were obtained. The NahK/NahL complex was also analyzed by Small-Angle X-Ray Scattering. NahK and NahK/NahL samples in their apo or ligated forms with different substrate analogues were crystallized and X-Ray Diffraction data were collected. NahK expressed alone and co-expressed with NahL was purified to homogeneity and both enzymes shown a k_{cat}/K_M values up to $10^7 \text{ M}^{-1}\text{s}^{-1}$. The crystalline structures of NahK in apo and ligated forms with Mg^{2+} and substrate analogues allowed the proposition of a reaction mechanism involving a lid domain. SAXS measurements suggested a particle of 275 kDa for the NahK/NahL complex, and the crystalline structure of this complex suggested a heterodecamer assembly. NahK and NahL face their active site cavities to each other in a large, elegant and functional quaternary structure. Together, these enzymes perform important steps during naphthalene degradation in *P. putida* G7. This work provided a deep structural and functional knowledge of this interesting protein complex.

Keywords: *Pseudomonas putida*, naphthalene, crystal structure, reaction mechanism.

LISTA DE FIGURAS

Figura 1: Estrutura química dos 16 HAP listados pela EPA como prioritários para monitoramento no meio ambiente.	3
Figura 2: Mecanismos de ativação dos HAP, formação de adutos e danos ao DNA.	5
Figura 3: Representação dos genes <i>nah</i> do plasmídeo NAH7.	11
Figura 4: Via de degradação do naftaleno em <i>P. putida</i> G7.	12
Figura 5: <i>Meta</i> e <i>orto</i> -clivagem do catecol.	14
Figura 6: Comparação dos diferentes produtos gerados pela degradação de naftaleno, fenantreno e antraceno pelas enzimas Nah codificadas pelo plasmídeo NAH7.	14
Figura 7: Via de <i>meta</i> -clivagem de catecóis com diferentes grupos substituintes a partir de diferentes compostos aromáticos.....	16
Figura 8: Reação de descarboxilação e hidratação catalisadas pelas enzimas NahK e NahL, respectivamente.....	20
Figura 9. Representação da estrutura da enzima FAH obtida a partir do código PDB 1QCN..	22
Figura 10. Representação em estéreo (<i>cross-eyed</i>) do barril- β incompleto da enzima FAH. .	23
Figura 11. Esquema das construções finais nos vetores de expressão.	33
Figura 12. Substrato de NahK (1) e de NahL (5); e algumas moléculas análogas selecionadas como ligantes para NahK.	43
Figura 13. Amplificação do gene <i>nahK</i> de <i>P. putida</i> G7 para clonagem no vetor pET28a-TEV.	46
Figura 14. Análise por SDS-PAGE da expressão de NahK separadamente de NahL.	47
Figura 15. Cromatograma e análise por SDS-PAGE da primeira etapa de purificação de 6xHis-NahK por IMAC.....	49
Figura 16. Cromatograma e análise por SDS-PAGE da segunda etapa de purificação de NahK por IMAC.	51
Figura 17. Cromatograma e análise por SDS-PAGE da terceira etapa de purificação de NahK por SEC.	52
Figura 18. Alinhamento das sequências de resíduos de aminoácidos de NahK e Xyll.	54
Figura 19. Distribuição de tamanhos de partícula de NahK recombinante obtido por DLS. ...	55
Figura 20. Dicroísmo circular de NahK na ausência de NahL.....	57
Figura 21. Curvas de absorvância da reação da enzima 4-OT ao converter 2HM em 4-oxalocrotonato.....	58
Figura 22. Curvas de absorvância da reação da enzima NahK ao descarboxilar 4-oxalocrotonato na ausência de NahL.....	59

Figura 23. Velocidades iniciais da reação de NahK na ausência de NahL em diferentes concentrações de substrato e de Mg^{2+} .	60
Figura 24. Influência do pH na cinética de NahK na ausência de NahL.	61
Figura 25. Cristais de NahK e padrão de difração de raios-X.	63
Figura 26. Representação das moléculas de NahK empacotadas nas células unitárias dos cristais provindos das condições CS1-16 (a) e CS1-22 (b).	65
Figura 27. Representação das estruturas secundárias e terciária de NahK.	69
Figura 28. Alinhamento de sequências de aminoácidos (a) e das estruturas terciárias (b) de NahK, HpcG e MhpD.	71
Figura 29. Comparação dos diferentes domínios N-terminal das enzimas da superfamília FAH.	74
Figura 30. Estrutura quaternária de HpcG.	75
Figura 31: Representação em estéreo (<i>cross-eyed</i>) da cadeia principal dos domínios N-terminal e barril- β de NahK.	76
Figura 32: Representação de alguns dos resíduos responsáveis pelas interações entre a fita 2, hélices E e H e uma volta do domínio barril com resíduos do domínio N-terminal em NahK.	77
Figura 33: Modelo de superfície de NahK exibindo em destaque a grande cavidade que forma o sítio ativo.	80
Figura 34: Mapas de densidade eletrônica ilustrando a posição do metal, moléculas de água e ligantes no sítio ativo dos modelos de NahK.	82
Figura 35: Representação em estéreo (<i>cross-eyed</i>) de alguns resíduos no sítio ativo de NahK dos modelos <i>Apo</i> (a) e <i>Mg</i> (b).	83
Figura 36: Representação em estéreo (<i>cross-eyed</i>) de alguns resíduos e dos ligantes no sítio ativo de NahK, modelos <i>AKG</i> (a) e <i>ACT</i> (b).	86
Figura 37: Representação em estéreo (<i>cross-eyed</i>) de alguns resíduos e dos ligantes no sítio ativo de NahK, cadeia A do modelo <i>AA</i> (a) e cadeia A do modelo <i>2OA</i> (b).	89
Figura 38. Representação em estéreo (<i>cross-eyed</i>) de alguns resíduos e dos ligantes no sítio ativo de NahK, cadeia B do modelo <i>2OA</i> .	90
Figura 39. Representação em 'worm' do modelo <i>Mg</i> e da cadeia A do modelo <i>2OA</i> .	91
Figura 40: Via de degradação por metaclivagem do catecol e intermediários incertos até meados de 1970.	93
Figura 41. Mecanismo de descarboxilação do oxaloacetato catalisada por metal.	94
Figura 42: Conformação favorável para descarboxilação.	95
Figura 43: Deslocação de cargas negativas em um ânion carboxilato genérico (a) e seu respectivo ânion carboxilato vinílico (b).	96
Figura 44: Oxaloacetato (1) e seu correspondente vinílico 4-oxalocrotonato (2).	96
Figura 45. Mecanismo proposto para a descarboxilação de 4-oxalocrotonato por NahK.	98

Figura 46. Amplificação do gene <i>nahL</i> de <i>P. putida</i> G7 para clonagem no vetor pET28a-TEV.	99
Figura 47. Cromatograma e análise por SDS-PAGE da purificação de 6xHis-NahL por IMAC.	100
Figura 48: Alinhamento das sequências de resíduos de aminoácidos de NahL e XylJ.	101
Figura 49: Amplificação do gene <i>nahK</i> e <i>nahL</i> de <i>P. putida</i> G7 para clonagem no vetor pETDuet-1.....	102
Figura 50: Análise por SDS-PAGE da coexpressão de NahK e NahL.	104
Figura 51: Cromatograma e análise por SDS-PAGE da primeira etapa de purificação do complexo NahK/NahL.....	105
Figura 52: Cromatograma e análise por SDS-PAGE da segunda etapa de purificação do complexo por IMAC.....	106
Figura 53: Cromatograma e análise por SDS-PAGE da etapa de purificação do complexo por SEC.....	107
Figura 54: Distribuição de tamanhos de partícula do complexo recombinante NahK/NahL obtido por DLS.....	108
Figura 55: Dicroísmo circular do complexo NahK/NahL.	109
Figura 56: Curvas de SAXS e processamento de dados.	111
Figura 57: Modelo de baixa resolução obtido por SAXS sem imposição de simetria.....	113
Figura 58: Modelo de baixa resolução obtido por SAXS com simetria imposta 'P5'	114
Figura 59: Modelo de baixa resolução obtido por SAXS com simetria imposta 'P52'	115
Figura 60: Superposição do modelo cristalográfico do homodecâmero de HpcG e o modelo de baixa resolução do complexo NahK/NahL.	116
Figura 61: Curvas de absorvância da reação da enzima NahK ao descarboxilar 4-oxalocrotonato na presença de NahL.	117
Figura 62: Velocidades iniciais da reação de NahK na presença de NahL em diferentes concentrações de substrato e de Mg ²⁺	118
Figura 63: Influência do pH na cinética de NahK na presença de NahL.....	119
Figura 64: Curvas de absorvância da reação da enzima NahL ao hidratar o substrato 2-hidroxi-2,4-pentadienoato.	120
Figura 65: Influência do pH na cinética de NahL na presença de NahK.....	121
Figura 66: Cristais (a) e padrão de difração de raios-X (b) do complexo NahK/NahL.....	123
Figura 67: Representação dos oito heterodecâmeros de NahK e NahL empacotados na célula unitária.	125
Figura 68: Representação das estruturas secundárias e terciária de NahL.	127
Figura 69: Modelo de superfície de NahL exibindo em destaque a cavidade do sítio ativo.	128
Figura 70: Representação em 'worm' do modelo de NahL.	129

Figura 71: Mecanismo de ação de XylJ proposto por Miller a partir da estrutura de MhpD.	131
Figura 72: Alinhamento de sequência primária de NahL e MhpD.	132
Figura 73: Modelo de superfície das subunidades formadoras dos heterodecâmeros de NahK e NahL.	134
Figura 74: Modelo de superfície do complexo evidenciando um monômero de NahK (azul), um monômero de NahL (cinza) e suas respectivas cavidades catalíticas (vermelho).	135
Figura 75: Modelo de superfície do complexo representando a posição de dois monômeros de NahK (em azul) e de dois monômeros de NahL (em cinza).	136
Figura 76: Elementos estruturais responsáveis pela formação do anel pentamérico.	137
Figura 77: Ajuste da curva experimental do complexo NahK/NahL em solução obtida por SAXS com a curva de espalhamento teórico do heterodecâmero cristalográfico.	138

LISTA DE TABELAS

Tabela 1: Análise de identidade das proteínas da via superior e inferior de degradação do naftaleno de <i>P. putida</i> G7 e de proteínas com estrutura tridimensional já elucidada depositada no PDB.	19
Tabela 2. Antibióticos utilizados para a seleção dos respectivos plasmídeos.	27
Tabela 3. Oligonucleotídeos sintéticos utilizados como iniciadores para amplificação dos genes <i>nahK</i> e <i>nahL</i>	29
Tabela 4. Reagentes e quantidades utilizadas nas reações de PCR.	31
Tabela 5. Programa para amplificação dos genes <i>nahK</i> e <i>nahL</i>	31
Tabela 6. Coeficientes de extinção molar e massas moleculares das proteínas recombinantes.	36
Tabela 7. Materiais e quantidades utilizadas na preparação dos géis para SDS-PAGE.	37
Tabela 8. Parâmetros de cinética para NahK na ausência de NahL utilizando-se diferentes concentrações de magnésio.....	60
Tabela 9. Parâmetros de cinética para NahK na ausência de NahL utilizando-se diferentes valores de pHs em tampão fosfato de sódio.	61
Tabela 10. Sumário das estatísticas dos conjuntos de dados obtidos a partir dos cristais de NahK crescidos na condição CS1-16.....	66
Tabela 11. Sumário das estatísticas dos conjuntos de dados obtidos a partir dos cristais de NahK crescidos em condição melhorada de CS1-16.	67
Tabela 12. Proteínas com estrutura conhecida pertencentes à superfamília FAH e similaridades com NahK.....	70
Tabela 13: Relação entre número de resíduos de aminoácidos que formam o domínio barril e a área de superfície enterrada entre os domínios barril dos monômeros.....	78
Tabela 14: Parâmetros de cinética para NahK na presença de NahL utilizando-se diferentes concentrações de magnésio.....	118
Tabela 15: Parâmetros de cinética para NahK na ausência de NahL utilizando-se diferentes valores de pHs em tampão fosfato de sódio.	119
Tabela 16: Parâmetros de cinética para NahL na presença de NahK utilizando-se diferentes valores de pHs em tampão fosfato de sódio.	121
Tabela 17: Sumário das estatísticas do conjunto de dados obtido a partir de um cristal do complexo NahK/NahL.....	124
Tabela 18: Parâmetros cinéticos de XylJ e mutantes correspondentes.	130

LISTA DE ABREVIATURAS E SIGLAS

Unidades de medida

Å	angstrom
cm	centímetro
mm	milímetro
µm	micrômetro
nm	nanômetro
g	gramas
mg	miligramas
µg	microgramas
ng	nanogramas
h	hora
min	minuto
s	segundo
L	litro
mL	mililitro
µL	microlitro
M	molar
mM	milimolar
µM	micromolar
%	por cento
°C	graus célsius
x <i>g</i>	vezes a força gravitacional
kDa	quilodalton
rpm	rotação por minuto
m/v	percentual massa por volume
v/v	percentual volume por volume
V	volts

Abreviações dos resíduos de aminoácidos

A	Ala	alanina
C	Cys	cisteína
D	Asp	aspartato
E	Glu	glutamato
F	Phe	fenilalanina
G	Gly	glicina
H	His	histidina
I	Ile	isoleucina
K	Lys	lisina
L	Leu	leucina
M	Met	metionina
N	Asn	asparagina
P	Pro	prolina
Q	Gln	glutamina
R	Arg	arginina
S	Ser	serina
T	Thr	treonina
V	Val	valina
W	Trp	triptofano
Y	Tyr	tirosina

2HM	2-hidroximuconato, substrato de 4-OT
4-OD	4-oxalocrotonato descarboxilase
4-OT	4-oxalocrotonato tautomerase
CD	Dicroísmo Circular, do inglês <i>Circular Dichroism</i>
DLS	Espalhamento de Luz Dinâmico, do inglês <i>Dynamic Light Scattering</i>
DNT	domínio N-terminal
DCT	domínio C-terminal

FAH	fumarilacetoacetato hidrolase
HAP	hidrocarbonetos aromáticos policíclicos
HPD	2-hidroxi-2,4-pentadienoato, substrato de VPH
IMAC	Cromatografia de Afinidade a Metal, do inglês <i>Immobilized Metal Ion Affinity Chromatography</i>
IPTG	isopropil β-D-1-tiogalactopiranosídeo
MR	Substituição Molecular, do inglês <i>Molecular Replacement</i>
OD ₆₀₀	densidade óptica medida a 600 nm
PCR	Reação em Cadeia da Polimerase, do inglês <i>Polymerase Chain Reaction</i>
PDB ID	identificador da estrutura proteica depositada no Banco de Dados de Proteínas
q.s.p.	quantidade suficiente para
SAXS	Espalhamento de Raios-X a Baixo Ângulo, do inglês <i>Small-Angle X-Ray Scattering</i>
SDS	dodecil sulfato de sódio
SDS-PAGE	Eletroforese em Gel de Poliacrilamida – SDS, do inglês <i>Polyacrylamide Gel Electrophoresis</i>
SEC	Cromatografia de Exclusão Molecular, ou Gel Filtração, do inglês <i>Size Exclusion Chromatography</i>
TEV	Tobacco Etch Virus
T_m	temperatura de anelamento
TRIS	tris(hidroximetil)aminometano
VPH	vinilpiruvato hidratase
XRD	Difração de Raios-X, do inglês <i>X-Ray Diffraction</i>

SUMÁRIO

1. INTRODUÇÃO E JUSTIFICATIVA	1
1.1. PETRÓLEO E HIDROCARBONETOS AROMÁTICOS POLICÍCLICOS	1
1.2. BIOREMEDIAÇÃO	6
1.3. NAFTALENO E <i>PSEUDOMONAS PUTIDA</i> G7	9
1.4. DEGRADAÇÃO DE XILENOS E TOLUENOS EM <i>P. PUTIDA</i> MT-2 E DE HOMOPROTOCATECUATO EM <i>E. COLI</i> C 15	15
1.5. CARACTERIZAÇÃO ESTRUTURAL DAS ENZIMAS DA VIA DE DEGRADAÇÃO DO NAFTALENO	17
1.6. NAHK, NAHL, ENZIMAS HOMÓLOGAS E A SUPERFAMÍLIA FAH	19
2. OBJETIVOS	24
2.1. OBJETIVO GERAL	24
2.2. OBJETIVOS ESPECÍFICOS	24
3. MATERIAIS E MÉTODOS	25
3.1. ELETROFORESE EM GEL DE AGAROSE	25
3.2. CEPAS BACTERIANAS E PREPARO DE CÉLULAS ELETROCOMPETENTES	25
3.3. TRANSFORMAÇÃO DE <i>ESCHERICHIA COLI</i> POR ELETROPORAÇÃO	27
3.4. PCR DE COLÔNIAS	28
3.5. EXTRAÇÃO DE DNA PLASMIDIAL	28
3.6. DESENHO DOS OLIGONUCLEOTÍDEOS INICIADORES	29
3.7. AMPLIFICAÇÃO DOS GENES E CLONAGEM NO VETOR pGEM-T	30
3.8. SUBCLONAGEM DOS GENES NO VETOR pET28A-TEV E pETDUET-1	31
3.9. SEQUENCIAMENTO DOS CLONES POSITIVOS	33
3.10. EXPRESSÃO DAS PROTEÍNAS RECOMBINANTES	34
3.11. LISE DAS CÉLULAS E PURIFICAÇÃO DA FRAÇÃO SOLÚVEL	35
3.12. ELETROFORESE EM GEL DE POLIACRILAMIDA-SDS	37
3.13. ESPALHAMENTO DE LUZ DINÂMICO	38
3.14. DICROÍSMO CIRCULAR	38
3.15. ESPALHAMENTO DE RAIOS-X A BAIXO ÂNGULO	39
3.16. CINÉTICA ENZIMÁTICA	40
3.17. ENSAIOS DE CRISTALIZAÇÃO	41
3.18. DIFRAÇÃO DE RAIOS-X	42
3.19. CONSTRUÇÃO, REFINAMENTO E ANÁLISE DOS MODELOS CRISTALOGRAFICOS	43
4. RESULTADOS E DISCUSSÃO	45
4.1. PRODUÇÃO DE NAHK RECOMBINANTE NA AUSÊNCIA DE NAHL	45
4.2. ESTADO OLIGOMÉRICO E ESTRUTURAS SECUNDÁRIAS DE NAHK	54
4.3. NAHK POSSUI ATIVIDADE ENZIMÁTICA NA AUSÊNCIA DE NAHL	58
4.4. DETERMINAÇÃO DA ESTRUTURA CRISTALOGRAFICA DE NAHK	62
4.5. ESTRUTURA GERAL DE NAHK	68
4.6. O SÍTIO ATIVO DE NAHK	79
4.7. NAHK POSSUI UM DOMÍNIO 'TAMPA'	91
4.8. MECANISMO DE DESCARBOXILAÇÃO DE NAHK	92
4.9. PRODUÇÃO DE NAHL	99

4.10. PRODUÇÃO DO COMPLEXO NAHK/NAHL	102
4.11. ESTADO OLIGOMÉRICO E ESTRUTURAS SECUNDÁRIAS DO COMPLEXO	107
4.12. ESPALHAMENTO DE RAIOS-X A BAIXO ÂNGULO	109
4.13. ATIVIDADE ENZIMÁTICA DE NAHK NA FORMA DE COMPLEXO	116
4.14. ATIVIDADE ENZIMÁTICA DE NAHL NA FORMA DE COMPLEXO	120
4.15. CRISTALIZAÇÃO DO COMPLEXO NAHK/NAHL	122
4.16. ANÁLISE GERAL DA ESTRUTURA DE NAHL	126
4.17. ESTRUTURA GERAL DO COMPLEXO NAHK/NAHL	132
4.18. AJUSTE DOS DADOS DE SAXS UTILIZANDO O MODELO CRISTALOGRÁFICO DO COMPLEXO NAHK/NAHL	138
5. CONCLUSÃO E PERSPECTIVAS	139
6. REFERÊNCIAS BIBLIOGRÁFICAS	142
7. ANEXO I	150

1. INTRODUÇÃO E JUSTIFICATIVA

1.1. Petróleo e Hidrocarbonetos Aromáticos Policíclicos

No último século o petróleo tem sido o principal recurso natural utilizado como matéria prima e fonte de energia pelo homem. Ele serve como base para a fabricação do diesel, gasolina e querosene, que são os principais combustíveis para veículos automotores, embarcações e aeronaves em todo o mundo. Através de seu refino também são obtidos outros materiais utilizados para a fabricação dos mais variados produtos de demanda constante, como plásticos, lubrificantes, parafinas, gás natural, gás de petróleo liquefeito, solventes orgânicos, etc.

O petróleo é composto por uma combinação complexa e de proporção variável de moléculas orgânicas, dentre elas estão os hidrocarbonetos alifáticos, alicíclicos e aromáticos, podendo também conter pequenas quantidades de nitrogênio, oxigênio, enxofre e íons metálicos (principalmente níquel e vanádio) [1,2]. A proporção dentre os tipos de hidrocarbonetos pode variar consideravelmente entre diferentes reservas de petróleo e áreas geográficas. No geral, estima-se que 93 a 97% da massa total do petróleo seja constituída por hidrocarbonetos comumente presentes nas seguintes proporções: alcanos 40 a 70%, cicloalcanos 20 a 50%, aromáticos 5 a 20% [3].

Infelizmente, a intensa exploração do petróleo traz sérias consequências: uma grande variedade de poluentes é liberada no meio ambiente em decorrência de sua extração e o uso de seus derivados. Derramamentos de óleo cru durante perfurações em poços marítimos, ou acidentes durante seu transporte, são bastante conhecidos e bem reportados pela mídia devido às grandes dimensões que tais contaminações podem tomar. Por exemplo, em 2010 a torre petrolífera *Deepwater Horizon* derramou no Golfo do México mais de 794 milhões de litros de petróleo, contaminando diretamente uma área de 180.000 km² de oceano e ameaçando o equilíbrio de inúmeros ecossistemas [4]. Entretanto, tais contaminações não ocorrem somente em ambientes marinhos e costeiros. Ar, solo e lençóis freáticos são frequentemente contaminados devido aos derramamentos em menores

escalas em refinarias, terminais de distribuição, tanques de armazenamento subterrâneos e pela combustão incompleta de combustíveis, ameaçando mais diretamente comunidades humanas [5,6]. Assim, contaminações do meio ambiente por hidrocarbonetos tornaram-se frequentes, sendo reconhecido como problema mundial a partir da metade dos anos 80 [7].

Os compostos aromáticos monocíclicos (tais como benzeno, tolueno, etilbenzeno e xileno), juntamente com os hidrocarbonetos aromáticos policíclicos (HAP, como, por exemplo, naftaleno, antraceno, fenantreno e pireno) estão entre os contaminantes mais comuns e tóxicos dentre os hidrocarbonetos totais do petróleo [8]. Os HAP se destacam dentre os demais hidrocarbonetos poluentes por possuírem dois ou mais anéis aromáticos fundidos, sendo classificados como leves (de 2 a 3 anéis) ou pesados (de 4 a 6 anéis) [9,10].

Devido à possibilidade da fusão de um número variável de anéis e das várias posições em que estes podem ligar entre si, há atualmente mais de 100 HAP reconhecidos pela União Internacional de Química Pura e Aplicada – IUPAC. Dentre esses, 32 foram listados pela Agência de Proteção Ambiental dos Estados Unidos – EPA como os principais HAP poluentes, sendo que 16 deles (Figura 1) são tomados como alvos frequentes para o monitoramento e dosagem de amostras em função de suas importâncias físico-químicas, toxicológicas, industriais e ambientais [11]. Além disso, os HAP são moléculas bastante estáveis quimicamente, e, somando-se a essa característica a sua hidrofobicidade, tendem a ser fortemente adsorvidos às partículas orgânicas e minerais do solo, o que contribui para sua persistência no ambiente e aumento da possibilidade de exposição desses compostos aos seres humanos, outros animais e vegetais [11].

O tempo de permanência dessas moléculas no ambiente depende de vários fatores, como a estrutura química do HAP, concentração e a biodisponibilidade do contaminante. Adicionalmente, fatores como tipo e estrutura do solo, pH, temperatura, bem como níveis de oxigênio, nutrientes e água podem limitar a capacidade da microbiota local em degradar essas moléculas [12,13].

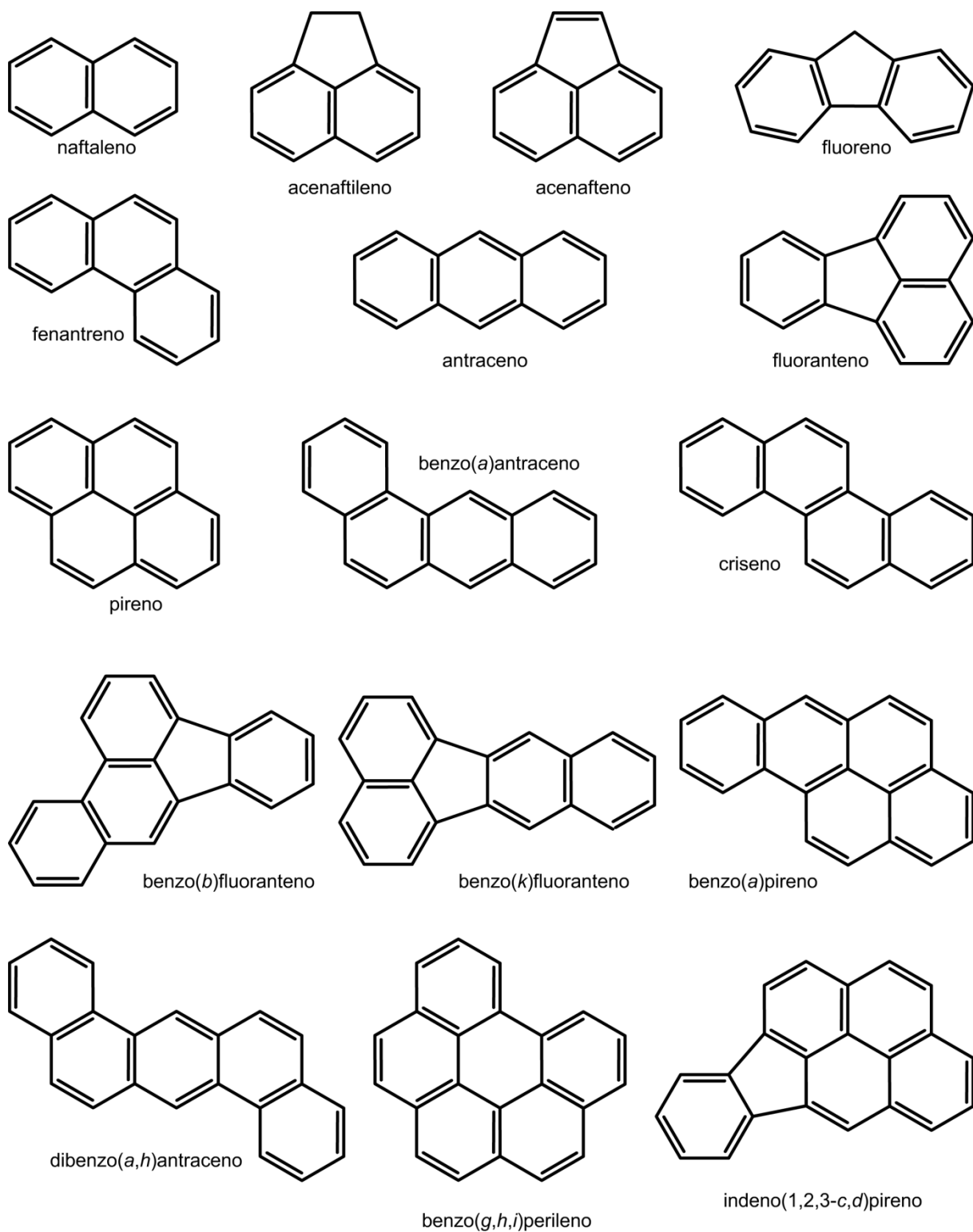


Figura 1: Estrutura química dos 16 HAP listados pela EPA como prioritários para monitoramento no meio ambiente.

Quando incorporados pela fauna e flora, os HAP ligam-se às moléculas proteicas impedindo seu correto funcionamento e também, por serem hidrofóbicos, atravessam facilmente a membrana celular. Adicionalmente, tornam-se parte da reserva lipídica e podem permanecer indefinidamente no organismo, podendo ser transferidos sem alteração de suas estruturas da presa ao predador e, eventualmente, ao homem. De fato, muitos seres expõem-se à contaminação por hidrocarbonetos derivados do petróleo principalmente pela ingestão de animais ou plantas que acumularam tais moléculas. Isto traz sérias implicações para a pesca e a saúde pública devido à subsistência a partir de áreas contaminadas [14]. HAP e seus derivados também podem ser absorvidos pela pele ou por inalação, sendo rapidamente distribuídos pelo organismo [9].

Uma vez dentro das células, o metabolismo dos HAP torna-os mais hidrossolúveis, permitindo serem mais facilmente excretados pela urina. Apesar disso, os metabólitos intermediários dessas moléculas são ainda mais tóxicos e carcinogênicos. Exposições a esses contaminantes foram reportadas como diretamente relacionadas ao desenvolvimento de câncer e outras doenças [15,16]. O metabolismo dos HAP no organismo envolve uma série de enzimas presentes em todos os tecidos e que catalisam reações de oxidação, redução, hidrólise e conjugação. Devido às diferenças estruturais dos diversos HAP, várias vias podem ser utilizadas em sua biomodificação, ocorrendo diferenças no metabolismo e na carcinogenicidade de cada molécula.

Alguns mecanismos têm sido propostos para explicar a transformação dos HAP em moléculas mais carcinogênicas. No mecanismo mais aceito, por formação de diol-epóxido (Figura 2 a), monoxigenases dependentes do citocromo P450 oxidam os HAP formando compostos epóxidos que podem espontaneamente produzir fenóis ou, por ação das epóxido hidrolases, produzir dihidrodíóis. Esses últimos podem sofrer nova epoxidação levando à formação de dihidrodíóis epóxidos, que são capazes de reagir com as bases nucleotídicas do DNA e, eventualmente, iniciar um processo mutagênico. No mecanismo via formação de quinona (Figura 2 b), a desidrogenação enzimática dos dihidrodíóis produzem hidroquinonas e quinonas capazes de reagir diretamente com o DNA gerando adutos, ou com O₂ gerando espécies reativas de oxigênio que atacam o DNA e ocasionam lesões que podem levar a mutações. Há ainda o mecanismo de formação de radicais catiônicos através de oxidação enzimática que contribui para a depuração de bases do DNA (Figura 2 c); e também o

mecanismo por oxidação benzílica, no qual há a formação de ésteres benzílicos, eletrofílicos, através de uma série de reações de substituição, e que também promovem a formação de adutos com o DNA (Figura 2 d). Estes mecanismos não são excludentes, podendo ocorrer simultaneamente [9].

Todos esses fatos despertam a atenção da indústria e da comunidade científica no sentido de desenvolver pesquisas que visam diminuir a concentração de HAP de áreas contaminadas a fim de evitar o contato de humanos e outros animais com esses compostos.

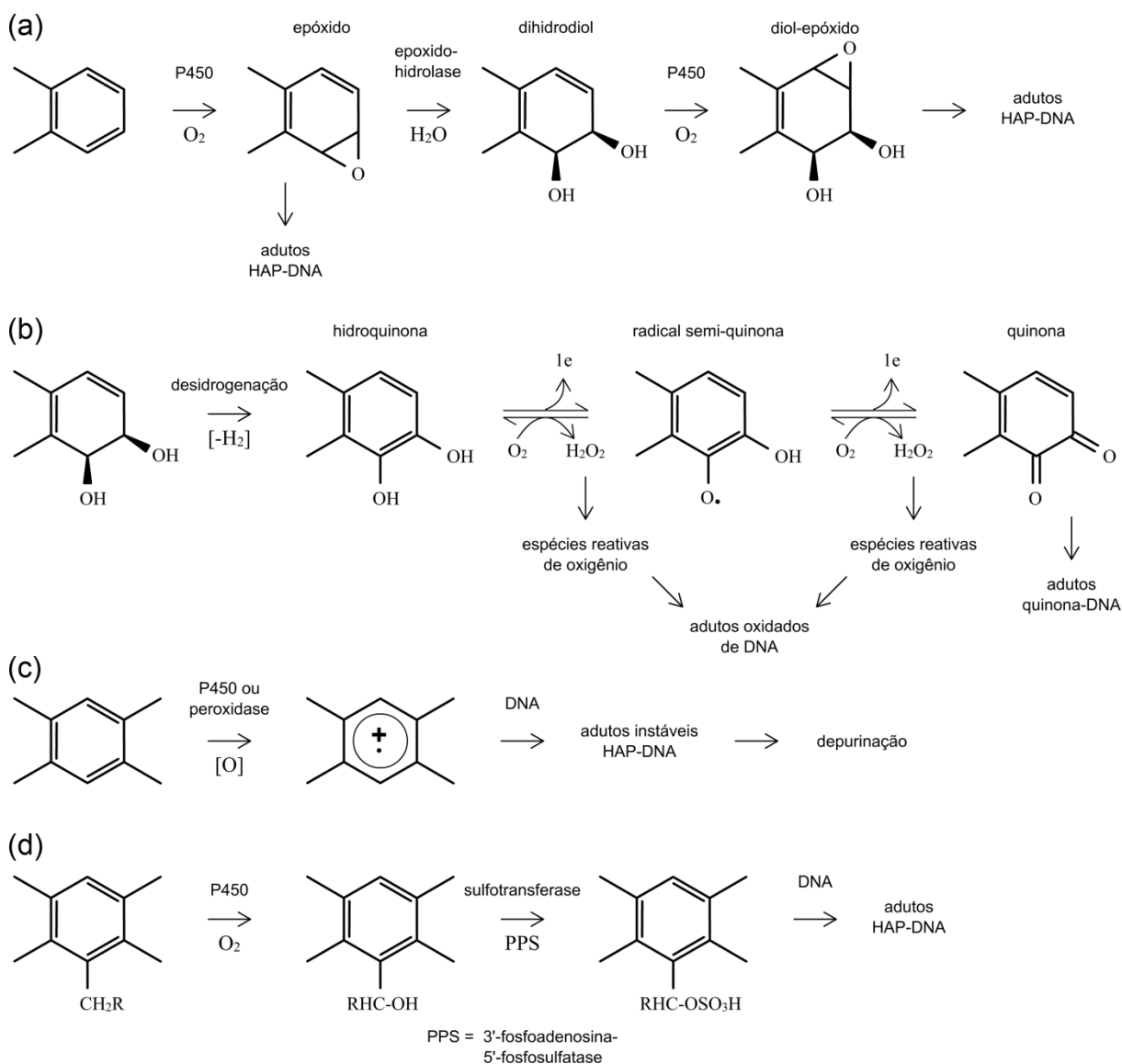


Figura 2: Mecanismos de ativação dos HAP, formação de adutos e danos ao DNA.

(a) Mecanismo por formação de diol-epóxido. **(b)** Mecanismo via formação de quinona. **(c)** Mecanismo por formação de radical-cátion. **(d)** Mecanismo por oxidação benzílica.

1.2. Bioremediação

Existem diversos processos que auxiliam na diminuição da concentração dos poluentes do ambiente. Processos de remediação físicos e químicos consistem na coleta do material contaminado, seja água ou solo, e remoção dos poluentes por adsorção em carvão ativado, oxidação por meio de aquecimento em temperaturas de até 500 °C, ou outros processos demorados e onerosos [14]. Devido à relação de custo-benefício e eficiência, outras estratégias para a remediação de áreas contaminadas têm sido buscadas.

Estima-se que 90% das emissões de HAP ocorram por atividade antropogênica através da queima de madeira para produção de carvão vegetal, uso de incineradores e outros processos de combustão incompleta, e, principalmente, devido à extração, transporte, refino, armazenamento e utilização do petróleo e seus derivados [9,17]. Apesar disso, os HAP também são naturalmente gerados e liberados no meio ambiente através da combustão incompleta de matéria orgânica, como em queimadas, incêndios florestais e atividade vulcânica [18]. Dessa maneira, tratando-se de compostos orgânicos ubíquos, diversos microrganismos desenvolveram a habilidade de utilizar essas moléculas como fonte de carbono e energia [19].

Desde a década de 1950 espécies de diversos gêneros de fungos (*Cunninghamella elegans*, *Aspergillus* sp., *Trichocladium* sp., *Fusarium oxysporum*, *Verticillium* sp., *Achremonium* sp.) e bactérias (*Pseudomonas* sp., *Agrobacterium* sp., *Bacillus* sp., *Burkholderia* sp., *Rhodococcus* sp., *Mycobacterium* sp., *Sphingomonas* sp.) vêm sendo isoladas e caracterizadas quanto à habilidade de consumir diferentes HAP, e novas vias de degradação têm sido elucidadas [20-23]. Assim, o uso desses microrganismos tornou-se uma estratégia interessante para o tratamento de áreas contaminadas por tais poluentes.

O termo bioremediação engloba diversas técnicas que fazem uso de processos de degradação biológicos *in* ou *ex-situ* para converter os contaminantes a moléculas menos tóxicas ou totalmente inertes (como por exemplo, CO₂ e H₂O). Tais técnicas são frequentemente consideradas tecnologias mais limpas e baratas quando comparadas a tratamentos físicos ou químicos de áreas contaminadas [23,24]. De fato, essa biotecnologia já vem sendo utilizada há vários anos em diversos países em escala comercial no tratamento

de diversos resíduos e na remediação de áreas contaminadas e, em certos casos, apresenta menor custo e maior eficiência na remoção dos contaminantes do que técnicas como incineração e lavagem do solo [13,25].

A degradação de HAP por microrganismos é limitada por uma série de fatores como a quantidade de água no solo, pH, temperatura, e níveis adequados de nutrientes como nitrogênio, fósforo, entre outros [26]. Tendo como premissa tais limitações, técnicas de bioremediação *in-situ* que adicionam nutrientes à área contaminada pode permitir o aumento do crescimento de microrganismos já presentes no local e, portanto, o aumento da velocidade de biodegradação dos poluentes. Essa técnica, chamada de bioestimulação, foi uma das estratégias utilizadas para tratar as linhas costeiras afetadas pelo derramamento de pelo menos 41 milhões de litros de óleo cru que vazaram do navio petroleiro *Exxon Valdez* na costa do Alasca em 1989 [27]. Apesar dos esforços no tratamento dessa região, estudos realizados quase 20 anos após o vazamento mostram que o solo arenoso das praias e costões contaminados ainda continha óleo proveniente do desastre, e que a contaminação em longo prazo, principalmente por HAP que persistem no local, causou uma taxa de mortalidade duradoura em diversos tipos de animais que habitam, alimentam ou se reproduzem na região [28].

Como observado, a comunidade microbiana presente naturalmente no local contaminado pode não exibir um potencial metabólico apropriado para a degradação e completa mineralização de alguns poluentes. Isso pode ocorrer, dentre diversos fatores, devido ao baixo número de microrganismos que possuem enzimas capazes de transformar o poluente em moléculas mais simples e menos tóxicas, ou pelo fato de que o poluente em si (ou a mistura de compostos) é complexo e só pode ser degradado por uma combinação específica de microrganismos. Nesse caso, a solução para uma bioremediação bem sucedida pode ser a inoculação de uma população de microrganismos exógenos que permita a degradação específica de determinados poluentes [26]. Essa técnica, chamada de bioaumento, tem sido muito debatida devido a estudos que comprovam tanto sua eficácia, quanto sua falha [29]. Além disso, a introdução de microrganismos exógenos em uma área delimitada pode sair do controle e implicar em uma competição com a microbiota nativa, gerando um desequilíbrio e alterando o ecossistema [30].

Ademais, condições adversas do ambiente podem não permitir a sobrevivência do microrganismo inoculado ou a degradação eficiente do poluente, ocasionando uma eficiência mais baixa quando comparada com resultados obtidos em condições laboratoriais [22]. Tais limitações abrem caminho para o estudo e desenvolvimento de organismos geneticamente modificados capazes de realizar a degradação de compostos tóxicos e que resistam a condições adversas. Nesse contexto, a combinação de vias catabólicas de diferentes microrganismos em uma única cepa é uma estratégia promissora para aumentar a eficiência de processos de bioaumento [30]. Entretanto, ainda ocorreriam os riscos associados com a transferência horizontal de genes e uma proliferação incontrolada do microrganismo inoculado, causando problemas com a microbiota nativa.

Frente a essas limitações, é levantado o fato de que a degradação de poluentes dentro dos microrganismos ocorre através de vias contendo séries de reações catalisadas por enzimas. Com as tecnologias atuais da genômica funcional, genomas completos de diversos microrganismos têm sido sequenciados e milhares de operons que codificam enzimas para o catabolismo de vários compostos vêm sendo elucidados [30]. Dessa maneira, uma opção interessante para apoiar e melhorar as técnicas atuais de bioremediação é o estudo, aprimoramento e a utilização das próprias enzimas catabólicas nos locais contaminados ou em técnicas *ex-situ*. Essa técnica, chamada de bioremediação enzimática, fornece uma série de vantagens frente aos demais processos de bioremediação [31].

O uso de enzimas provê um sistema mais simples e controlado do que o uso de um microrganismo inteiro. Como as enzimas catalisam reações específicas, a biotransformação dos poluentes não gera subprodutos indesejáveis, e, após algum tempo, as enzimas são removidas do ambiente pela degradação realizada pelos próprios microrganismos nativos. Além disso, através da tecnologia do DNA recombinante e métodos de engenharia de proteínas, é possível produzir enzimas em grande escala com atividade e estabilidade aumentadas [32]. Entretanto, alguns fatores limitantes devem ser considerados, como a necessidade do uso de cofatores (NAD(P)H, glutatona, flavina, entre outros) que elevaria os custos para o uso de enzimas em nível comercial, e que as condições ótimas para a ação das enzimas devem ser mantidas durante o processo [33].

Algumas dessas implicações podem ser contornadas através de um amplo conhecimento da via de degradação analisada e de como essas determinadas enzimas funcionam, e isso, na maioria das vezes, é conseguido pelo estudo cinético e pela resolução da estrutura tridimensional das mesmas. A partir do entendimento da estrutura tridimensional e função de enzimas é possível identificar os resíduos importantes para ligação ao substrato e catálise. Com o uso de técnicas de biologia molecular e de engenharia de proteínas, como desenho racional e mutações sítio-dirigidas, é possível aumentar a estabilidade e atividade enzimática para que essa possa agir em maiores faixas de temperatura, pH e força iônica [19,33]. Esses estudos abrem portas para o aprimoramento e o uso de enzimas diversas, o que permite tornar os processos de bioremediação mais baratos e efetivos.

1.3. Naftaleno e *Pseudomonas putida* G7

O estudo de enzimas para bioremediação envolve etapas cruciais como: selecionar e isolar microrganismos capazes de degradar compostos de interesse, sequenciar seu genoma e identificar as proteínas responsáveis pelas vias de degradação, isolar os genes codificadores dessas enzimas, expressar e purificar as mesmas para estudos cinéticos e estruturais, testar e aprimorar as propriedades enzimáticas em laboratório [33].

O naftaleno é o HAP mais simples, formado por apenas dois anéis aromáticos fundidos. Ele possui baixa solubilidade em água (30 mg/L), sublima facilmente à temperatura ambiente e tende a ser fortemente absorvido pelo solo. Apesar de sua hidrofobicidade, dentre os HAP, o naftaleno é o que possui a maior solubilidade em água, sendo um dos primeiros dentre esses compostos a atingir lençóis freáticos em caso de contaminação do solo, e o prioritário da listagem dos 16 HAP a terem seus níveis de concentração monitorados no ambiente [34]. Além disso, devido à sua estrutura mais simples, o naftaleno é frequentemente utilizado como molécula modelo em estudos que visam caracterizar a habilidade de microrganismos em degradar HAP mais complexos, auxiliando também na compreensão de vias catabólicas de HAP com três ou mais anéis benzênicos [35].

Apesar de sua ubiquidade, atividades antropogênicas continuam sendo a principal fonte de contaminação local por naftaleno. Como exemplos podemos citar as indústrias químicas e siderúrgicas, queima de biomassa, fumaça de cigarro, repelente de insetos, desodorizantes, inseticidas e diversas outras fontes em adição as atividades relacionadas à extração e uso de petróleo e derivados [36]. Exposição ao naftaleno foi associada a diversos casos de toxicidade e outros efeitos agudos e crônicos em humanos e animais em laboratório, sendo reportada anemia hemolítica, atividade cataratogênica, dano aos pulmões, rim, cérebro, baço e fígado, incluindo tumores em ratos e camundongos [37]; além de formação de estresse oxidativo nas células e depuração de bases do DNA, aumentando as chances de mutações e desenvolvimento de câncer [38]. Assim, o naftaleno foi reclassificado pela EPA como um possível carcinógeno humano, sendo essencial o seu controle e eliminação de áreas contaminadas [39].

Como apresentado, fungos e bactérias estão ativamente envolvidos na degradação de poluentes orgânicos de áreas contaminadas. Diversas espécies bacterianas, principalmente do gênero *Pseudomonas*, são conhecidas e estudadas pela capacidade de degradação de HAP, sendo que a maioria das que possuem uma eficiente atividade de degradação foram isoladas justamente de solos contaminados [20]. Nesse sentido, a degradação de naftaleno por *Pseudomonas putida* G7 tem sido bastante estudada devido à presença de um plasmídeo de 83 kb chamado NAH7 [40]. Esse plasmídeo codifica mais de 20 proteínas que juntas são capazes de degradar completamente o naftaleno, permitindo ao hospedeiro utilizar essa molécula (ou o salicilato) como única fonte de carbono [41]. O plasmídeo NAH7 também possui genes para outras 60 proteínas com funções diversas, como, por exemplo, genes para replicação e manutenção do plasmídeo, conjugação bacteriana, resistência à luz ultravioleta, recombinação e outras funções [40].

Os genes para a degradação de naftaleno, chamados de *nah*, estão organizados em dois operons: operon *nah* superior, que codifica as enzimas da via superior de degradação (NahAaAbAcAdBCDEF), responsáveis pela abertura e degradação do primeiro anel aromático, convertendo o naftaleno a salicilato e piruvato; e o operon *nah* inferior (ou operon *sal*), que codifica as enzimas da via inferior de degradação (NahGHIJKLMNO), que abrem e degradam o segundo anel aromático por meio da via de *meta*-clivagem do catecol, convertendo o salicilato a acetil-CoA e mais uma molécula de piruvato [40] (Figura 3). Ambos

os operons são regulados positivamente por um terceiro operon, que codifica apenas uma proteína (NahR) responsável pela regulação positiva da transcrição dos outros dois operons [42]. NahR é regulada positivamente por salicilato, o último produto da via superior de degradação do naftaleno. Existem outros genes nos operons superior e inferior: *nahX* e *nahQ*, que codificam proteínas de função desconhecida, *nahT* que codifica uma ferredoxina hipotética, e *nahY*, que codifica uma proteína responsável por prover quimiotaxia ao naftaleno [43].

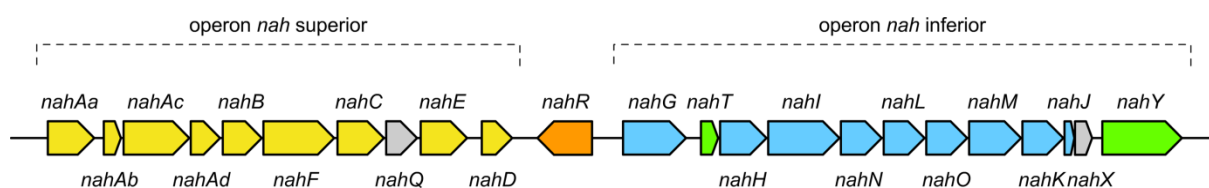


Figura 3: Representação dos genes *nah* do plasmídeo NAH7.

Os polígonos indicam o tamanho e sentido de transcrição dos genes *nah*: em amarelo, genes que pertencem à via superior de degradação; em azul, genes que pertencem à via inferior de degradação; em laranja está o gene para a proteína regulatória da expressão dos demais genes *nah*; em verde e cinza estão outros genes para proteínas com outras funções.

O primeiro passo nessa via de degradação, e talvez o mais complexo, é a incorporação de um oxigênio molecular num dos anéis aromáticos do naftaleno para a produção de *cis*-1,2-dihidronaftaleno-1,2-diol (**2**, Figura 4). Essa reação é catalisada pelo complexo proteico naftaleno dioxigenase e utiliza NADH como cofator. Esse complexo é constituído por três componentes proteicos: ferredoxina redutase (**NahAa**), ferredoxina (**NahAb**) e uma proteína ferro-enxofre composta de subunidades α e β (**NahAc** e **NahAd**). NahAa transfere dois elétrons a partir de NADH ao FAD, gerando uma forma totalmente reduzida do FAD, que, por sua vez, transfere os dois elétrons para o centro ferro-enxofre de NahAb. Esses elétrons são, por fim, transferidos para o sítio ativo de NahAcAd para facilitar a adição do oxigênio molecular ao naftaleno [42].

A próxima etapa é uma dupla desidrogenação realizada pela enzima *cis*-naftaleno dihidrodiool desidrogenase (**NahB**), utilizando NAD^+ como cofator e gerando 1,2-dihidroinaftaleno (**3**, Figura 4). Esse produto é *meta*-clivado pela adição de mais um oxigênio molecular através da enzima 1,2-dihidroinaftaleno dioxigenase (**NahC**), gerando 2-hidroxi-2*H*-cromeno-2-carboxilato (**4**). Essa reação abre o primeiro anel aromático, mas o produto da reação se recicla espontaneamente, formando novamente um anel. A quarta

reação da via é realizada pela enzima 2-hidroxicromeno-2-carboxilato isomerase (**NahD**), que abre definitivamente o anel, produzindo *trans*-*o*-hidroxibenzilidenopiruvato (**5**). Na sequência, essa molécula é clivada pela enzima *trans*-*o*-hidroxibenzilidenopiruvato hidratase-aldolase (**NahE**) pela adição de uma molécula de água ao substrato, gerando uma molécula de salicilaldeído (**6**) e uma de piruvato [42]. Finalizando as etapas da via superior de degradação, a enzima salicilaldeído desidrogenase (**NahF**), utilizando NAD^+ como cofator, converte salicilaldeído a salicilato (**7**) [42].

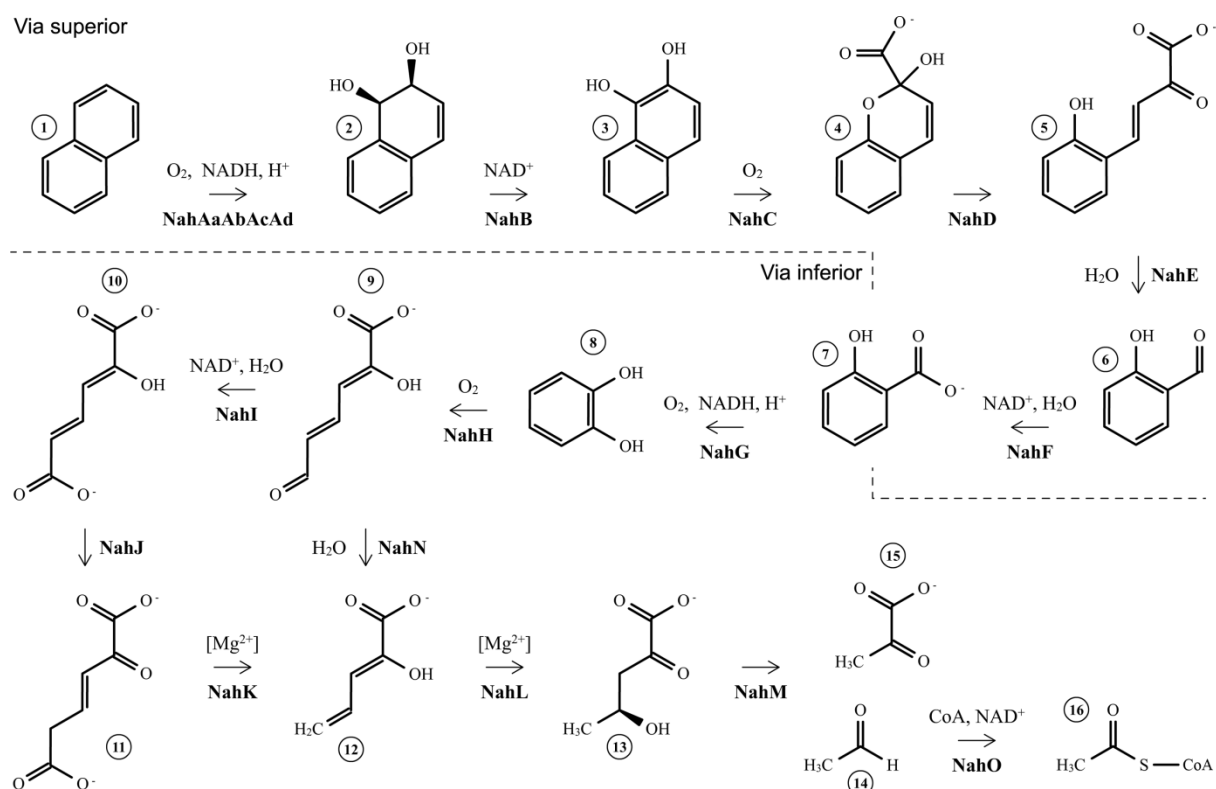


Figura 4: Via de degradação do naftaleno em *P. putida* G7.

A via superior de degradação converte naftaleno (**1**) a salicilato (**7**) e piruvato (**15**) através das enzimas NahAaAbAcAdBCDEF. Os intermediários dessa conversão são: (**2**) *cis*-1,2-dihidronaftaleno-1,2-diol, (**3**) 1,2-dihidroinaftaleno, (**4**) 2-hidroxi-2*H*-cromeno-2-carboxilato, (**5**) *trans*-*o*-hidroxibenzilidenopiruvato, (**6**) salicilaldeído. A via de degradação inferior converte salicilato (**7**) a piruvato (**15**) e acetil-CoA (**16**) através das enzimas NahGHIJKLMNO. Os intermediários dessa conversão são: (**8**) catecol, (**9**) 2-hidroxiuconato semialdeído, (**10**) 2-hidroxiuconato, (**11**) 4-oxalocrotonato, (**12**) 2-hidroxi-2,4-pentadienoato, (**13**) 2-oxo-4-hidroxipentanoato, (**14**) acetaldeído.

A via inferior de degradação continua as etapas de clivagem do naftaleno através da enzima salicilato hidroxilase (**NahG**), que utiliza NADH e oxigênio molecular para converter salicilato (**7**, Figura 4) a catecol (**8**). O catecol, através da *meta*-clivagem, é oxidado pela enzima catecol 2,3-dioxigenase (**NahH**), abrindo o segundo anel aromático, e formando 2-

hidroximuconato semialdeído (**9**). Esse produto, por sua vez, pode seguir dois rumos para continuar a catálise. Através do ramo multi-enzimático ocorre mais uma etapa de oxidação, no qual a enzima 2-hidroximuconato semialdeído desidrogenase (**NahI**), utilizando NAD^+ como cofator, oxida seu substrato a 2-hidroximuconato (**10**). Esse produto sofre uma etapa de isomerização, catalisada pela enzima 4-oxalocrotonato tautomerase (**NahJ**), gerando 4-oxalocrotonato (**11**). Na sequência, utilizando íon magnésio como cofator, ocorre uma etapa de descarboxilação seguida de hidratação, catalisadas pelo complexo proteico formado pelas enzimas 4-oxalocrotonato descarboxilase (**NahK**) e vinilpiruvato hidratase (**NahL**), gerando os produtos intermediários 2-hidroxi-2,4-pentadienoato (**12**) e 2-oxo-4-hidroxipentanoato (**13**). Através de um atalho nessa via, chamado de ramo hidrolítico, a enzima 2-hidroximuconato semialdeído hidrolase (**NahN**) é capaz de clivar o 2-hidroximuconato semialdeído (**9**) gerando também o substrato de NahL (**12**) [44]. O penúltimo passo da via é catalisado pela enzima 2-oxo-4-hidroxipentanoato aldolase (**NahM**), que cliva seu substrato em uma molécula de acetaldeído (**14**) e uma de piruvato (**15**). Por fim, a enzima acetaldeído desidrogenase (**NahO**), utilizando NAD^+ como cofator, produz acetil-CoA (**16**), o produto final dessa via de degradação [42]. Assim, vemos que a degradação do naftaleno, através das enzimas do plasmídeo NAH7, gera como principais produtos duas moléculas de piruvato e uma molécula de acetil-CoA, ambos compostos são direcionados para o ciclo de Krebs para produção de energia.

O catecol também pode ser degradado pela via de *orto*-clivagem (Figura 5), que ocorre pelas enzimas codificadas pelo cromossomo bacteriano, resultando nos produtos finais succinil-CoA e acetil-CoA. Também é interessante destacar que as enzimas da via superior, que degradam o naftaleno até salicilato, também são capazes de degradar fenantreno (**4**, Figura 6) e antraceno (**7**) a 1-hidroxi-2-naftoato (**5**) e 2-hidroxi-3-naftoato (**8**), respectivamente. O composto **5** (Figura 6) pode ser convertido pela enzima 1-hidroxi-2-naftoato hidroxilase a 1,2-dihidroxi-naftaleno (**6**), que entra na via de degradação do naftaleno já apresentada previamente. O composto **8** (Figura 6) é convertido a 2,3-dihidroxi-naftaleno (**9**), que por sua vez é convertido por outras enzimas a salicilato e catecol, entrando na via inferior de degradação do naftaleno [42]. Não está claro se o plasmídeo NAH7, com mais de 25 proteínas hipotéticas [40], ou se a bactéria *P. putida* G7 possui genes

para essas outras enzimas em seu cromossomo, as quais permitiriam também a completa degradação de fenantreno e antraceno por esse microrganismo.

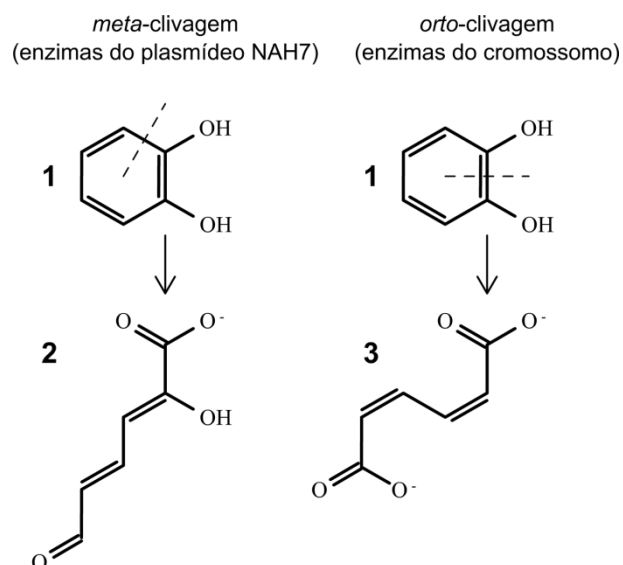


Figura 5: *Meta* e *orto*-clivagem do catecol.

A *meta*-clivagem converte catecol (1) a 2-hidroxi-3-oxo-6-oxohept-2-enoato (2); enquanto a *orto*-clivagem converte catecol a cis,cis-muconato (3). Os produtos finais de ambas as vias entram no ciclo de Krebs para produção de energia.

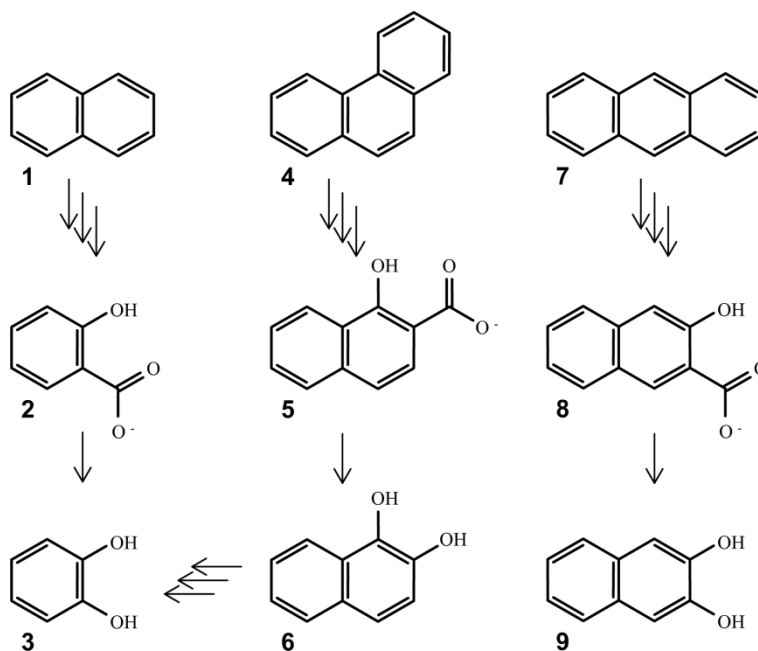


Figura 6: Comparação dos diferentes produtos gerados pela degradação de naftaleno, fenantreno e antraceno pelas enzimas Nah codificadas pelo plasmídeo NAH7.

(1) naftaleno, (2) salicilato, (3) catecol, (4) fenantreno, (5) 1-hidroxi-2-naftoato, (6) 1,2-dihidroxi-naftaleno, (7) antraceno, (8) 2-hidroxi-3-naftoato, (9) 2,3-dihidroxi-naftaleno.

1.4. Degradação de xilenos e toluenos em *P. putida* mt-2 e de homoprotocatecuato em *E. coli* C

Como apresentado, alguns microrganismos possuem um impressionante potencial metabólico para a degradação de vários compostos poluentes. Esses diferentes compostos são inicialmente degradados por diferentes vias metabólicas, ditas vias periféricas, que convergem a um número limitado de moléculas chave, que são, por fim, processadas através de vias de degradação centrais. Um exemplo é a degradação de diversos compostos aromáticos, como o triptofano, naftaleno, tolueno e fenantreno, que são todos degradados até catecol (ou catecóis com diferentes grupos substituintes), que, por sua vez, é degradado através de vias comuns [45].

Dentro desse contexto, pode-se destacar o plasmídeo TOL pWW0 de *P. putida* mt-2. Esse plasmídeo possui 115 kb e codifica diversas enzimas para a degradação de toluenos e xilenos [46]. Os genes de degradação do plasmídeo TOL pWW0, denominados genes *xyl*, também estão organizados em operons. O operon superior, composto pelos genes *xy/CMABN*, codifica enzimas que degradam tolueno, 3-etil tolueno, *m*- e *p*-xileno, e 1,2,4-trimetil benzeno a benzoato, 3-etil benzoato, *m*- e *p*-toluato, e 3,4-dimetilbenzoato, respectivamente. O operon inferior pode ser fisicamente e funcionalmente dividido entre os genes *xy/XYZL*, envolvidos na conversão desses benzoatos/toluatos a (metil)catecóis; e os genes *xy/ITEGFJQKIH*, envolvidos na conversão dos (metil)catecóis a intermediários do ciclo de Krebs através da via de *meta*-clivagem do catecol [47]. A expressão desses genes é controlada pelo produto de dois genes regulatórios, *xy/R* e *xy/S*.

A degradação de tirosina e de fenilalanina em diversas bactérias, como por exemplo, em *E. coli* C, pode ocorrer pela via de degradação do ácido 4-hidroxifenilacético através de enzimas codificadas pelo cromossomo bacteriano. O ácido 4-hidroxifenilacético é convertido a homoprotocatecuato que, por sua vez, é *meta*-clivado numa via bastante semelhante à via de degradação por *meta*-clivagem do catecol [48-50].

Ao se fazer a comparação das reações enzimáticas da via inferior de degradação do naftaleno do plasmídeo NAH7, com o segundo grupo de genes da via inferior de degradação de xilenos e toluenos do plasmídeo TOL pWW0, juntamente com a via de clivagem do

homoprotocatecuato em *E. coli* C, percebe-se que a série de reações que catalisa a degradação dos respectivos catecóis com seus diferentes grupos substituintes até intermediários do ciclo de Krebs é praticamente idêntica (Figura 7) [45,49].

Assim, pode-se destacar a existência de diversos genes que codificam enzimas isofuncionais, como NahJ / XylH / HpcD, NahK / XylI / HpcE, e NahL / XylJ / HpcG. Tal fato tem sido atribuído à transferência horizontal de uma via ancestral e sua integração em vários operons responsáveis pela degradação de diferentes compostos aromáticos [47]. Sendo assim, a elucidação da estrutura-função de uma das enzimas isofuncionais também auxilia na compreensão das demais enzimas homólogas em suas respectivas vias de degradação.

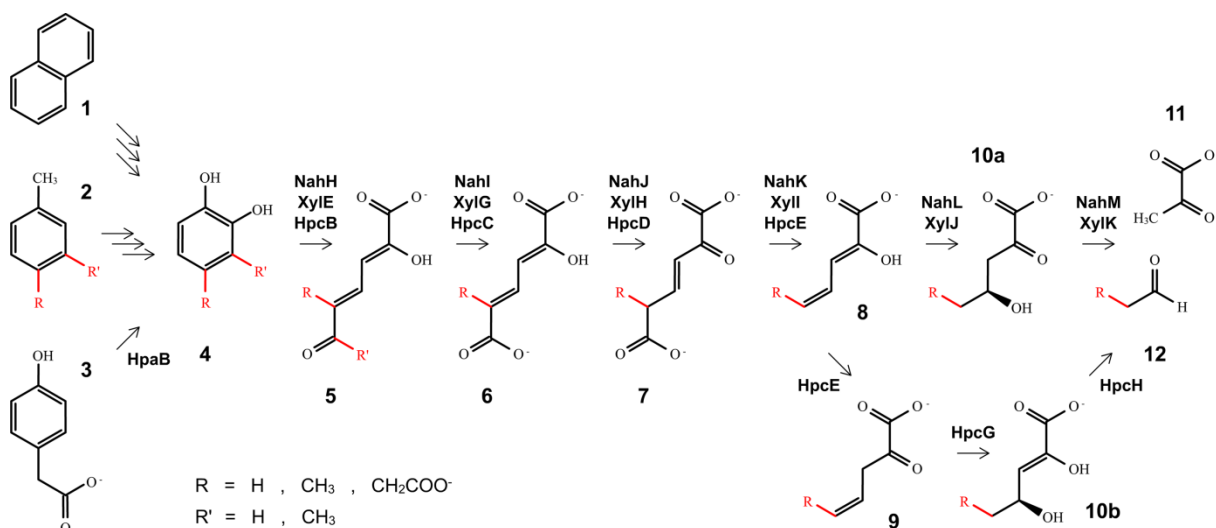


Figura 7: Via de *meta*-clivagem de catecóis com diferentes grupos substituintes a partir de diferentes compostos aromáticos.

O naftaleno através de diversas reações é convertido a catecol; xilenos e toluenos também são convertidos por várias enzimas a (metil)catecóis; aminoácidos aromáticos, como fenilalanina e tirosina, são degradados a 4-hidroxifenilacetato, que por sua vez, é convertido a homoprotocatecuato. Todos esses catecóis são degradados de forma bastante semelhante através da via de *meta*-clivagem do catecol.

(1) naftaleno, (3) 4-hidroxifenilacetato, (11) piruvato.

Para $R = H$ e $R' = H$: (2) tolueno, (4) catecol, (5) 2-hidroxi-6-oxohepta-2,4-dienoato, (6) 2-hidroxi-5-metil-6-oxohepta-2,4-dienoato, (7) 4-oxalocrotonato, (8) 2-hidroxi-2,4-pentadienoato, (10a) 2-oxo-4-hidroxipentanoato, (12) acetaldeído.

Para $R = H$ e $R' = CH_3$: (2) *m*-xileno, (4) 3-metilcatecol, (5) 2-hidroxi-6-oxohepta-2,4-dienoato.

Para $R = CH_3$ e $R' = H$: (2) *p*-xileno, (4) 4-metilcatecol, (5) 2-hidroxi-5-metil-6-oxohepta-2,4-dienoato, (6) 2-hidroxi-5-metil-6-oxohepta-2,4-dienoato, (7) 5-metil-2-oxohex-4-eno-1,6-dienoato, (8) 2-hidroxi-2,4-pentadienoato, (10a) 4-hidroxi-2-oxohexanoato, (12) propionaldeído.

Para $R = CH_2COO^-$ e $R' = H$: (4) homoprotocatecuato, (5) 5-carboximetil-2-hidroxi-6-oxohepta-2,4-dienoato, (6) 5-carboximetil-2-hidroxi-6-oxohepta-2,4-dienoato, (7) 5-oxopent-3-eno-1,2,5-tricarboxilato, (8) 2-hidroxi-2,4-pentadienoato, (9) 2-oxohept-3-eno-1,7-dienoato, (10b) 2,4-dihidroxihept-2-ene-1,7-dienoato, (12) succinato semialdeído.

1.5. Caracterização estrutural das enzimas da via de degradação do naftaleno

Como sugerido, um amplo conhecimento de uma determinada via de degradação e de como suas enzimas funcionam pode ser alcançado pela resolução da estrutura tridimensional dessas enzimas.

O método experimental mais comum para obter a estrutura tridimensional detalhada de uma macromolécula é através da interpretação do padrão de difração de raios-X de uma amostra cristalina [51]. Portanto, determinar a estrutura tridimensional de uma proteína por XRD implica, primeiramente, em obter cristais de alta qualidade da proteína alvo. Os cristais são obtidos pela precipitação lenta da proteína em forma pura a partir de uma solução aquosa em condições que não desnatura a mesma. Para coletar os dados de difração, o cristal é posicionado entre uma fonte e um detector de raios-X. Ao atravessar o cristal, os raios-X são difratados e emergem como se fossem ondas refletidas por centenas de planos cristalinos imaginários do cristal, aparecendo como pontos espalhados na imagem capturada pelo detector. Cada “ponto de difração” possui uma amplitude e uma fase, porém, durante o processo, somente as informações da amplitude e direção de cada onda são coletadas. A informação da fase, essencial para resolver a estrutura da proteína, é perdida. Um dos métodos utilizados para contornar o problema das fases é o de Substituição Molecular, que consiste em utilizar as fases do modelo tridimensional de uma proteína conhecida como estimativa inicial para as fases da molécula em estudo. A proteína conhecida é referida como modelo de fases e deve apresentar certa identidade com a proteína estudada [51]. Posteriormente, após o tratamento inicial dos dados e a obtenção das fases, deve-se estimar a distribuição das nuvens eletrônicas das moléculas que constituem o cristal e, a partir da interpretação deste dado, construir um modelo atômico que justifique a distribuição de densidade eletrônica (e conseqüentemente os dados de difração medidos). Ao final do processo, obtém-se um modelo tridimensional com as posições atômicas das moléculas em estudo. Obs.: mais detalhes no livro “Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models” [51] – além do propósito desta tese.

O Banco de Dados de Proteínas RCSB é um portal de informação de estruturas macromoleculares biológicas no qual são depositados modelos de estruturas tridimensionais de proteínas e ácidos nucleicos [52]. De um total de 113.816 estruturas tridimensionais atualmente depositadas no PDB (dados de 2015), 89% foram obtidas por XRD, e destas, aproximadamente 72% foram de proteínas obtidas de forma recombinante através de um sistema de expressão heterólogo em *E. coli*. No portal do PDB, utilizando a ferramenta *Search/Sequences* juntamente com a sequência de aminoácidos da proteína alvo, é possível localizar outras proteínas com estrutura tridimensional já resolvida e que possuem certa identidade com a proteína em estudo.

Dessa maneira, uma análise das sequências de aminoácidos das enzimas da via superior e inferior de degradação do naftaleno em *P. putida* G7 indicou que as proteínas NahAb, NahAc e NahAd do complexo enzimático da naftaleno dioxigenase, bem como as enzimas NahC, NahD, NahH e NahJ apresentam proteínas homólogas com mais de 85% de identidade já depositadas no banco de dados. De fato, essas proteínas já caracterizadas estruturalmente (também de espécies do gênero *Pseudomonas*) possuem a mesma função e substrato das enzimas de *P. putida* G7. Portanto, essas enzimas não seriam, inicialmente, candidatas prioritárias para a determinação de suas estruturas tridimensionais. Por outro lado, as enzimas NahAa, NahB, NahE, NahG, NahI, NahK, NahL, NahM, NahN e NahO apresentam macromoléculas semelhantes com estruturas já resolvidas, porém com baixa identidade de sequência de aminoácidos (cerca de 26 a 61%). Logo, a resolução de suas estruturas tridimensionais seria mais interessante por apresentar um maior potencial para gerar conhecimento, principalmente relacionado à elucidação da relação estrutura-função destas enzimas.

Das enzimas listadas na Tabela 1, o grupo de Biologia Estrutural – BioEst do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, liderado pelo prof. Dr. Ronaldo Alves Pinto Nagem, já resolveu a estrutura de seis enzimas (destacadas em negrito), as quais estão em processo de publicação (NahB, NahG, NahI, NahK e NahL) ou já foram publicadas (NahF). Duas delas, NahK e NahL são os objetos de estudo deste trabalho.

Tabela 1: Análise de identidade das proteínas da via superior e inferior de degradação do naftaleno de *P. putida* G7 e de proteínas com estrutura tridimensional já elucidada depositada no PDB.

Enzima	PDB ID	Identidade	Enzima	PDB ID	Identidade
NahAa	1KRH	30%	NahG	4BJZ	27%
NahAb [53]	2QPZ	94%	NahH [54]	1MPY	85%
NahAc [55]	1EG9	97%	Nahl [56]	4I26	61%
NahAd [55]	1EG9	95%	NahJ [57]	1OTF	85%
NahB	2Y93	37%	NahK	2EB4	41%
NahC	2EHX	97%	NahL	1SV6	39%
NahD [58]	2IMD	87%	NahM	1NVM	55%
NahE	3PUD	30%	NahN	1UKB	59%
NahF [59]	4JZ6	100%	NahO	4JN6	58%

1.6. NahK, NahL, enzimas homólogas e a superfamília FAH

A quinta reação na via inferior de degradação do naftaleno em *P. putida* G7 é realizada pela enzima NahK, uma 4-oxalocrotonato descarboxilase (EC 4.1.1.77) [40]. Nessa reação, utilizando um íon de magnésio como cofator, a molécula 2-oxo-3-hexenediotato é convertida a 2-hidroxi-2,4-pentadienoato com a consequente liberação de uma molécula de CO₂ (Figura 8). Através da tradução da sequência nucleotídica do gene *nahK* do plasmídeo NAH7 é obtida uma sequência de 264 resíduos de aminoácidos para uma proteína de massa teórica de 28.435 kDa. NahK é uma enzima homóloga a Xyll, uma vez que ambas possuem 94% de identidade entre suas sequências de aminoácidos, agem sobre o mesmo substrato, e provavelmente compartilham o mesmo enovelamento proteico.

NahK também catalisa uma reação de descarboxilação bastante semelhante à da enzima HpcE, presente na via de degradação do homoprotocatecuato (Figura 7). Entretanto, HpcE possui 427 resíduos de aminoácidos e uma massa teórica de 46.878 kDa. Foi reportado que essa proteína é uma enzima bifuncional com função de descarboxilase e isomerase, possuindo para isso, dois domínios catalíticos independentes [60]. NahK compartilha baixa identidade de sequência de aminoácidos com HpcE (a ferramenta Protein BLAST [61] não consegue alinhar as sequências proteicas de NahK e HpcE, mesmo considerando apenas a sequência de aminoácidos do domínio catalítico responsável pela descarboxilação em HpcE). Apesar da estrutura de HpcE já ter sido descrita, não se sabe ao certo o mecanismo de ação dessa enzima. NahK, Xyll e HpcE fazem parte da superfamília FAH, e como catalisam reações

similares, a resolução da estrutura de NahK também pode fornecer informações a respeito do mecanismo de ação de XylI e HpcE.

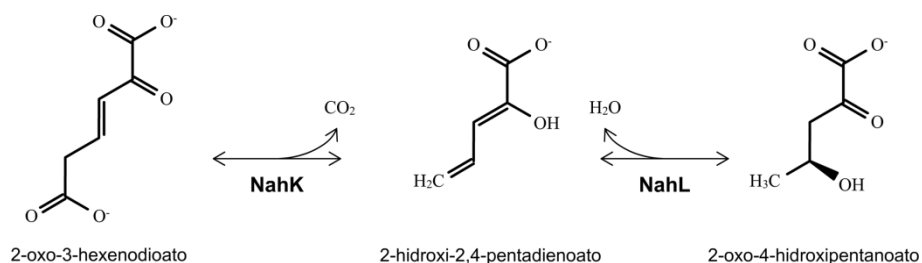


Figura 8: Reação de descarboxilação e hidratação catalisadas pelas enzimas NahK e NahL, respectivamente.

NahL, a próxima enzima na via de degradação do naftaleno, é denominada vinilpiruvato hidratase (EC 4.2.1.80). A tradução da sequência de nucleotídeos do gene *nahl* gera uma proteína com 261 resíduos de aminoácidos com uma massa teórica de 27.899 kDa. NahL hidrata o produto de NahK, gerando a molécula 2-oxo-4-hidroxi-pentanoato (Figura 8), e é homóloga à XylJ, compartilhando 90% de identidade entre suas sequências de aminoácidos. Ambas agem sobre o mesmo substrato e catalisam a mesma reação de hidratação da enzima MhpD de *E. coli* [62], presente na via de degradação do *m*-hidroxifenilpropionato; reação semelhante também àquela catalisada pela enzima HpcG da via do homoprotocatecuato (Figura 7) [63]. Apesar disso, NahL possui apenas 41 e 37% de identidade com MhpD e HpcG, respectivamente. NahL, XylJ, MhpD e HpcG também pertencem à superfamília FAH. NahK e NahL compartilham 39% de identidade entre suas sequências de aminoácidos, podendo também compartilhar o mesmo enovelamento proteico.

As enzimas XylI e XylJ são estudadas há mais de 30 anos devido à via de *meta*-clivagem do catecol. Em 1989 foi reportado que XylI e XylJ formam um complexo *in vivo*, mas o tamanho de sua estrutura quaternária era incerto [64]. É provável que NahK e NahL também formem um complexo proteico. Sabe-se também que o produto da reação de NahK é instável, sofrendo uma isomerização não enzimática em meio aquoso. Dessa forma, a formação do complexo entre uma 4-oxalocrotonato descarboxilase e uma vinilpiruvato hidratase auxiliaria no rápido processamento desse intermediário entre a reação de descarboxilação e de hidratação. Assim, há pelo menos 15 anos, grupos de pesquisa tentam obter a estrutura cristalográfica desse complexo enzimático para elucidar os mecanismos de

reação dessas duas enzimas, para saber como ocorre a montagem de sua estrutura quaternária, e também se há comunicação entre os sítios ativos [65]. Estudos para tentar resolver essas questões são dificultados pelas baixas quantidades de enzima purificada de extrato bruto de *Pseudomonas sp.*, pela incerteza sobre os intermediários da reação, e também pelo fato das duas enzimas existirem sob a forma de um complexo.

A enzima fumarilacetoacetato hidrolase de camundongo (PDB ID: 1QCN [66]), relacionada ao metabolismo de aminoácidos aromáticos, foi resolvida estruturalmente e descrita em 1999 como uma proteína com um enovelamento único até então. Posteriormente, por meio da elucidação de estruturas proteicas de outros organismos, esse enovelamento foi descrito como conservado de bactérias a mamíferos e, dessa forma, a estrutura da enzima FAH acabou definindo o enovelamento geral característico de sua superfamília. A superfamília FAH tem crescido consideravelmente e atualmente inclui enzimas com diversas funções como, por exemplo, hidrólise [66,67], isomerização [60], desidratação [68], hidratação [62,63] e descarboxilação [60,69]. A maioria dessas proteínas liga-se a um íon metálico em seu sítio ativo, mas os mecanismos de reação são bastante variados dentre os diferentes tipos enzimáticos.

FAH, assim como a maioria dos membros de sua superfamília, é um homodímero (Figura 9 a) e cada subunidade possui 46 kDa de massa molecular [66]. A estrutura tridimensional da enzima FAH exibe dois domínios principais, o primeiro é um domínio N-terminal composto por aproximadamente 120 resíduos que se enovelam em um sanduíche α/β (destacado em azul na Figura 9 a); o segundo é um grande domínio C-terminal contendo cerca de 300 resíduos (destacados em amarelo e salmão) enovelados em uma estrutura semelhante a um barril formado por uma folha- β torcida envolta por hélices e voltas (amarelo), e que também contém um segundo elemento estrutural formado por uma segunda folha- β mais plana (destacada em salmão).

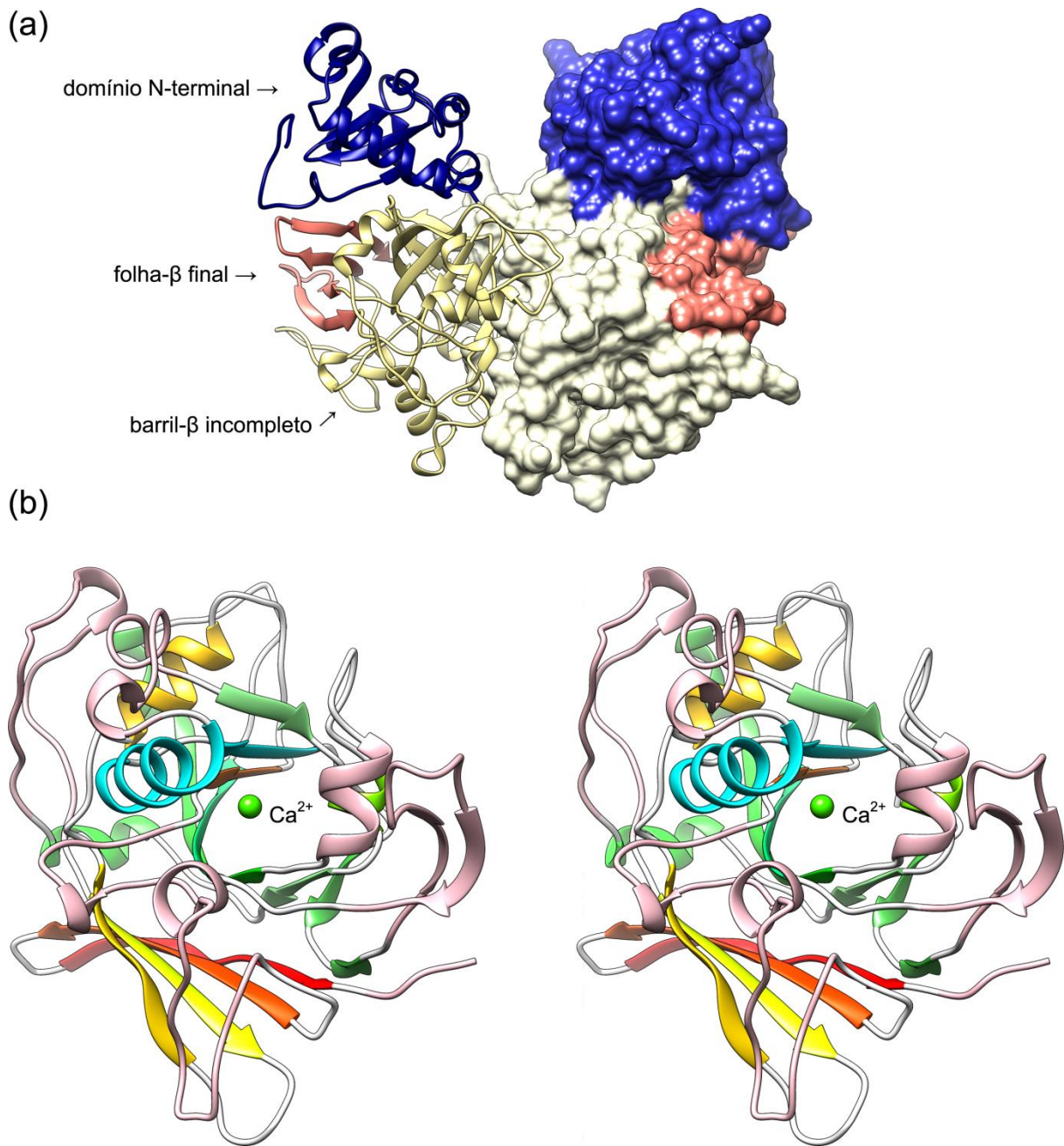


Figura 9. Representação da estrutura da enzima FAH obtida a partir do código PDB 1QCN. **(a)** representação em *ribbon* do enovelamento da cadeia A e da superfície da cadeia B do dímero de FAH. O domínio N-terminal está representado em azul, enquanto o domínio C-terminal está representado em amarelo (barril-β incompleto e anexos) e salmão (folha-β final). **(b)** representação em estéreo (*cross-eyed*) da cadeia principal de FAH compreendida entre os resíduos 121 e 417. O domínio N-terminal não está exibido nesta representação. Um íon cálcio (esfera verde) ligado à proteína se encontra no topo da estrutura em barril-β. O esquema de cores segue o padrão da Figura 27 para facilitar a comparação visual entre as estruturas de FAH e NahK. As partes em rosa são estruturas em FAH que não possuem contraparte em NahK.

A estrutura em barril não possui as ligações de hidrogênio entre suas fitas- β inicial e final, sendo chamado então de barril- β incompleto. Na enzima FAH as laterais da folha- β do barril são fechadas por uma volta β tipo IV formada pelos resíduos 162 a 165 (Figura 10, região destacada em vermelho). Essa mesma região do barril é responsável pela maior parte das interações intermoleculares da interface do dímero e encontra-se protegida do solvente. No topo do barril há um átomo de cálcio sendo coordenado pelos resíduos Asp126, Glu199, Glu201 e Asp233, e também por uma molécula de água e um íon acetato. Esse íon metálico, reportado como essencial para a reação enzimática, se encontra no fundo de um bolsão de aproximadamente 15 Å de profundidade. Portanto, o domínio contendo o barril- β incompleto foi reportado como possuindo os papéis de ligação ao átomo de metal, catálise e dimerização [66].

A ferramenta Protein BLAST (em sua configuração *default*) também não consegue fazer o alinhamento entre as sequências de aminoácidos de NahK e NahL com a enzima FAH. Apesar disso, sabe-se que as enzimas HpcG e MhpD, que possuem certa identidade com NahK e NahL, fazem parte da superfamília FAH, e também possuem em suas estruturas terciárias o barril- β incompleto anexado à folha- β C-terminal, característico dessa superfamília. Assim, supõe-se que NahK e NahL também possuam esse enovelamento característico.

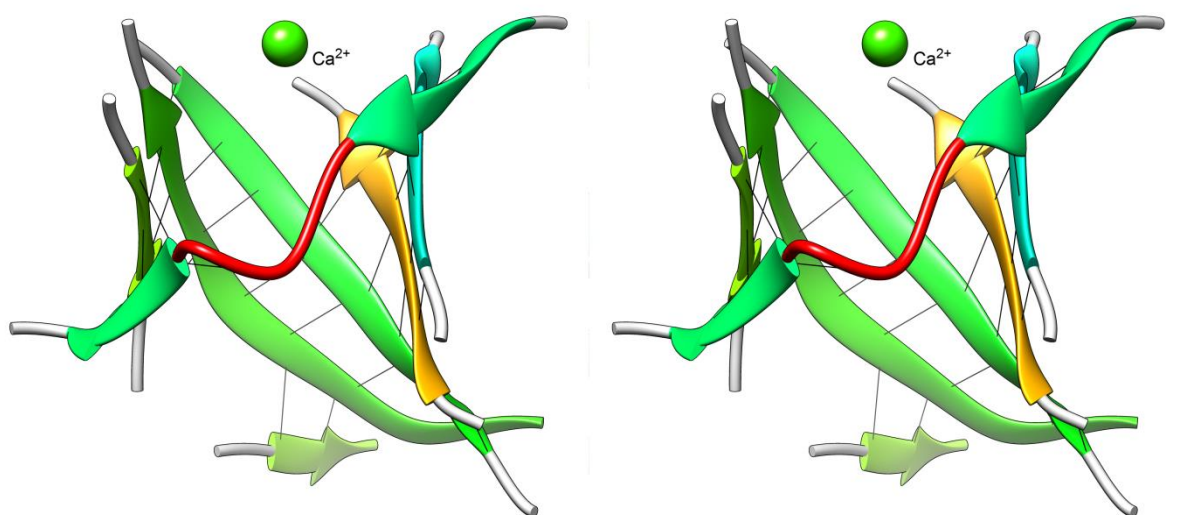


Figura 10. Representação em estéreo (*cross-eyed*) do barril- β incompleto da enzima FAH. O íon cálcio (esfera verde) ligado à proteína se encontra no topo do barril. As ligações de hidrogênio entre as fitas- β do barril estão representadas por linhas contínuas. A volta β tipo IV que fecha as extremidades da folha- β do barril está destacada em vermelho.

2. OBJETIVOS

2.1. Objetivo geral

Caracterizar bioquímica e estruturalmente as enzimas recombinantes 4-oxalocrotonato descarboxilase e vinilpiruvato hidratase, assim como seu complexo, codificadas pelos genes *nahK* e *nahL* provindos do plasmídeo NAH7 de *P. putida* G7.

2.2. Objetivos específicos

- Obter as proteínas recombinantes NahK e NahL, tanto separadamente quanto em sua forma complexada, puras e em grandes quantidades (escala de mg).
- Caracterizar o estado oligomérico do complexo NahK/NahL.
- Determinar os parâmetros cinéticos das enzimas utilizando seus substratos naturais.
- Cristalizar as proteínas recombinantes isoladas e em sua forma complexada.
- Resolver e analisar as estruturas cristalográficas das proteínas através de Difração de raios-X para entender melhor a relação estrutura-função das mesmas.

3. MATERIAIS E MÉTODOS

A metodologia para o cultivo de *Pseudomonas putida* G7 e a extração do plasmídeo NAH7, utilizado para as amplificações dos genes *nahK* e *nahL*, foi realizada durante o mestrado e apresentados na forma de uma dissertação [70], portanto não serão apresentados aqui.

3.1. Eletroforese em gel de agarose

Todas as amostras de DNA analisadas por eletroforese foram feitas em géis de agarose 1% (m/v) preparados como descrito a seguir.

Agarose (1% m/v) foi dissolvida por aquecimento em tampão TAE (TRIS-acetato 40 mM pH 8,0 e EDTA 1 mM). Para visualização das bandas de DNA foi adicionado SYBR Safe 0,1 µL/mL (Invitrogen). O tampão utilizado na corrida eletroforética foi o mesmo utilizado na confecção do gel. Após a polimerização do gel, as amostras foram aplicadas nas canaletas e a eletroforese executada a 80 V por 45 min utilizando o sistema Mini-Sub Cell GT (Bio-Rad) e fonte FB300 (Fisher Scientific). Para visualização do resultado, o gel foi exposto à luz ultravioleta no transiluminador UV 302 nm T26M (Bio Agency) e fotodocumentados utilizando câmera Vilber Lourmat, ou utilizando o sistema Alphamager HP (ProteinSimple). O SYBR Safe foi utilizado como fluoróforo no lugar do brometo de etídio, pois ele apresenta baixa ou nenhuma atividade mutagênica [71].

3.2. Cepas bacterianas e preparo de células eletrocompetentes

Células da linhagem *E. coli* DH5α (Life Technologies) foram utilizadas para replicação das construções plasmidiais. Essa cepa de *E. coli* possui uma mutação no gene *endA*, representada no genótipo por *endA1*. O gene *endA* selvagem codifica a Endonuclease I, que é uma proteína periplasmática de 12 kDa capaz de degradar DNA de dupla fita. Cepas de *E.*

coli com essa mutação produzem uma forma inativa da nuclease e são referidas como *EndA-*, permitindo assim um significativo aumento no rendimento e qualidade de DNA plasmidial usado para sequenciamentos ou outros ensaios [72]. Essa cepa também carrega outros genes modificados, como a mutação *hsdR17* que elimina a endonuclease de restrição do sistema de restrição-modificação *EcoKI*, permitindo que DNA não metilado por esse sistema não seja degradado [73]. Também está presente o alelo $\Delta(lacZ)M15$, permitindo a coloração azul/branca para identificação de colônias positivas para genes clonados em vetores baseados no sistema *lacZ* [74].

Células da linhagem *E. coli* BL21(DE3) (Novagen) foram utilizadas para expressão das proteínas recombinantes. As principais características dessa são: derivada da cepa *E. coli* B834 que é deficiente das proteases *lon* e *ompT*, as quais podem degradar proteínas durante a etapa de purificação; essa cepa também carrega o gene para a enzima T7 RNA Polimerase, essencial para a superexpressão de genes clonados em vetores que carregam o promotor T7 [75]. O gene da T7 RNA polimerase está sob o controle do promotor *Lac* e sua expressão é induzida por lactose ou IPTG.

Essas cepas foram utilizadas no preparo de células eletrocompetentes de acordo com o protocolo presente no manual do eletroporador MicroPulser (Bio-Rad), como descrito a seguir.

Inicialmente, foi feito um preinóculo da cepa de interesse em 5 mL de meio de cultura Luria-Bertani (meio LB) líquido que foi incubado 16 h a 37 °C sob agitação de 150 rpm. Após o crescimento, o preinóculo foi utilizado como inóculo em 500 mL de meio LB líquido. As células do inóculo cresceram a 37 °C sob agitação de 200 rpm, sendo o crescimento monitorado em espectrofotômetro até a cultura alcançar uma leitura de OD_{600} entre 0,5 e 0,7. Ao alcançar o nível de crescimento desejado, a cultura foi resfriada em gelo por 20 min e em todos os passos subsequentes, sempre que possível, as células e todos os recipientes foram mantidos em gelo.

As próximas etapas consistem em 5 ciclos de centrifugações a 4.000 x *g*, por 15 min a 4 °C, sendo o sobrenadante cuidadosamente descartado, seguidos de ressuspensões das células em volumes cada vez menores de glicerol 10% estéril e gelado. Inicialmente em volume de 500 mL, sendo o volume de ressuspensão reduzido para 250 mL, 20 mL e por fim

2 mL. Após serem ressuspendidas no volume final, as células foram imediatamente armazenadas em microtubos contendo alíquotas de 50 µL e incubadas em gelo até serem transportadas e armazenadas a -80 °C, estando assim prontas para uso.

3.3. Transformação de *Escherichia coli* por eletroporação

As transformações das cepas foram realizadas através de eletroporação em aparelho eletroporador MicroPulser (Bio-Rad), utilizando o protocolo descrito em seguida.

Cubetas para eletroporação de 0,2 cm (Bio-Rad) e alíquotas de células eletrocompetentes foram mantidas em gelo por cerca de 5 min. Posteriormente, 3 a 5 µL de produtos de ligação de fragmentos a vetores, ou plasmídeos purificados, foram incubados com 50 µL de células eletrocompetentes e mantidos em gelo por mais 2 min. Em seguida, as células foram transferidas para a cubeta e a eletroporação realizada utilizando o programa EC2 do aparelho. Imediatamente após a eletroporação, foram acrescentados à cubeta 500 µL de meio LB líquido sem antibiótico e as células foram transferidas para um microtubo de 1,5 mL. A cultura foi incubada a 37 °C por 1 h sob agitação de 200 rpm.

Finalizada a incubação, as células foram plaqueadas em meio seletivo LB-ágar 1,5% contendo o antibiótico para o qual o vetor inserido e/ou a cepa conferem resistência. As placas foram incubadas a 37 °C por 16 h. Os antibióticos utilizados, bem como suas respectivas concentrações de estoque e de uso são exibidos na Tabela 2.

Tabela 2. Antibióticos utilizados para a seleção dos respectivos plasmídeos.

Antibiótico	Plasmídeo	Concentração estoque	Concentração de uso
Ampicilina	pGEM-T / pETDuet-1	100 mg/mL em água	100 µg/mL
Kanamicina	pET28a-TEV	50 mg/mL em água	50 µg/mL

3.4. PCR de colônias

Após a eletroporação das células e crescimento de colônias nas placas contendo meio seletivo, algumas das colônias foram selecionadas, replicadas em uma nova placa e analisadas por PCR de colônias utilizando o mesmo protocolo de amplificação dos respectivos genes de interesse (Tabela 5). Na PCR de colônias não se coloca amostra de DNA e em seu lugar é utilizada uma pequena amostra da colônia que se deseja analisar. As amostras amplificadas por PCR foram analisadas através de eletroforese em gel de agarose. Se a colônia analisada for proveniente de um clone transformado pelo vetor que contém o gene de interesse haverá a amplificação desse gene e a visualização de uma banda no gel de agarose na altura esperada de acordo com o tamanho do gene. Dessa forma, colônias positivas foram selecionadas, cultivadas por 16 h em meio LB líquido contendo o antibiótico respectivo para cada vetor (Tabela 2) e, posteriormente, seus plasmídeos foram extraídos e purificados utilizando kit. Colônias positivas também foram armazenadas a -80 °C em glicerol 30%.

3.5. Extração de DNA plasmidial

Clones positivos ou células contendo os plasmídeos de expressão foram inoculados em 10 mL de meio LB líquido contendo o antibiótico específico para cada vetor e cultivados por 16 h a 37 °C sob agitação de 200 rpm para posterior extração dos plasmídeos. As extrações e purificações foram realizadas utilizando kits AxyPrep Plasmid Miniprep (Axygen Biosciences), illustra plasmidPrep Mini Spin Kit (GE Healthcare), QIAprep Spin Miniprep Kit (QIAGEN), ou Wizard Plus Minipreps DNA Purification System (Promega). Em todos os casos foi seguido o protocolo do fabricante.

Todos os kits possuem um procedimento semelhante, que consiste no crescimento das colônias, centrifugação das células e descarte do sobrenadante, ressuspensão das células e posterior lise alcalina. A lise faz extravasar o conteúdo celular juntamente com os plasmídeos. Após a lise, neutraliza-se o pH da solução com uma base fraca, comumente

acetato de potássio. A amostra é então centrifugada e a fase aquosa aplicada na coluna do kit, que possui a capacidade de reter ácidos nucleicos. Após uma ou mais etapas de lavagem para remover contaminantes, o DNA é eluído da coluna e sua quantidade e pureza analisada em gel de agarose ou aparelho espectrofotométrico (Nanodrop).

3.6. Desenho dos oligonucleotídeos iniciadores

As sequências nucleotídicas dos genes *nahK* e *nahL* foram obtidas a partir da sequência do plasmídeo NAH7 de *P. putida* G7 depositada no banco de dados nucleotídicos do *National Center for Biotechnology Information* - NCBI (GenBank: AB237655.1) [40]. Os iniciadores foram desenhados no programa Primer Premier 5.0. De acordo com a estratégia de clonagem utilizando o vetor de expressão pET28a-TEV [14] e pETDuet-1 (Novagen), foram inseridas nos iniciadores pequenas sequências reconhecidas por enzimas de restrição compatíveis com os sítios múltiplos de clonagem desses vetores. Cada par de iniciadores direto e reverso foram ajustados para que suas T_m fossem próximas e para minimizar a formação de estruturas secundárias e anelamentos em regiões inespecíficas do plasmídeo NAH7. Os iniciadores, sítios das enzimas de restrição e T_m estão representados na Tabela 3. Os nucleotídeos sublinhados correspondem às sequências reconhecidas pelas respectivas enzimas de restrição.

Tabela 3. Oligonucleotídeos sintéticos utilizados como iniciadores para amplificação dos genes *nahK* e *nahL*.

Iniciadores	Sequências nucleotídicas	Enzimas de restrição	T_m (°C)
nahK-NheI-F*	5' <u>GCT AGC</u> ATG AAC CGT ACC CTG ACC CGC 3'	<i>NheI</i>	65,3
nahK-HindIII-R*	5' <u>AAG CTT</u> TCA GAT AAA GCG CGC GGA TAC CG 3'	<i>HindIII</i>	66,6
nahK-F-pETd	5' <u>GGA TCC</u> <u>G GAA AAC CTG TAT TTT CAG GGC</u> ATG AAC CGT ACC CTG ACC C 3'	<i>BamHI</i>	56,6
nahK-R-pETd	5' <u>AAG CTT</u> CGG GAT TTC AGA TAA AG 3'	<i>HindIII</i>	56,0
nahL-F-pET	5' <u>CAT ATG</u> GAC CAG ATA TTG ATC AAG CAG 3'	<i>NdeI</i>	60,0
nahL-R-pET	5' <u>CTC GAG</u> GCT CAT TCG CTT GTC CTC TG 3'	<i>XhoI</i>	58,2

*Oligos desenhados e utilizados durante o mestrado [70].

Os iniciadores nahK-NheI-F e nahK-HindIII-R são para amplificação e clonagem do gene *nahK* no vetor pET28a-TEV, enquanto que os iniciadores nahK-F-pETd e nahK-R-pETd são para amplificação e clonagem do gene *nahK* no primeiro sítio múltiplo de clonagem do vetor

pETDuet-1. Os iniciadores nahL-F-pET e nahL-R-pET servem tanto para amplificação e clonagem do gene *nahL* no vetor pET28a-TEV quanto para clonagem no segundo sítio múltiplo de clonagem do vetor pETDuet-1.

O iniciador direto nahK-F-pETd contém uma sequência nucleotídica adicional destacada em vermelho que corresponde a sequência de aminoácidos reconhecida e clivada pela protease TEV. Para manter a fase de leitura, o nucleotídeo destacado em azul foi adicionado entre a sequência da enzima de restrição e a sequência do sítio de clivagem da protease. A adição desse sítio de clivagem permitirá a remoção da cauda de afinidade de NahK expressa em pETDuet-1. O vetor pET28a-TEV já possui essa sequência incorporada no vetor, não sendo necessária sua inclusão nos seus respectivos iniciadores.

Os iniciadores foram recebidos liofilizados. O material foi ressuscitado em H₂O milli-Q estéril a 37 °C e posteriormente armazenado a -20 °C em pequenas alíquotas estoque na concentração de 100 µM. Para uso dos iniciadores, uma alíquota foi diluída para 10 µM utilizando H₂O milli-Q autoclavada.

3.7. Amplificação dos genes e clonagem no vetor pGEM-T

As amplificações dos genes *nahK* e *nahL* foram realizadas através de PCR tendo como molde de DNA o plasmídeo NAH7, e a temperatura durante a etapa de anelamento foi determinada a fim de se obter o mínimo de amplificações inespecíficas e o máximo de amplificação no tamanho esperado do gene. Os reagentes, suas quantidades e suas respectivas concentrações utilizadas nas amplificações estão representados na Tabela 4. O programa utilizado para as amplificações está representado na Tabela 5. Os fragmentos amplificados dos genes foram visualizados através de eletroforese em gel de agarose.

As amostras contendo o material amplificado utilizando os respectivos iniciadores para *nahK* e *nahL*, após purificadas diretamente da solução de PCR utilizando kit e dosagem em aparelho espectrofotômetro, foram ligadas ao vetor pGEM-T (Promega) conforme instruções do fabricante. Após a reação de ligação, os plasmídeos foram inseridos por eletroporação

em células de *E. coli* DH5 α eletrocompetentes e clones positivos foram selecionados por PCR de colônias.

Tabela 4. Reagentes e quantidades utilizadas nas reações de PCR.

Reagente	Volume	Concentração final
Plasmídeo NAH7 de <i>P. putida</i> G7 a 30 ng/ μ L	5 μ L	3 ng/ μ L
Tampão 10X	5 μ L	1X
MgSO ₄ 50 mM	2 μ L	2 mM
Platinum® Taq DNA Polymerase High Fidelity 5 U/ μ L (Invitrogen)	0,2 μ L	1 U
dNTPs 10 mM	1 μ L	0,2 mM
Iniciador direto 10 μ M	1 μ L	0,2 μ M
Iniciador reverso 10 μ M	1 μ L	0,2 μ M
H ₂ O milli-Q autoclavada	q.s.p. 50 μ L	-

Tabela 5. Programa para amplificação dos genes *nahK* e *nahL*.

Etapa	Temperatura	Tempo
1. Início	94 °C	5 min
2. Desnaturação	94 °C	1 min
3. Anelamento	58 °C (<i>nahK</i> para pET28a-TEV)	1 min
	53 °C (<i>nahL</i> para pET28a-TEV)	
	53 °C (<i>nahK</i> e <i>nahL</i> para pETDuet-1)	
4. Extensão	72 °C	1 min
5. Repetição das etapas 2 a 4 (24 vezes)	-	-
6. Extensão final	72 °C	5 min

3.8. Subclonagem dos genes no vetor pET28a-TEV e pETDuet-1

Aproximadamente 5 μ g dos vetores pGEM-T contendo os genes *nahK* e *nahL* foram digeridos por 16 h a 37 °C utilizando as respectivas enzimas de restrição na concentração de 10 unidades por μ g de DNA. A fim de gerar pontas coesivas compatíveis com os insertos, o vetor pET28a-TEV também foi digerido por 16 h a 37 °C utilizando as mesmas enzimas de restrição na concentração de 20 unidades por μ g de DNA. Todas as reações de digestão foram feitas em tampão Tango 2X (Fermentas). Insertos e vetores digeridos foram submetidos à eletroforese em gel de agarose e as bandas respectivas foram excisadas do gel, purificadas utilizando kit e sua concentração foi dosada em espectrofotômetro. Antes das reações de digestão final, a digestão dos vetores foi testada em pequena escala com cada enzima separadamente e em reação de dupla digestão. Posteriormente, os insertos foram

ligados separadamente ao vetor de expressão pET28a-TEV utilizando a enzima T4 DNA ligase (Promega), conforme recomendação do fabricante, e inseridos por eletroporação em células de *E. coli* BL21(DE3) eletrocompetentes. Clones positivos foram selecionados por PCR de colônias e armazenados em glicerol 30% a -80 °C. Clones selecionados também tiveram seus plasmídeos extraídos e purificados para sequenciamento.

Para a construção do vetor de coexpressão o inserto contendo o gene *nahL*, previamente digerido e extraído do vetor pGEM-T, foi ligado ao vetor pETDuet-1 previamente digerido com as enzimas *NdeI* e *XhoI* na concentração de 20 unidades por µg de DNA, e inserido por eletroporação em células de *E. coli* DH5α. Colônias positivas para o gene *nahL*, selecionadas por PCR de colônias, foram crescidas em meio LB líquido contendo ampicilina e tiveram seus plasmídeos extraídos e purificados utilizando kit. Foi realizado um sequenciamento para confirmar a sequência e correta inserção do gene *nahL* no segundo sítio múltiplo de clonagem do vetor pETDuet-1. Posteriormente, esse vetor, já contendo o gene *nahL*, foi digerido utilizando as enzimas *BamHI* e *HindIII*, ligado ao inserto contendo o gene *nahK* proveniente da digestão do vetor pGEM-T e inserido por eletroporação em células de *E. coli* BL21(DE3). Colônias positivas para ambos os genes foram selecionadas por PCR de colônias, crescidas em meio LB líquido com ampicilina e armazenadas em glicerol 30% a -80 °C; essas colônias também tiveram seus plasmídeos extraídos, purificados e sequenciados para confirmar a sequência e correta inserção do gene *nahK* no primeiro sítio múltiplo de clonagem do vetor pETDuet-1.

A Figura 11 exibe o esquema das construções finais em cada vetor de expressão e as respectivas enzimas de restrição utilizadas para a correta ligação de cada inserto nos devidos sítios.

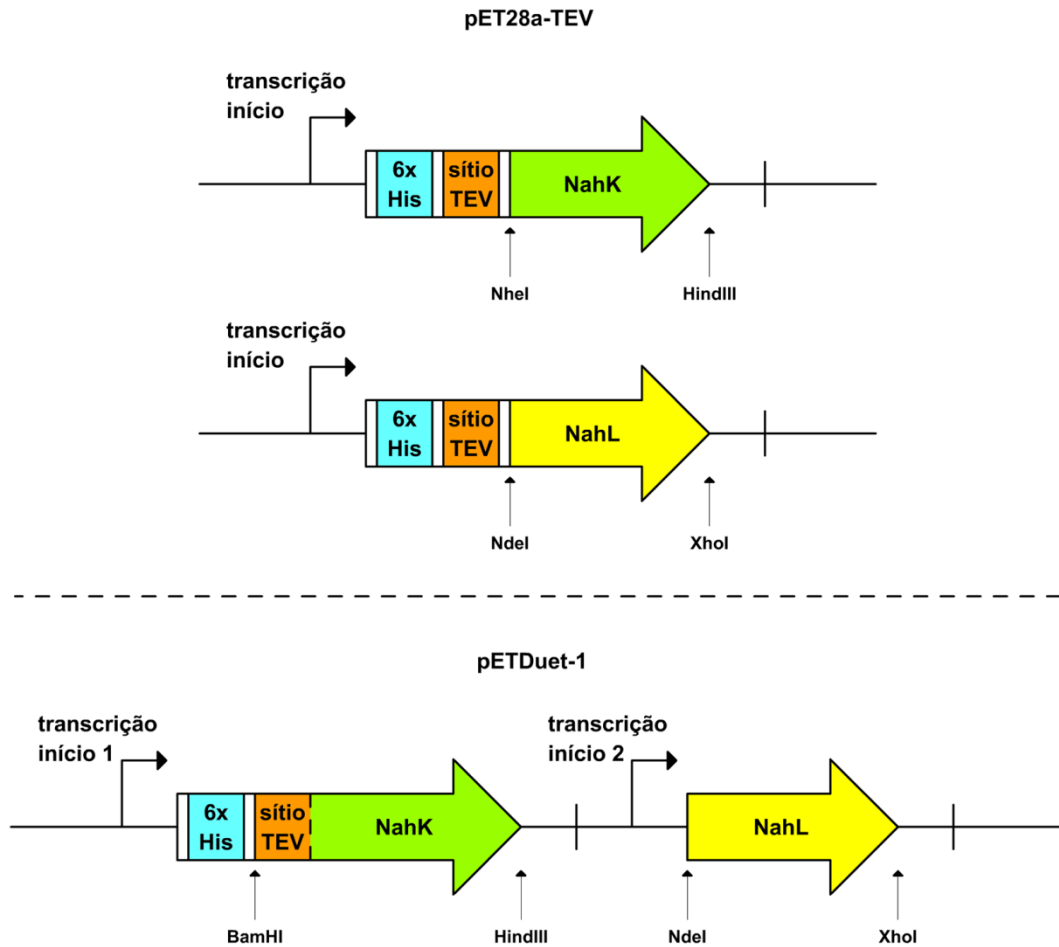


Figura 11. Esquema das construções finais nos vetores de expressão.

3.9. Sequenciamento dos clones positivos

Os sequenciamentos dos plasmídeos recombinantes foram feitos pelo método didesoxiterminal [76] no Núcleo de Análise do Genoma e Expressão Gênica (NAGE / ICB / UFMG) com auxílio do DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE DNA Analysis (Amersham Biosciences). Para confirmação da integridade das sequências alvo clonadas no vetor pET28a-TEV foram utilizados os iniciadores T7 direto (5' TAA TAC GAC TCA CTA TAG GG 3') e T7 reverso (5' GCT AGT TAT TGC TCA GCG G 3'). Para confirmação da integridade da sequência do gene *nahK* inserida no primeiro sítio múltiplo de clonagem do vetor pETDuet-1 foram utilizados os iniciadores T7 direto e DuetDOWN1 (5' GAT TAT GCG GCC GTG TAC AA 3'); para o gene *nahL*, inserido no segundo sítio múltiplo de clonagem do

vetor pETDuet-1, foram utilizados os iniciadores DuetUP2 (5' TTG TAC ACG GCC GCA TAA TC 3') e T7 reverso. A análise do sequenciamento foi feita no programa Chromas Lite 2.01 (Technelysium).

3.10. Expressão das proteínas recombinantes

Clones selecionados para expressão foram preinoculados em 20 mL de meio LB líquido contendo o respectivo antibiótico para cada vetor de expressão e incubados a 37 °C por 16 h sob agitação de 200 rpm. Em seguida, a cultura foi utilizada como inóculo na proporção de 1:100 de cultura em 2 L de meio LB líquido com antibiótico. As células do inóculo foram cultivadas a 37 °C sob agitação de 200 rpm e seu crescimento foi monitorado em espectrofotômetro até a cultura alcançar OD₆₀₀ entre 0,5 e 0,7. Nesse momento, foi adicionado o indutor de expressão IPTG para a concentração final de 0,5 mM. A expressão foi feita a 30 °C por 5 h sob agitação de 200 rpm. Após a etapa de expressão, a cultura foi centrifugada a 5.000 x *g* a 4 °C por 5 min, sendo o sobrenadante descartado e as células armazenadas a -20 °C.

Para verificar a expressão da proteína recombinante, 1 mL do inóculo foi retirado imediatamente antes da indução por IPTG, centrifugado a 10.000 x *g* por 5 min e as células ressuspensas em 50 µL de água e 10 µL de tampão de amostra de proteína 6X (SDS 10% m/v, glicerol 30% v/v, azul de bromofenol 1% m/v, TRIS-HCl 250 mM pH 6,8). Após o término da expressão, a OD₆₀₀ da cultura foi novamente medida, e uma quantidade equivalente de células foi coletada (quando comparada com a OD₆₀₀ antes da indução). A amostra foi centrifugada e as células ressuspensas em 50 µL de água e 10 µL de tampão de amostra de proteína 6X. Posteriormente as amostras foram incubadas em banho-maria a 100 °C por 5 min e submetidas à SDS-PAGE para verificar a presença de bandas de superexpressão.

3.11. Lise das células e purificação da fração solúvel

Para a etapa de lise, as células foram ressuspensas em *tampão de ligação A* (tampão fosfato de sódio 20 mM pH 7,4, NaCl 500 mM, imidazol 45 mM). O volume de tampão utilizado foi na proporção de 20 mL por litro de expressão para lise utilizando ciclos de congelamento e sonicação, ou na proporção de 40 mL por litro de expressão para lise utilizando homogeneizador. Para o processo de lise por ciclos de congelamento e sonicação foi adicionada lisozima à suspensão de células para a concentração final de 100 µg/mL, sendo as células mantidas em gelo por 30 min antes do início dos ciclos de congelamento. Em seguida, o material foi congelado a -20 °C e descongelado a temperatura ambiente por três ou quatro vezes. Após os ciclos de congelamento, o material lisado foi mantido em gelo e sonicado utilizando o aparelho Sonic Dismembrator 500 (Fisher Scientific) em amplitude de 30% com 5 pulsos de 10 s ligado e 1 min desligado. Para o processo de lise utilizando homogeneizador, as células foram ressuspensas em *tampão de ligação A* e passadas três ou quatro vezes no homogeneizador EmulsiFlex-C3 (Avestin) seguindo instruções do fabricante.

Após o processo de lise, o material lisado foi centrifugado duas vezes a 12.000 x *g* a 4 °C por 15 min, sendo que o sobrenadante foi transferido para um novo tubo entre as etapas de centrifugação. Quando necessário, para clarificar a amostra, o sobrenadante foi filtrado utilizando filtro de 0,45 µm. A fração insolúvel, correspondente ao precipitado após o lisado ser centrifugado, foi ressuspensa em água no mesmo volume utilizado para a lise e uma amostra foi retirada para análise por SDS-PAGE. Uma amostra para análise também foi retirada da fração solúvel, correspondente ao sobrenadante do material lisado.

Após a centrifugação, a fração solúvel do lisado foi imediatamente aplicada a 1 mL/min em coluna de afinidade Histrap HP de 5 mL (GE Healthcare), previamente equilibrada com *tampão de ligação A*, a fim de realizar a primeira etapa de purificação por IMAC. Posteriormente, a coluna foi lavada a 5 mL/min utilizando 50 mL do mesmo tampão de ligação. Para recuperar as proteínas que ficaram ligadas à resina, foi passado pela coluna a 5 mL/min 50 mL de *tampão de ligação A* em um gradiente crescente com o *tampão de eluição A* (tampão fosfato de sódio 20 mM, NaCl 500 mM, imidazol 500 mM). Frações

correspondentes ao pico de eluição contendo a proteína de interesse foram unidas e desalinizadas em coluna HiPrep 26/10 (GE Healthcare) previamente equilibrada e eluída com *tampão final* (tampão TRIS-HCl 50 mM pH 7,4 e NaCl 50 mM). A concentração das proteínas foi estimada através da absorbância a 280 nm utilizando para o cálculo os coeficientes de extinção molar e massas moleculares exibidos na Tabela 6 e calculados utilizando a ferramenta ProtParam do portal ExPASy [77].

Tabela 6. Coeficientes de extinção molar e massas moleculares das proteínas recombinantes.

Proteína	Coefficiente de extinção molar ($M^{-1} cm^{-1}$)	Massa molecular (kDa)
NahK com 6xHis-tag	21.430	30,725
NahK sem cauda	19.940	28,919
NahL com 6xHis-tag	18.450	29,913
NahL sem cauda	16.960	28,107
NahK/NahL com 6xHis-tag	38.390	58,800
NahK/NahL sem cauda	36.900	56,392

Para NahK recombinante e para o complexo NahK/NahL, após primeira etapa de purificação, foi realizada a clivagem da cauda de histidinas pela protease TEV. Para isso a protease foi aplicada diretamente na amostra utilizando a proporção de 5% w/w de protease/proteína recombinante, e incubado durante 4 horas a 25 °C; ou na proporção de 20% w/w e incubado por 16 h a 4 °C. Em ambas as condições também foi adicionado à amostra o agente redutor ditioneitol (DTT) na concentração final de 1 mM. Após a reação de clivagem, as amostras foram novamente aplicadas a 1 mL/min em coluna HisTrap HP para uma segunda etapa de purificação por IMAC. A coluna foi previamente equilibrada e lavada a 5 mL/min utilizando 50 mL de *tampão de ligação B* (TRIS-HCl 50 mM pH 7,5 e NaCl 50 mM); e o gradiente de eluição com 50 mL foi feito com *tampão de ligação B* e *tampão de eluição B* (TRIS-HCl 50 mM pH 7,5, NaCl 50 mM e imidazol 500 mM). As frações contendo as proteínas recombinantes sem a cauda de histidinas foram unidas e submetidas a dessalinização em coluna HiPrep 26/10 (GE Healthcare) previamente equilibrada com *tampão final*. Para um polimento na purificação, as amostras foram submetidas a uma etapa final de purificação por SEC realizada em coluna Superdex 200 10/300 GL ou em coluna HiLoad Superdex 16/600 75 PG (GE Healthcare). Amostras foram recolhidas ao longo dos processos de purificação, misturadas com tampão de amostra de proteína 6X, incubadas em banho-maria a 100 °C por 5 min e submetidas à análise por SDS-PAGE.

3.12. Eletroforese em Gel de Poliacrilamida-SDS

As amostras das expressões proteicas, lise e frações durante as etapas de purificação foram analisadas através de SDS-PAGE. Os géis foram preparados da mistura de N,N'-metileno-bis-acrilamida 0,8% e acrilamida 30% m/v. O gel de separação (12 ou 15%) foi feito adicionando-se a essa mistura tampão TRIS-HCl 1,5 M pH 8,8 e SDS 0,01% m/v. O gel foi polimerizado pela adição de persulfato de amônio (APS) a 10% m/v (concentração final de 0,5% v/v) e N,N,N',N'-tetrametiletilenodiamina (TEMED) 0,05% v/v. O gel de concentração (4%) foi preparado como descrito para o gel de separação, porém utilizando tampão TRIS-HCl 0,5 M pH 6,8. As quantidades dos materiais estão exibidas na Tabela 7, e são suficientes para montagem de 2 géis de 0,75 mm de espessura utilizando-se o aparato de montagem para a cuba Mini-PROTEAN Tetra Cell (Bio-Rad).

Tabela 7. Materiais e quantidades utilizadas na preparação dos géis para SDS-PAGE.

Reagente	Gel de separação 12 e 15%		Gel de concentração 4%
Água destilada	3,35 mL	2,29 mL	1,22 mL
TRIS-HCl 1,5 M pH 8,8	2,5 mL	2,5 mL	-
TRIS-HCl 0,5 M pH 6,8	-	-	0,5 mL
SDS 10%	100 µL	100 µL	20 µL
Bis-acrilamida 0,8%, acrilamida 30%	4 mL	5 mL	260 µL
APS 10%	100 µL	100 µL	20 µL
TEMED	10 µL	10 µL	2 µL

Para a corrida eletroforética foi utilizado tampão de eletrodo contendo TRIS-HCl 25 mM pH 8,3, glicina 192 mM, e SDS 0,1%, e a voltagem foi estabelecida a 110 V até que as amostras entrassem no gel de concentração e alcançassem o gel de separação. A partir desse momento, a voltagem foi elevada para 150 V até o final da corrida. Para geração de corrente e controle de voltagem foi utilizada a fonte FB300 (Fisher Scientific).

Para visualização das bandas, os géis foram colocados em solução corante por 2 horas (comassie blue R-250 0,5% m/v, ácido acético 10% e etanol 50%), e depois deixados em solução descorante (ácido acético 10% e etanol 25%) até o aparecimento das bandas proteicas e o descoramento das regiões do gel que não tinham proteínas. Os géis foram fotodocumentados utilizando o sistema Alphamager HP (ProteinSimple).

3.13. Espalhamento de Luz Dinâmico

A determinação do estado de agregação e provável oligomerização das proteínas NahK e do complexo NahK/NahL em solução foi estimada por ensaios de DLS. Os experimentos foram realizados utilizando o equipamento Zetasizer Nano ZS (Malvern Instruments). A amostra de NahK utilizada para esse experimento foi coletada após a segunda etapa de purificação por IMAC, enquanto que a amostra do complexo NahK/NahL foi coletada após a etapa final de purificação por SEC. As amostras proteicas foram mantidas a 4 °C até o momento do ensaio e, logo antes da análise, foram centrifugadas a 15.000 x *g* por 10 min para remoção de qualquer material suspenso. Os ensaios foram realizados na temperatura de 25 °C em cubeta de quartzo com 1 cm de caminho óptico e utilizando 1 mL de amostra a aproximadamente 1 mg/mL de concentração proteica. Foram coletadas automaticamente cerca de 100 a 250 medidas (dependendo da concentração proteica e do tamanho das partículas) e os dados foram analisados no programa Malvern Zetasizer software v. 6.01. Com essa análise, foram obtidos valores para a polidispersividade das proteínas em solução, o raio hidrodinâmico da molécula em nm e o peso molecular estimado da molécula em kDa. Para verificar se havia alguma partícula em solução que não fosse da proteína alvo foi feito, antes do experimento com a amostra proteica, uma coleta de 20 medidas do tampão no qual a proteína estava diluída e da cubeta vazia. Nestes casos, os valores do raio médio e do peso molecular médio devem ser próximos de zero.

3.14. Dicroísmo Circular

Para caracterização da estrutura secundária de NahK e do complexo NahK/NahL, assim como suas estabilidades térmicas, foram realizados experimentos de CD utilizando o aparelho Jasco J-815 (Jasco corporation). A concentração das proteínas neste experimento foi de aproximadamente 0,2 mg/mL em tampão TRIS-HCl 5 mM pH 7,4 e NaCl 5 mM. As amostras utilizadas, tanto de NahK quanto do complexo NahK/NahL, foram coletadas após purificação por SEC e foram mantidas a 4 °C até o momento do ensaio. Para a coleta de dados foram utilizados cubetas de quartzo com 1 ou 0,5 mm de caminho óptico,

comprimentos de onda de 190 a 260 nm e temperaturas crescentes de 20 a 80 °C em intervalos de 5 ou 10 °C. As medidas finais de cada curva são a média de no mínimo 10 acumulações para cada temperatura. Para estimar a porcentagem de cada tipo de estrutura secundária presente nas amostras, a deconvolução dos dados foi feita pelo programa CDNN v. 2.1 [78].

3.15. Espalhamento de Raios-X a Baixo Ângulo

Ensaio de SAXS do complexo NahK/NahL nas concentrações de aproximadamente 1,5, 3 e 6 mg/mL foram realizados na linha SAXS-2 do Laboratório Nacional de Luz Síncrotron (LNLS) em Campinas-SP utilizando um detector bidimensional MAR-165 para registrar as intensidades de espalhamento como função do vetor \mathbf{q} . O comprimento de onda utilizado foi $\lambda = 1,55 \text{ \AA}$ e a distância amostra-detector foi de 964 mm, de forma que o intervalo do módulo do vetor de espalhamento foi de $0,014 < q < 0,334 \text{ \AA}^{-1}$. Duas curvas de espalhamento da solução de proteína e do tampão foram coletadas em etapas de 300 e 600 segundos para monitoramento dos danos causados pela radiação à amostra, além da estabilidade do feixe, sendo as amostras mantidas em um porta amostras com 1 mm de espessura de janelas de mica à temperatura constante de 25 °C. Duas curvas foram coletadas para a concentração de 1,5 mg/mL em 300 e 600 segundos, uma curva para a concentração de 3 mg/mL em 600 segundos, e quatro curvas para a concentração de 6 mg/mL em 600 segundos. Subtraiu-se das curvas de espalhamento da solução de proteína o espalhamento da solução tampão nos respectivos tempos de 300 e 600 segundos (fosfato de sódio 50 mM pH 7,4, NaCl 50 mM), e a diferença entre as curvas foi normalizada pela concentração proteica.

A análise dos dados foi realizada utilizando uma série de programas dentro do pacote ATSAS [79]. A função de distribuição de distâncias $P(r)$ e o raio de giro foram avaliados pelo programa GNOM, a partir do qual também se obteve o diâmetro máximo do oligômero. Inicialmente foram gerados 12 modelos de átomos *dummy* no programa DAMMIN no modo *slow* utilizando uma simetria de partícula 1 (opção p1 no programa). Posteriormente à análise desses primeiros modelos, foram gerados outros 16 modelos com os mesmos

parâmetros de entrada exceto que foi imposta a simetria de partícula 5 (opção p5 no programa), e também 20 modelos utilizando a simetria de partícula 5-2 (opção p52 no programa). O processamento desses modelos foi feito com o pacote DAMAVER para avaliação de similaridade e clusterização. Para a superposição dos modelos de baixa (SAXS) e alta resolução (XRD), utilizou-se o programa Chimera [80]. Estimativas da massa molecular do modelo de baixa resolução obtido por SAXS foram obtidas pela ferramenta SAXS MoW (disponível no sítio <http://www.if.sc.usp.br/~saxs/saxsmow.html>) [81].

3.16. Cinética enzimática

Os ensaios de atividade enzimática para NahK com cauda de histidinas clivada e para o complexo NahK/NahL (também com cauda de histidinas clivada) foram realizados em espectrofotômetro Hewlett-Packard 8452A, ou Agilent 8453 em temperatura ambiente. A atividade de NahK na ausência de NahL foi monitorada pelo aumento da absorbância a 265 nm ($\epsilon = 7.850 \text{ M}^{-1} \text{ cm}^{-1}$) como resultado do aumento da concentração do produto da enzima. Para a atividade de NahK complexada com NahL foi monitorada o decréscimo da absorbância a 236 nm ($\epsilon = 6.580 \text{ M}^{-1} \text{ cm}^{-1}$) como resultado da diminuição da concentração do substrato, e para NahL complexada com NahK foi monitorado o decréscimo da absorbância a 265 nm ($\epsilon = 7.850 \text{ M}^{-1} \text{ cm}^{-1}$) também como resultado da diminuição da concentração do substrato. Os valores de coeficientes de extinção molar dos substratos e produtos das reações foram os mesmos utilizados por Stanley [65].

O substrato natural de NahK foi gerado pela incubação de 2HM com a enzima 4-OT de *P. putida* mt-2 (previamente clonada do plasmídeo TOL pWW0, expressada e purificada no laboratório do prof. Dr. Christian P. Whitman). 4-OT é uma enzima homóloga à NahJ, situada anteriormente à NahK na mesma via de degradação. Tanto a molécula 2HM quanto o substrato de NahL também foram produzidos pelo grupo do prof. Dr. Whitman, da Divisão de Química Médica da Faculdade de Farmácia da Universidade do Texas, nos Estados Unidos da América. Para cada ensaio de cinética da enzima NahK, foi utilizada uma pequena quantidade (cerca de 0,0030 g) da molécula 2HM solubilizada em etanol para concentração final de 20 mM. Para produzir o substrato de NahK, foram misturados 100 μL da solução de

2HM juntamente com 2 μL da enzima 4-OT a aproximadamente 1 mg/mL em 1 mL total de reação em tampão fosfato de sódio 10 mM pH 7,4. Após 5 minutos de reação, a solução contendo o substrato de NahK foi mantida em gelo até ser utilizada para os ensaios.

Para medir a atividade de NahL foi preparada uma solução estoque de HPD a 20 mM em etanol, e, a partir desta, foi feita uma solução de uso de HPD a 2 mM também em etanol. O uso de etanol na preparação da solução estoque e de uso de substrato deve-se ao fato do HPD ser isomerizado e hidratado não enzimaticamente em soluções aquosas, e mantendo-o em etanol conserva-se o substrato intacto por mais tempo já que a taxa de isomerização e hidratação nesse solvente é extremamente baixa.

As reações de cinética foram feitas em tampão fosfato de sódio 20 mM em uma faixa de pH de 6 a 8, utilizando 5 mM de MgCl_2 e aproximadamente 0,5 $\mu\text{g}/\text{mL}$ de NahK, e 0,3 $\mu\text{g}/\text{mL}$ do complexo NahK/NahL. A reação foi iniciada pela adição de diferentes quantidades de substrato (1,6 a 48 μM para NahK, e 2 a 60 μM para NahL) num volume final de reação de 1 mL. Para cada reação foi gerado um gráfico da absorbância em função do tempo e determinada a inclinação inicial da curva, a qual corresponde à velocidade inicial da reação. Utilizando a velocidade inicial da reação para diversas concentrações de substrato foi possível gerar um gráfico de Michaelis-Menten (velocidade inicial (v) *versus* concentração de substrato [S]) para estimar os parâmetros cinéticos das enzimas.

3.17. Ensaios de cristalização

Os ensaios de cristalização da proteína NahK e do complexo NahK/NahL, na presença ou não de ligantes, foram realizados manualmente segundo a técnica de difusão de vapor em gota suspensa em placas de 24 poços utilizando soluções do Kit Crystal Screen I e II (Hampton Research), num total de 96 condições. Sobre uma lamínula silanizada foram depositadas duas gotas de amostra proteica a aproximadamente 10 mg/mL, sendo uma de 1 μL e outra de 2 μL . Posteriormente, 1 μL da solução da condição de cristalização do poço foi adicionado e misturado a cada gota de proteína. A lamínula contendo as gotas foi virada e colocada sobre o poço com a condição de cristalização correspondente, de forma que as gotas ficassem suspensas sobre o poço. O fechamento hermético do poço foi feito utilizando

graxa de silicone. As placas de cristalização foram incubadas a 18 °C por várias semanas. O monitoramento dos resultados em cada condição de cristalização ocorreu quase que diariamente. Condições promissoras, nas quais houve nucleações, formação de agulhas ou crescimento de pequenos cristais, foram amplificadas manualmente alterando a concentração de agentes de precipitação e a faixa de pH utilizada de forma a prover uma maior gama de condições para os testes de cristalização.

Ensaio automatizados de cristalização para o complexo NahK/NahL também foram realizados utilizando o ROBOLAB – Laboratório Automatizado de Cristalização de Macromoléculas do Laboratório Nacional de Biociências – LNBio, em Campinas-SP. Para isso, a amostra já na concentração adequada para cristalização foi encaminhada para o ROBOLAB onde foram realizados ensaios de cristalização segundo a técnica de difusão de vapor em gota sentada, depositada em placas de 96 poços. Foi realizada também a captura automática de imagens das gotas com 0, 1, 2, 3, 5, 8, 13, 21, 34 e 55 dias após a montagem das placas com sensores no espectro visível e ultravioleta. As imagens puderam ser acessadas remotamente. Utilizando o ROBOLAB, mais de 500 condições de cristalização foram testadas utilizando-se apenas 0,5 µL de solução de proteína por condição. Condições promissoras foram selecionadas e amplificadas manualmente utilizando-se a técnica de difusão de vapor em gota suspensa em placas de 24 poços.

3.18. Difração de Raios-X

Os cristais obtidos de NahK e do complexo NahK/NahL foram transportados até o LNLS para serem submetidos à XRD nas linhas D03B-MX1 e W01B-MX2. Para a coleta de dados, os cristais foram capturados da gota de cristalização através de um loop de nylon e incubados por alguns segundos em uma gota de 1 µL de solução crioprotetora formada pela mistura da solução da condição de cristalização respectiva e etileno-glicol na concentração final de 10% v/v. Posteriormente, o cristal foi montado na cabeça de um goniômetro e centralizado no local onde passa o feixe de raios-X. Os cristais, a partir desse momento, ficaram sob um fluxo de nitrogênio constante, sendo resfriados à -173 °C. Inicialmente, para cada coleta, o detector foi colocado a uma distância de 100 mm do cristal e a qualidade de difração foi

testada recolhendo imagens a 0, 45 e 90°. Essas imagens foram analisadas no programa iMosflm [82], que localiza as reflexões do padrão de XRD e fornece estimativas prévias sobre parâmetros de célula unitária, grupo espacial do cristal e mosaicidade. Além disso, o programa também apresenta estratégias para a coleta de dados, sugerindo ângulos preferenciais para o início e o término para que a coleta seja mais eficiente. Para a aquisição de cada imagem do conjunto de dados, o cristal foi submetido ao feixe de raios-X durante alguns segundos, sendo rotacionado de 0,1 a 1° ao longo de cada exposição.

Quando necessário, foi realizada a técnica de *soaking* que consiste em mergulhar o cristal por um tempo determinado numa gota de 1 µL contendo a solução da condição de cristalização na presença de algum ligante, seja esse algum íon ou uma pequena molécula, e posteriormente, o cristal foi mergulhado na solução de crioproteção antes da montagem no goniômetro. A estrutura química dos substratos naturais de NahK e NahL, assim como os ligantes selecionados para *soaking* são exibidos na Figura 12.

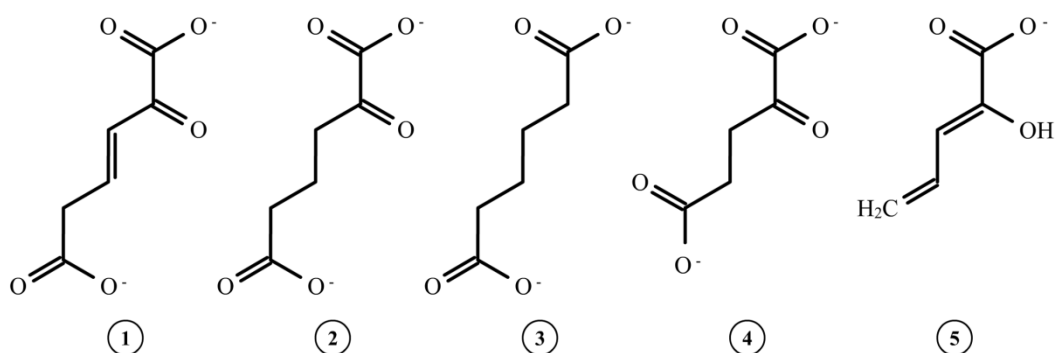


Figura 12. Substrato de NahK (1) e de NahL (5); e algumas moléculas análogas selecionadas como ligantes para NahK.

(1) 4-oxalocrotonato; (2) 2-oxoadipato; (3) adipato; (4) α-cetoglutarato; (5) 2-hidroxi-2,4-pentadienoato.

3.19. Construção, refinamento e análise dos modelos cristalográficos

Os conjuntos de dados de difração foram processados no programa HKL2000 [83], o qual gera um arquivo com extensão '.sca' para cada conjunto de dados, contendo parâmetros como os índices hkl com suas respectivas intensidades e erros associados aos pontos de difração, grupo espacial, tamanho e ângulos da célula unitária. Esses arquivos

foram processados posteriormente pelo programa Scalepack2mtz presente no pacote CCP4 [84], o qual gera um arquivo com extensão '.mtz' para cada conjunto de dados, que possui os módulos dos fatores de estrutura $|F(hkl)|$ para cada reflexão hkl. Os arquivos '.mtz' foram então utilizados pelo software Phaser MR [85] (pacote CCP4) para obtenção das fases por substituição molecular (MR, *molecular replacement*). A molécula utilizada como modelo para MR para os dados de NahK foi um monômero de HpcG (PDB ID: 2EB4 [63]), e para os dados do complexo NahK/NahL foi uma molécula do modelo de NahK já refinado.

Detectores utilizados atualmente em linhas de difração de raios-X medem as intensidades dos pontos de difração, mas toda informação a respeito das fases de cada onda é perdida. Entretanto, para construir um mapa de densidade eletrônica, e, portanto, construir o modelo atômico da proteína estudada, a informação das fases é necessária. Para solucionar esse problema, a técnica MR faz uso da estrutura de uma proteína homóloga já resolvida para estipular e utilizar as fases iniciais juntamente com as intensidades experimentais no cálculo do mapa de densidade eletrônica. Com sorte, essas fases serão acuradas o suficiente para gerar um primeiro mapa de densidade eletrônica, a partir do qual o modelo atômico da proteína que se deseja resolver possa ser construído e refinado.

Após a etapa de MR, o modelo preliminar, ainda contendo o monômero de HpcG (ou de NahK, no caso do complexo), foi reconstruído utilizando a sequência nativa de resíduos de aminoácidos de NahK ou NahL, e refinado parcialmente utilizando o programa Buccaneer [86] (pacote CCP4). Após a obtenção desse primeiro modelo tridimensional parcialmente construído, o arquivo contendo as coordenadas atômicas '.pdb' do modelo e o arquivo contendo os módulos dos fatores de estrutura experimentais '.mtz' foram processados utilizando ciclos de refinamento automático e manual utilizando a ferramenta *phenix.refine* do pacote PHENIX [87] e o programa Coot [88], respectivamente. Os ciclos de refinamento consistem, basicamente, em analisar se as posições atômicas estão de acordo com os mapas de densidade eletrônica gerados, e corrigi-las quando necessário. Nesses ciclos são impostas distâncias interatômicas consistentes com ligações covalentes e os ângulos entre essas ligações também são considerados a fim de gerar um modelo coerente.

Após o refinamento final, as estruturas dos 6 modelos de NahK em sua forma nativa e complexada com ligantes, e do modelo do complexo NahK/NahL foram analisadas através do programa Chimera [80] e depositadas no PDB.

4. RESULTADOS E DISCUSSÃO

A parte inicial deste trabalho, no que diz respeito ao processo de clonagem e parte da padronização da produção e cristalização da enzima NahK na ausência de NahL, foi realizada durante o final do mestrado e apresentado sob a forma de uma dissertação [70]. Como o trabalho que levou ao desenvolvimento desta tese de doutorado é uma continuação direta dos estudos durante o mestrado, alguns dos resultados já apresentados na dissertação serão reapresentados nesta tese para uma melhor compreensão da apresentação e discussão sobre NahK.

4.1. Produção de NahK recombinante na ausência de NahL

O gene *nahK* de *P. putida* G7 com aproximadamente 810 pares de base (pb), incluindo as sequências para enzimas de restrição contidas nos iniciadores, foi previamente amplificado por PCR utilizando os iniciadores nahK-NheI-F e nahK-HindIII-R (Figura 13), ligado ao vetor de clonagem pGEM-T e inserido em células de *E. coli* DH5 α [70]. A amplificação do gene ocorreu de forma mais eficiente utilizando a T_m de 58 °C e na concentração de DNA molde de aproximadamente 3 ng/ μ L. Alguns clones obtidos foram analisados por PCR de colônias e um clone foi selecionado para sequenciamento e subclonagem do gene no vetor pET28a-TEV.

Vários clones contendo o gene *nahK* subclonado em pET28a-TEV foram identificados através de PCR de colônias. Com base no sequenciamento do clone selecionado para expressão da proteína recombinante foi confirmada a correta inserção do gene em fase no sítio múltiplo de clonagem do vetor. Foram realizados diversos sequenciamentos, mas mesmo utilizando o iniciador reverso nahK-HindIII-R e iniciador reverso T7, não foi possível concluir o sequenciamento da região gênica que comporta os últimos 350 nucleotídeos correspondentes a parte mais próxima ao C-terminal da proteína. O sequenciamento cobriu cerca de 60% do gene *nahK* a partir da região correspondente ao N-terminal, e até essa porção não ocorreram mutações na sequência nucleotídica.

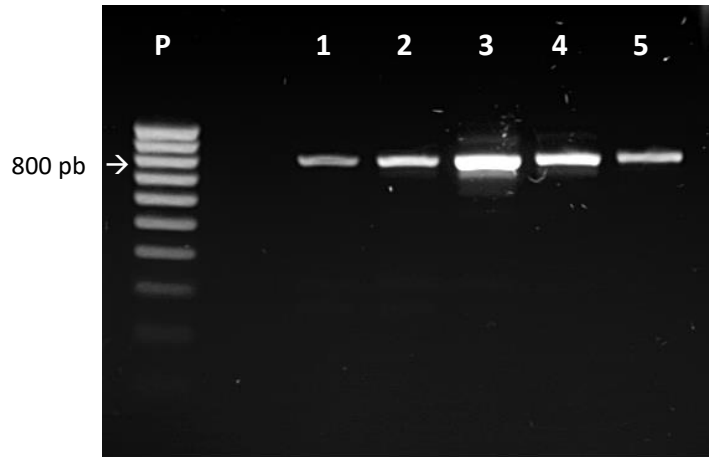


Figura 13. Amplificação do gene *nahK* de *P. putida* G7 para clonagem no vetor pET28a-TEV. **P** – padrão 100 pb EZ Load Molecular Ruler (Bio-Rad); **1 a 5** – amplificações do gene *nahK* contendo, respectivamente, DNA plasmidial NAH7 numa concentração de 30, 15, 3, 0,6 e 0,3 ng/μL. Imagem retirada de Guimarães, 2011 [70].

Após subclonagem do gene *nahK* no vetor pET28a-TEV, a proteína recombinante 6xHis-NahK (aproximadamente 31 kDa) foi expressa em células de *E. coli* BL21(DE3) por 5 h a 30 °C [70]; como pode ser observado pelo aparecimento de uma banda de superexpressão após o término da indução presente na altura entre os marcadores de massa molecular de 25 e 35 kDa (Figura 14, 0h e 5h). 6xHis-NahK foi expressa de forma parcialmente solúvel e, apesar da fração insolúvel aparentemente conter maior quantidade de proteína do que a fração solúvel (Figura 14, S e I), um litro de cultura tipicamente rendeu de 40 a 60 mg totais de proteína recombinante após primeira etapa de purificação por IMAC, não sendo necessária a adoção de estratégias para aumentar a solubilidade dessa proteína.

De acordo com a fase de leitura do vetor pET28a-TEV e os iniciadores utilizados para a amplificação do gene *nahK*, a seguinte sequência de resíduos de aminoácidos foi expressa:

MGHHHHHENLYFQ/GHMASMNRTLREQVLALAEHIENAELNVHDIGKVTNDFPEMTFADAYDVQWEIRRRKEARGN
 KIVGLKMGLTSWAKMAQMGVETPIYGFADYFSVPDGGVVDCSKLIHPKIEAIEISVVTKAPLHGPGCHLGDVIAAIDYVIPTV
 EVIDSRyenFKFDLISVVADNASSTRFITGGRMASLEEVDLRTLGVVMEKNGEVVELGAGAAVLGHPLSSVAMLANLLAERG
 EHIPAGTFIMTGGITAAPVAPGDNITVRYQGLGSVSARFI

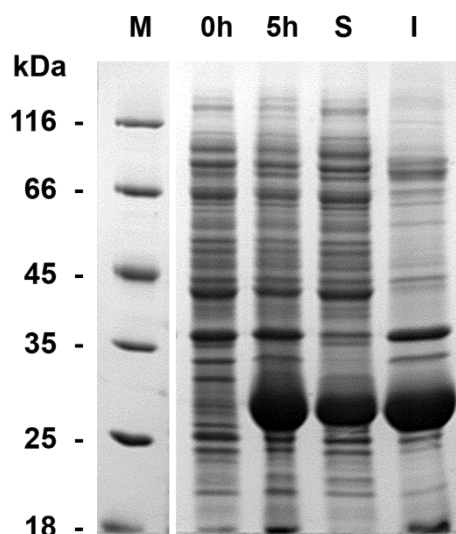


Figura 14. Análise por SDS-PAGE da expressão de NahK separadamente de NahL.

M – Padrão de massa molecular (Fermentas); **0h** – extrato proteico antes da indução por IPTG; **5h** – extrato proteico após 5h de indução por IPTG 0,5 mM; **S** e **I** – frações solúvel e insolúvel, respectivamente, após lise e centrifugação das células. Imagem adaptada de Guimarães, 2011 [70].

NahK expressa através desta maneira possui uma cauda de 6 histidinas consecutivas em sua porção N-terminal (sequência destacada em azul). Essa cauda permitiu a purificação parcial de NahK por IMAC em coluna HisTrap HP que contém íons de níquel imobilizados em sua resina. O perfil cromatográfico típico da primeira etapa de purificação de 6xHis-NahK pode ser observado na Figura 15 (a). Amostras foram coletadas durante esta etapa de purificação e analisadas por SDS-PAGE (Figura 15 b). Durante a injeção na coluna da fração solúvel da expressão de NahK, proteínas que não se ligaram à resina passaram diretamente pela coluna e são observadas como um grande platô na curva de absorção (linha contínua) entre os volumes de eluição 0 e 60 mL (fração I). Após a lavagem para a remoção do restante das proteínas que não se ligaram à resina, que ocorreu entre os volumes de eluição 60 e 90 mL (fração L), há um gradiente crescente com tampão de eluição que contém imidazol em maior concentração (linha tracejada). O imidazol compete com a cauda de histidinas pelas interações com os íons metálicos imobilizados na resina, e uma grande concentração de imidazol é capaz de desligar as proteínas que ficaram retidas na coluna por este motivo. O pico observado entre os volumes de eluição 90 e 120 mL é em grande parte constituído por 6xHis-NahK, mas também possui contaminantes (fração P). Esses contaminantes são em grande parte decorrentes de proteínas nativas de *E. coli* que possuem afinidade ao níquel, principalmente por conter grande quantidade de histidinas em suas superfícies. Para

diminuir a quantidade de contaminantes é adicionada uma pequena quantidade de imidazol no tampão de ligação. Inicialmente, durante o mestrado, a concentração de 30 mM de imidazol foi estabelecida como recomendação do fabricante da coluna. Entretanto, como ainda havia uma grande quantidade de contaminantes na eluição da proteína (Figura 15 b, fração P), no doutorado, a concentração de imidazol foi elevada para 45 mM, e isso permitiu um grande aumento no nível de pureza de 6xHis-NahK com apenas uma etapa de purificação por IMAC (Figura 15 c, fração P).

O vetor pET28a-TEV também adiciona na porção N-terminal de NahK uma sequência de resíduos de aminoácidos reconhecida e clivada pela protease TEV (Glu-Asn-Leu-Tyr-Phe-Gln-Gly, sequência destacada em verde na sequência expressa). A clivagem ocorre entre os resíduos de glutamina e glicina (posição destacada pelo símbolo /). Essa estratégia para clivagem da cauda de histidinas foi utilizada devido à facilidade de expressão e purificação da protease TEV, que é produzida em grandes quantidades no próprio Laboratório de Biologia Estrutural – ICB / UFMG. Além disso, essa protease é ativa numa variedade de condições e temperaturas, além de reconhecer uma sequência bastante específica de resíduos, diminuindo a possibilidade de clivagem em sítios inespecíficos e deixando apenas um resíduo da sequência reconhecida ainda ligado na porção N-terminal da proteína alvo após a clivagem [89]. Sendo assim, utilizou-se a protease TEV para a clivagem da cauda de histidinas a 25 ou 4 °C por 4 ou 16 h, respectivamente.

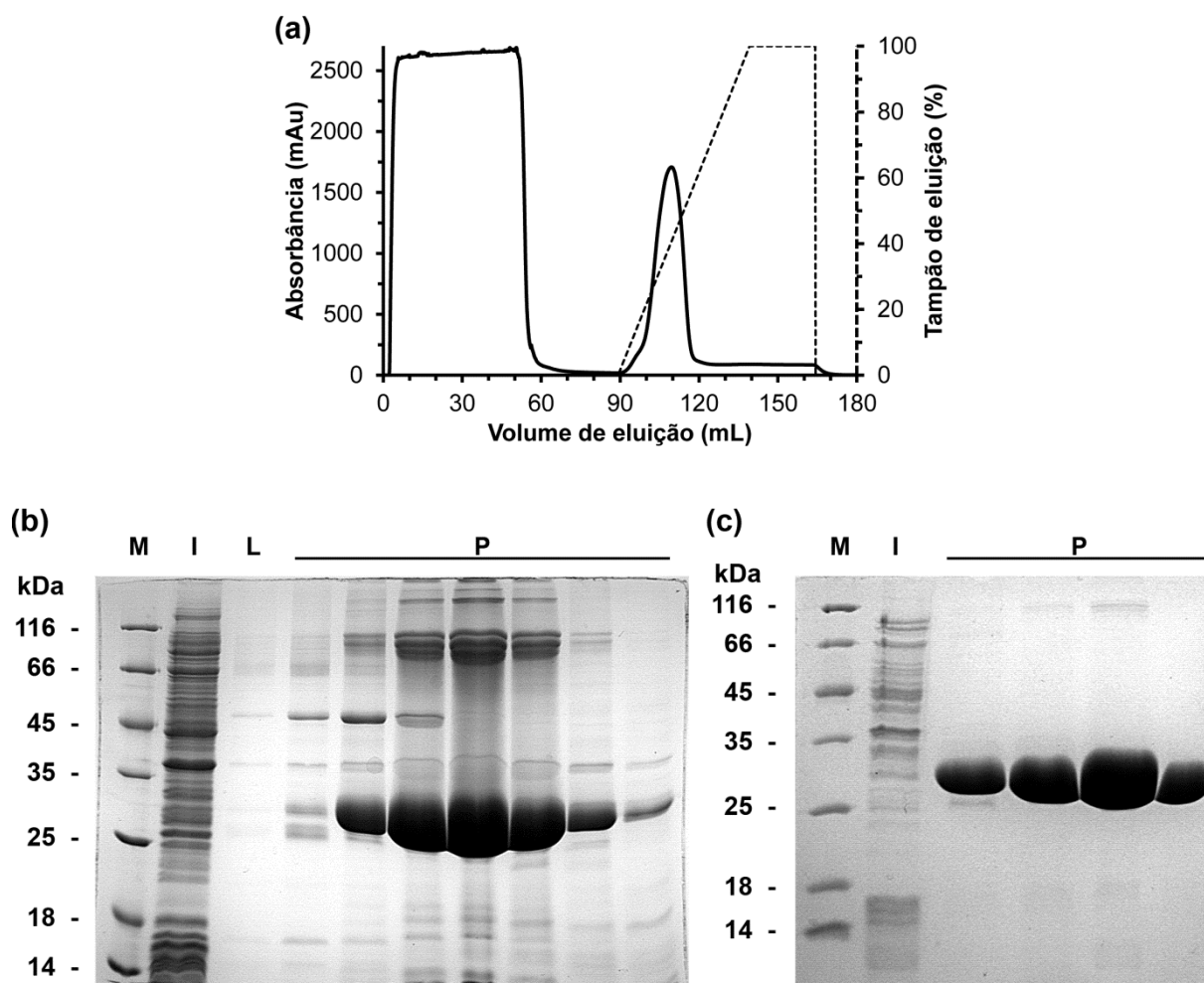


Figura 15. Cromatograma e análise por SDS-PAGE da primeira etapa de purificação de 6xHis-NahK por IMAC. **(a)** Cromatograma da primeira etapa de purificação de 6xHis-NahK por IMAC. A absorção foi medida a 280 nm. **(b)** SDS-PAGE da purificação de 6xHis-NahK utilizando 30 mM de imidazol no *tampão de ligação A*. **M** – Padrão de massa molecular (Fermentas); **I** – Fração de proteínas que não se ligaram à coluna durante a injeção da amostra; **L** – fração obtida durante a lavagem da coluna; **P** – frações constituintes do pico de eluição. **(c)** SDS-PAGE da purificação de 6xHis-NahK utilizando 45 mM de imidazol no *tampão de ligação A*. **M** – Padrão de massa molecular (Fermentas); **I** – Fração de proteínas que não se ligaram à coluna durante a injeção da amostra; **P** – frações constituintes do pico de eluição.

Após a reação de clivagem, a amostra de proteína contendo NahK recombinante foi aplicada novamente em coluna HisTrap HP para um segundo passo de purificação por IMAC, mas nesta etapa foi utilizando apenas 30 mM de imidazol no *tampão de ligação*. A estratégia foi usar uma quantidade menor de imidazol para que os contaminantes que foram capazes de se ligar na resina durante a primeira etapa de IMAC, utilizando 45 mM de imidazol no *tampão de ligação*, continuassem sendo retidos na coluna enquanto que NahK, já sem cauda de histidinas, passasse diretamente pela coluna, aumentando assim o nível de pureza da proteína recombinante. Como a protease TEV também possui uma cauda de histidinas, ela

também fica retida na coluna, juntamente com contaminantes e algumas moléculas de 6xHis-NahK que ocasionalmente não foram clivadas.

O perfil cromatográfico típico da segunda etapa de purificação de NahK já sem cauda de histidinas pode ser observado na Figura 16 (a). Amostras foram coletadas durante esta etapa de purificação e analisadas por SDS-PAGE (Figura 16 b). Durante a injeção da amostra, entre os volumes de eluição 0 e 15 mL quase não houve eluição de proteínas da coluna, como pode ser observado pela baixa absorbância a 280 nm nesses volumes (linha contínua). Após o início da lavagem da coluna com tampão de ligação contendo 30 mM de imidazol, observou-se a eluição de um grande pico de proteína constituído principalmente por NahK sem cauda de histidinas (volume de eluição 20 mL, fração P). Isso ocorreu porque não havia imidazol no tampão da reação de clivagem com a protease TEV, e quando essa amostra foi injetada na coluna, a própria proteína NahK ficou retida na mesma. Isso evidencia uma pequena afinidade dessa proteína ao níquel, já que NahK possui 7 resíduos de histidina, e como foi observado posteriormente em sua estrutura tridimensional, todos esses resíduos estão na superfície da macromolécula e são passíveis de interação com a resina. Com o término da injeção da amostra e o início da lavagem da coluna com tampão de ligação, a pequena quantidade de imidazol utilizada já foi suficiente para desligar NahK da resina. Após eluição de NahK e lavagem da coluna (volumes de eluição de 15 a 50 mL), há um gradiente crescente com tampão de eluição contendo imidazol em maior concentração (linha tracejada). Frações coletadas durante esse gradiente apresentaram várias bandas de proteína de massas moleculares próximas de 28 e 30 kDa (frações E). Essas bandas podem se tratar de moléculas de 6xHis-NahK que não foram clivadas (aproximadamente 31 kDa), e uma mistura de NahK clivada (aproximadamente 29 kDa) e protease TEV (aproximadamente 28 kDa), assim como contaminantes.

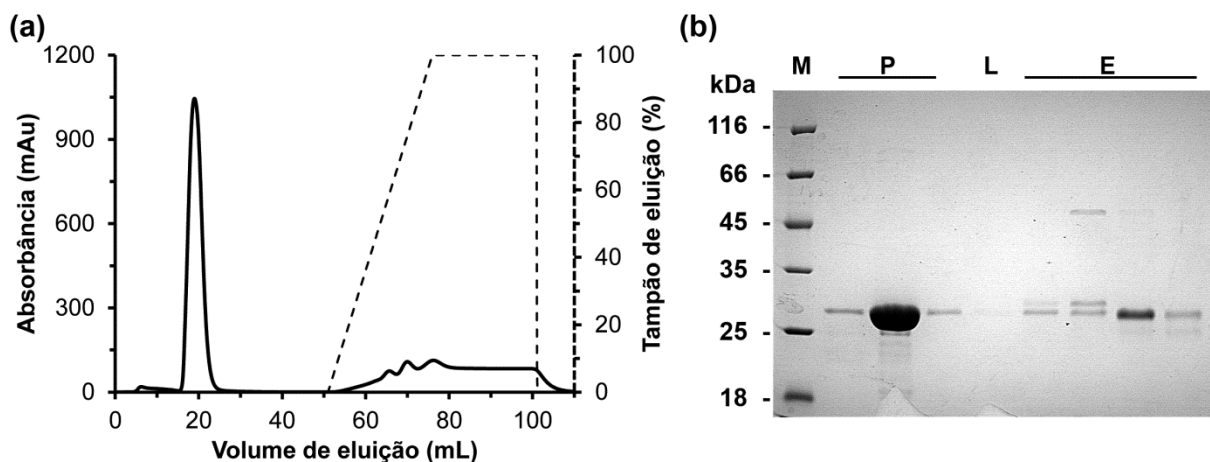


Figura 16. Cromatograma e análise por SDS-PAGE da segunda etapa de purificação de NahK por IMAC. **(a)** Cromatograma da segunda etapa de purificação de NahK por IMAC. A absorção foi medida a 280 nm. **(b)** Purificação de NahK utilizando 30 mM de imidazol no *tampão de ligação B*. **M** – Padrão de massa molecular (Fermentas); **P** – fração obtida durante a eluição do pico principal constituída principalmente por NahK sem cauda de histidinas; **L** – fração obtida durante a lavagem da coluna após eluição de NahK; **E** – frações constituintes dos pequenos picos de eluição obtidos durante o gradiente com imidazol. Imagem adaptada de Guimarães, 2011 [70].

Após a segunda etapa de purificação de NahK por IMAC a amostra já se encontrava com uma alta pureza e foi submetida aos ensaios de cristalização. Entretanto, uma etapa final de purificação por SEC foi realizada numa tentativa de caracterizar o estado oligomérico de NahK [70]. Essa técnica utiliza colunas contendo uma resina com diferentes tamanhos de poros, permitindo que moléculas maiores atravessem a resina com maior velocidade, enquanto moléculas menores vão passando através dos poros, atravessando a coluna em menor velocidade. Sendo assim, esta estratégia de purificação separa as proteínas por tamanho. A Figura 17 (a) exibe o perfil cromatográfico típico de NahK obtido durante esta etapa de purificação. As frações constituintes do pico principal eluído em aproximadamente 17 mL foram unidas e analisadas por SDS-PAGE (Figura 17 b). O resultado final foi uma amostra de alta pureza de NahK recombinante, como pode ser observado nas canaletas P e [P], nas quais foram aplicadas concentrações diferentes da amostra final de NahK.

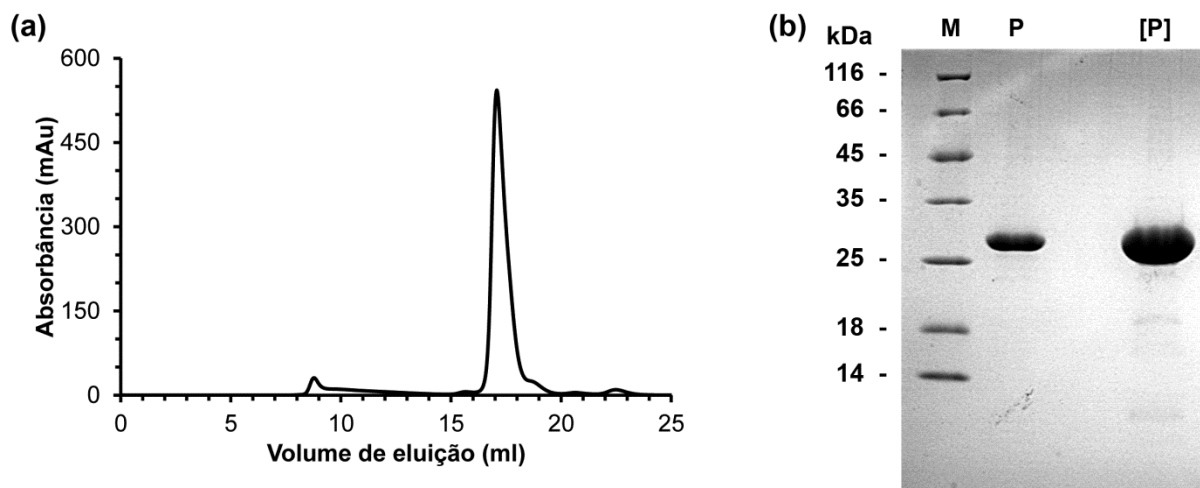


Figura 17. Cromatograma e análise por SDS-PAGE da terceira etapa de purificação de NahK por SEC. **(a)** Cromatograma da terceira etapa de purificação de NahK por SEC. A absorção foi medida a 280 nm. **(b)** SDS-PAGE da purificação de NahK por SEC. **M** – Padrão de massa molecular (Fermentas); **P** e **[P]** – fração obtida durante a eluição do pico principal entre 17 e 18 mL constituída principalmente por NahK recombinante.

NahK produzida desta maneira possui 5 resíduos de aminoácidos não nativos em sua porção N-terminal: Gly-His-Met-Ala-Ser. Apesar de a protease TEV manter apenas o resíduo de glicina após a clivagem, os resíduos de histidina, metionina, alanina e serina foram adicionados pelo vetor pET28a-TEV devido ao uso da enzima de restrição *NheI* no desenho do iniciador direto para a amplificação do gene (resíduos destacados em amarelo na sequência de aminoácidos listada anteriormente). A proteína se manteve solúvel e estável após ser concentrada até 12 mg/mL e armazenada a 4 °C. De acordo com a calibração da coluna de gel filtração e o volume de eluição do pico de NahK recombinante obtido durante a purificação por SEC, foi estimado que essa proteína comporta-se como uma partícula proteica de aproximadamente 25 a 30 kDa, sugerindo que NahK expressa na ausência de NahL é um monômero em solução.

Em 1989, Harayama e colaboradores reportaram que XylI é fisicamente associada a XylJ *in vivo* [64]. XylI e NahK compartilham 94% de identidade entre suas sequências de resíduos de aminoácidos (Figura 18); enquanto XylI e NahL compartilham 90% de identidade (Figura 48). De fato, XylI e NahK descarboxilam o mesmo substrato (4-oxalocrotonato), assim como XylJ e NahL hidratam o produto dessa descarboxilação. Harayama, no mesmo trabalho, expressou XylI em células de *E. coli* BL21(DE3) na ausência de XylJ, e realizou uma purificação parcial dessa enzima através de cromatografia de troca iônica e precipitação por sulfato de amônio. No entanto, essa 4-OD expressa e parcialmente purificada na ausência de

sua respectiva VPH apresentou baixa atividade de descarboxilação quando comparada com a atividade da mesma enzima expressa e purificada enquanto complexada com XylJ. Além disso, as frações proteicas coletadas após uma etapa de purificação por cromatografia de interação hidrofóbica não apresentaram atividade enzimática. Stanley e colaboradores também tentaram purificar e estabilizar Xyll na ausência de XylJ adicionando íons Mg^{2+} e glicerol nos tampões de purificação, sem sucesso [65]. Foi reportado que Xyll parcialmente purificada rapidamente perdia sua atividade enzimática quando concentrada ou armazenada a 4 °C.

Neste trabalho foi utilizada estratégia diferente para expressão e purificação de uma 4-OD na ausência da respectiva VPH de sua via, ambas provenientes do plasmídeo NAH7. Assim, uma forma estável de NahK foi expressa e purificada na ausência de NahL. Como será apresentado adiante nos testes cinéticos, NahK expressa dessa forma manteve-se ativa mesmo após concentrada ou armazenada a 4 °C. Apesar de NahK e Xyll serem enzimas homólogas e de mesma função, 17 de seus 264 resíduos de aminoácidos são diferentes (Figura 18). A maioria das substituições dentre esses resíduos são equivalentes em tamanho, carga elétrica e/ou hidrofobicidade (T6N, N23Q, V24A, F34Y, V45I, L127I, I134V, F168Y, S176N, E179D e I264V). Apesar disso, surpreendentemente, com a resolução da estrutura de NahK, é possível notar que 15 dos 17 resíduos diferentes entre essas enzimas estão localizados na superfície dessas proteínas e suas cadeias laterais interagem com o solvente. Portanto, apesar da alta identidade e similaridade de sequências entre NahK e Xyll, essa leve diferença poderia explicar uma maior solubilidade e estabilidade em NahK. Outra possível explicação poderia ser o protocolo de purificação mais ameno utilizando a cauda de histidinas e sem precipitação por sulfato de amônio.

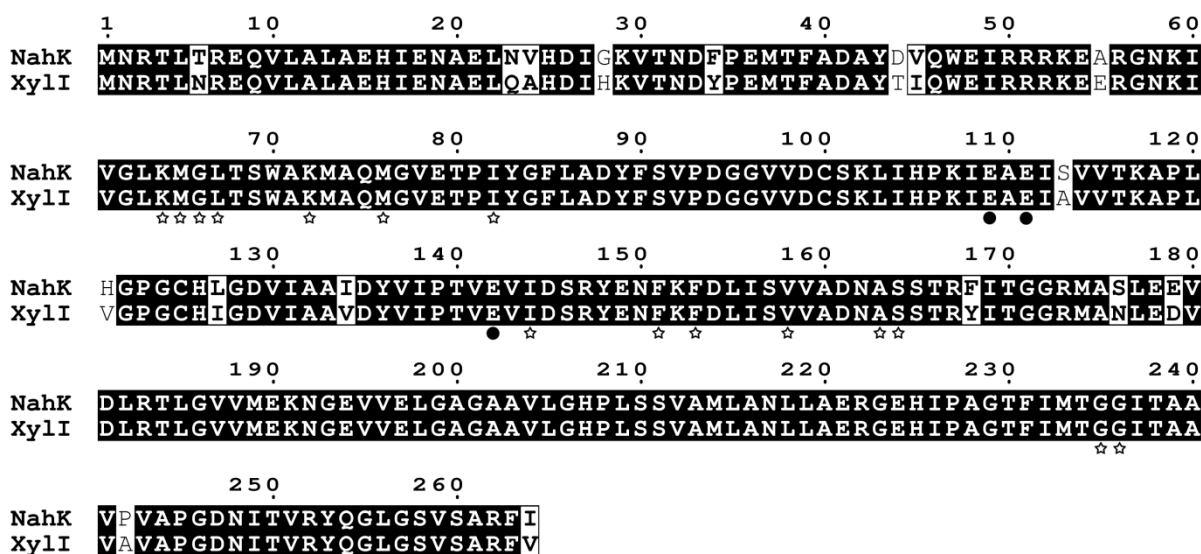


Figura 18. Alinhamento das sequências de resíduos de aminoácidos de NahK e XylI.

Os resíduos que coordenam o íon metálico no sítio ativo estão destacados por círculos, enquanto que os outros resíduos que também compõe o sítio ativo estão destacados por estrelas.

4.2. Estado oligomérico e estruturas secundárias de NahK

Através da técnica de DLS foi possível averiguar o estado de oligomerização de NahK em solução em sua forma isolada de NahL. O experimento de DLS mede as flutuações da intensidade da luz espalhada por partículas em movimento browniano em solução. A intensidade de movimentação das partículas é inversamente proporcional ao tamanho das mesmas. Logo, uma molécula proteica monomérica e pequena move-se mais rapidamente do que um dímero ou uma molécula maior. Além disso, proteínas de mesma massa molecular, mas de diferentes formas oligoméricas são caracterizadas por diferentes coeficientes de difusão. Os coeficientes de difusão das partículas são obtidos a partir das análises das flutuações instantâneas da luz espalhada por essas partículas. O raio hidrodinâmico das partículas é obtido a partir da equação de Stokes-Einstein, na qual se utiliza o coeficiente de difusão da partícula em questão [90,91].

Essa técnica é útil para calcular a polidispersidade de uma amostra já que é sensível aos tamanhos de partículas em solução. Também é possível sugerir o nível de agregação da proteína de acordo com sua massa molecular e conseqüentemente, julgar se a amostra tem

melhor capacidade de formar cristais. Amostras monodispersas são de boa qualidade para ensaios de cristalização [92]. Se a amostra estiver polidispersa, serão observados picos largos ou vários picos indicando partículas de diferentes tamanhos, o que pode sugerir uma amostra com impurezas ou contendo vários estados de agregação da mesma proteína. Quando a amostra está monodispersa, observa-se um único pico indicando que a solução apresenta partículas mais semelhantes, indicando alta pureza.

A Figura 19, obtida através de ensaio de DLS para NahK após duas etapas de purificação por IMAC, apresenta a distribuição de tamanhos de partícula pela porcentagem de massa (linha contínua) e pela intensidade de espalhamento da luz (linha tracejada).

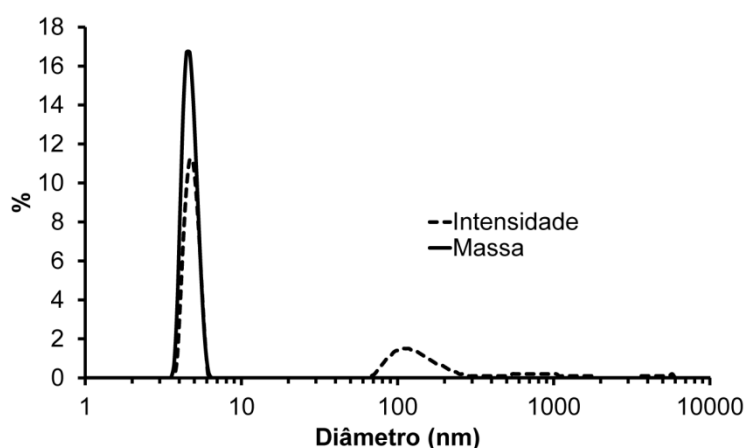


Figura 19. Distribuição de tamanhos de partícula de NahK recombinante obtido por DLS.

Apesar de a curva de intensidade assumir a presença de uma família de partículas com um diâmetro de aproximadamente 100 nm, a curva de porcentagem de massa indica que a amostra de NahK analisada era composta quase exclusivamente por uma população de partículas com cerca de 5 nm de diâmetro e aproximadamente 10% de polidispersidade, indicando que a amostra estava monodispersa e com uma alta pureza. Utilizando um modelo de proteína globular para estimativa de massa molecular, calculou-se que uma partícula proteica com tal diâmetro possui aproximadamente 26 kDa, indicando que NahK recombinante, que possui uma massa teórica de aproximadamente 29 kDa, se comporta como monômero em solução na ausência de NahL. Esse dado confirma o resultado obtido através de SEC, que indicou, pelo volume de retenção do pico de NahK, que a proteína comportava-se como uma partícula proteica de aproximadamente 25 a 30 kDa.

Uma busca no PDB [52] utilizando a sequência de resíduos de aminoácidos de NahK identificou as proteínas HpcG (2-oxo-hept-3-eno-1,7-dioato hidratase, PDB ID: 2EB4 [63]) e MhpD (ácido 2-hidroxipentadienoico hidratase, PDB ID: 2WQT [62]) como sendo as duas proteínas com estrutura tridimensional conhecida mais similares em sequência de aminoácidos à NahK. Essas enzimas compartilham com NahK, respectivamente, 41 e 34% de identidade entre suas sequências de aminoácidos. Foi reportado que MhpD (monômeros de aproximadamente 29 kDa) forma pentâmeros em solução, e que esses pentâmeros também foram visualizados na estrutura cristalográfica. Além disso, a unidade biológica de HpcG (monômeros de aproximadamente 30 kDa) é um decâmero formado por dois pentâmeros similares aos pentâmeros de MhpD.

Devido à similaridade de NahK com essas enzimas, e como sabe-se que NahK pode formar um complexo com NahL, assim como reportado para XylI e XylJ [64], podia-se esperar que NahK formasse estruturas quaternárias, talvez pentâmeros, assim como HpcG e MhpD. Entretanto, NahK comporta-se como monômero em solução, e nenhuma estrutura em pentâmero foi observada no empacotamento cristalino das moléculas de NahK (como poderá ser observado adiante). Sendo assim, podemos assumir que NahK por si só não tem as propriedades necessárias para formar estruturas quaternárias em solução. Além disso, a presença de NahL seria essencial para a formação de uma estrutura quaternária maior que dímeros de NahK e NahL.

A estrutura secundária de NahK e sua estabilidade térmica foram caracterizadas através de CD. Essa técnica é outro excelente método para o estudo de proteínas, sendo possível avaliar rapidamente a presença de estruturas secundárias, enovelamento e até mesmo interação de proteínas com ligantes. As proteínas são formadas por moléculas quirais, e as diferentes estruturas secundárias nas proteínas absorvem diferentemente a luz circularmente polarizada à direita e à esquerda, sendo o espectro de CD definido pela diferença dessa absorção [93]. O espectro de CD de NahK em solução a 20 °C (Figura 20 a) exibiu bandas negativas a 220 e 210 nm, e uma banda positiva a 195 nm. Esta forma característica de um espectro de CD indica que a proteína possui uma mistura de estruturas secundárias em hélices- α e fitas- β . A deconvolução dos dados através do programa CDNN mostra uma estimativa de aproximadamente 37% de hélices- α , 31% de fitas- β e 31% de estrutura randômica a 20 °C (Figura 20 b).

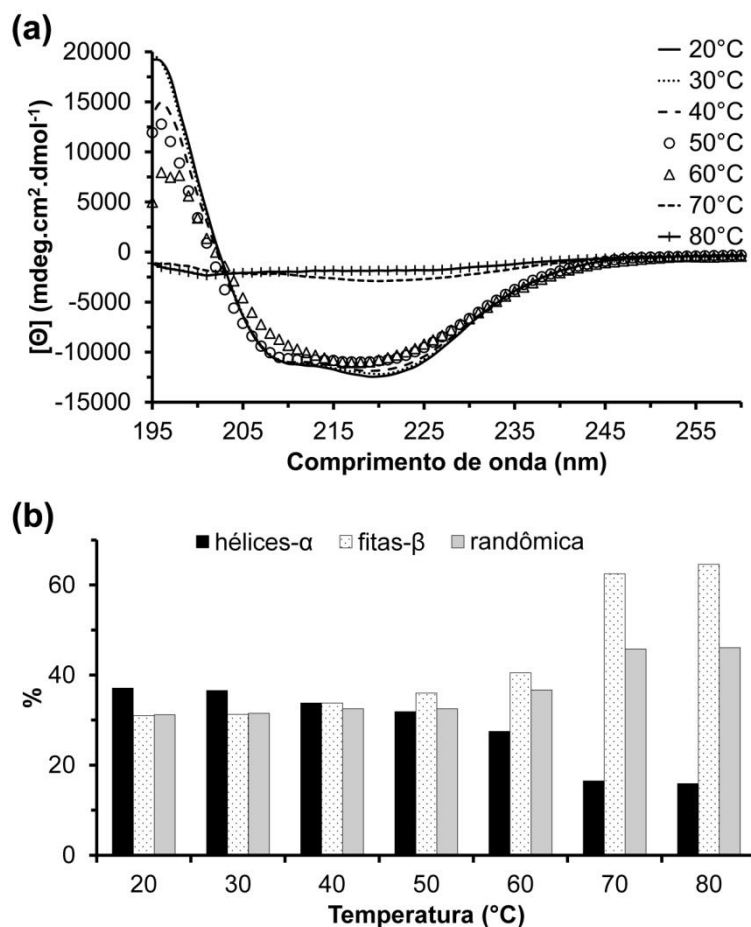


Figura 20. Dicroísmo circular de NahK na ausência de NahL.

(a) Espectro de dicroísmo circular de NahK em diferentes temperaturas. **(b)** Porcentagem de estruturas secundárias de acordo com a temperatura obtidas pela deconvolução dos dados.

De acordo com o servidor STRIDE [94], o resultado obtido para NahK é semelhante ao conteúdo de estruturas secundárias da estrutura cristalográfica de HpcG, que possui 34% de hélices- α e 32% de fitas- β ; e de MhpD, que possui 32% de hélices- α e 31% de fitas- β . Percebe-se então que a amostra de NahK recombinante expressa na ausência de NahL é constituída pela enzima em um enovelamento característico e semelhante ao de HpcG e MhpD, evidenciando o provável enovelamento similar dessas proteínas.

É possível observar na Figura 20 (b) que, com o aumento da temperatura, há uma perda gradual de estruturas em hélices- α acompanhada por um ganho de estruturas em fitas- β e estruturas randômicas indicadas pelo aparecimento de uma banda negativa a 217 nm ao longo da transição entre 20 e 60 °C. Uma mudança abrupta no formato da curva, indicando um maior grau de desenovelamento, ocorreu entre as temperaturas 60 e 70 °C, sugerindo que a T_m (temperatura de *melting*) para esta proteína esteja em torno de 65 °C.

4.3. NahK possui atividade enzimática na ausência de NahL

NahK é uma descarboxilase que remove o segundo grupo carboxilato do substrato 4-oxalocrotonato, presente no carbono 6 da molécula, transformando-o em uma molécula de CO₂. O primeiro grupo carboxilato é denominado como parte do grupo α -cetoácido do substrato; enquanto que o grupo que é descarboxilado é denominado como β -cetoácido vinílico. A reação catalisada por NahK é dependente de magnésio ou manganês como cofator [64]. Como o substrato e o produto de NahK possuem absorvâncias características a 236 e 265 nm, respectivamente, foi possível determinar os parâmetros cinéticos desta enzima através da análise da absorvância da luz ultravioleta pela amostra durante os ensaios de atividade enzimática.

Para testar a atividade de NahK foi preciso, previamente, gerar o substrato 4-oxalocrotonato a partir de 2HM. Como pode ser observado na Figura 21, que representa o andamento da reação da enzima 4-OT, a molécula 2HM, que tem alta absorvância a 295 nm, é isomerizada a 4-oxalocrotonato, que tem alta absorvância a 236 nm. A concentração final teórica do produto de 4-OT seria a mesma do substrato utilizado se a eficiência da reação fosse de 100%, entretanto, o equilíbrio da reação de 4-OT faz com que cerca de 80% do substrato 2HM seja convertido em produto; o substrato de NahK. Os cerca de 20% restantes de 2HM podem ser observados como um pequeno pico de absorvância residual a 296 nm após a reação chegar ao equilíbrio.

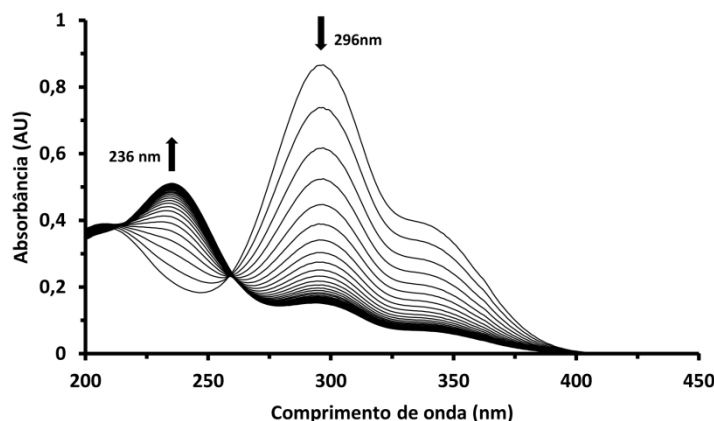


Figura 21. Curvas de absorvância da reação da enzima 4-OT ao converter 2HM em 4-oxalocrotonato.

Por conseguinte, a Figura 22 exibe o comportamento geral da curva de absorção ao longo do tempo durante a reação da enzima NahK na ausência de NahL. Ao descarboxilar 4-oxalocrotonato (que possui alta absorbância a 236 nm), são formados CO_2 e 2-hidroxi-2,4-pentadienoato (que possui alta absorbância a 265 nm). Dessa forma, com a diminuição da concentração do substrato há um decréscimo na absorção a 236 nm com o consequente acúmulo do produto, gerando o pico de absorção a 265 nm.

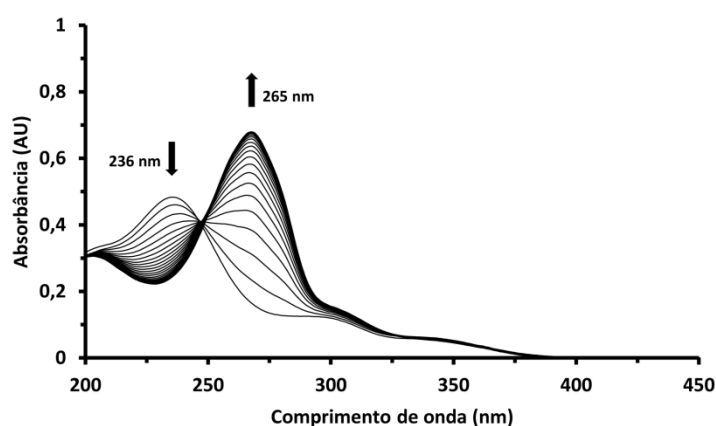


Figura 22. Curvas de absorbância da reação da enzima NahK ao descarboxilar 4-oxalocrotonato na ausência de NahL.

Inicialmente a atividade de NahK (a $0,3 \mu\text{g/mL}$) foi avaliada em pH 7,4 a temperatura ambiente em três diferentes concentrações de cloreto de magnésio. A Figura 23 mostra o gráfico obtido a partir dos valores da velocidade inicial da reação em função da concentração de substrato para diferentes concentrações de magnésio. A absorbância foi medida a 265 nm de acordo com o acúmulo do produto da reação. A Tabela 8 exibe os dados de cinética obtidos desse experimento.

As curvas tracejadas referentes aos gráficos de cinética foram obtidas a partir do ajuste dos pontos à equação de Michaelis Mentem.

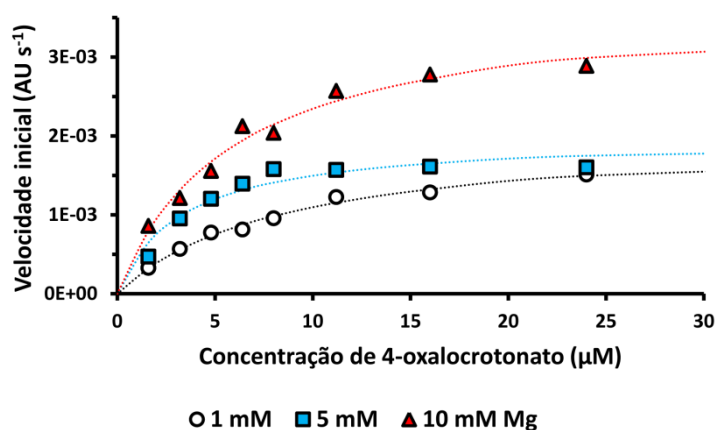


Figura 23. Velocidades iniciais da reação de NahK na ausência de NahL em diferentes concentrações de substrato e de Mg^{2+} .

Tabela 8. Parâmetros de cinética para NahK na ausência de NahL utilizando-se diferentes concentrações de magnésio.

[Mg^{2+}] (mM)	V_{max} ($AU s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1} s^{-1}$)
1	$0,0020 \pm 0,0001$	$8,42 \pm 0,96$	$19,75 \pm 0,99$	$2,3 \times 10^6$
5	$0,0020 \pm 0,0001$	$3,23 \pm 0,79$	$19,75 \pm 0,99$	$6,1 \times 10^6$
10	$0,0037 \pm 0,0002$	$5,84 \pm 0,89$	$36,54 \pm 1,98$	$6,3 \times 10^6$

O parâmetro V_{max} é a velocidade máxima estimada que a reação teoricamente pode alcançar (para determinada concentração de enzima) considerando que a enzima se encontra totalmente saturada com seu substrato. O parâmetro K_M é o valor de concentração de substrato necessário para que a velocidade da reação seja a metade da máxima; esse valor simboliza, para reações enzimáticas simples, a afinidade entre a enzima e um substrato, e são inversamente proporcionais (um alto valor de K_M significa uma baixa afinidade entre enzima e substrato). O parâmetro k_{cat} corresponde ao valor de quantas unidades de substrato cada sítio catalítico consegue processar por segundo, no máximo. O parâmetro k_{cat}/K_M é uma relação que expressa a eficiência da enzima e é um importante valor para se comparar a eficiência de enzimas diferentes, ou da mesma enzima em condições ou substratos diferentes.

Com base nos valores de cinética exibidos na Tabela 8, pode-se observar que a concentração de Mg^{2+} na solução é um fator limitante para a eficiência da enzima, sendo que com 1 mM o valor de k_{cat}/K_M foi o mais baixo dentre os três obtidos. Utilizando Mg^{2+} a 10 mM obteve-se um V_{max} mais elevado, entretanto o valor de K_M também aumentou na

mesma proporção, fazendo com que o valor de k_{cat}/K_M fosse praticamente o mesmo para o ensaio com Mg^{2+} a 5 mM. Sendo assim, adotou-se a concentração de Mg^{2+} 5 mM para realização dos demais ensaios de cinética.

A atividade da enzima NahK (a 0,5 $\mu\text{g}/\text{mL}$) na ausência de NahL foi testada em valores de pH variando entre 6 e 8 em temperatura ambiente utilizando MgCl_2 a 5 mM (Figura 24). Os parâmetros de cinética podem ser visualizados na Tabela 9.

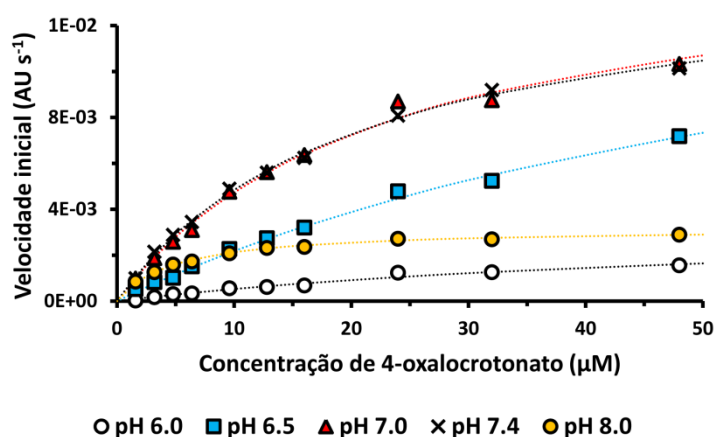


Figura 24. Influência do pH na cinética de NahK na ausência de NahL.

Tabela 9. Parâmetros de cinética para NahK na ausência de NahL utilizando-se diferentes valores de pHs em tampão fosfato de sódio.

pH	V_{max} (AU s^{-1})	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
6,0	$0,0033 \pm 0,0006$	$51,65 \pm 14,16$	$22,48 \pm 4,09$	$4,4 \times 10^5$
6,5	$0,0167 \pm 0,0015$	$64,71 \pm 8,53$	$108,35 \pm 9,73$	$1,7 \times 10^6$
7,0	$0,0153 \pm 0,0008$	$22,24 \pm 2,41$	$104,21 \pm 5,45$	$4,7 \times 10^6$
7,4	$0,0145 \pm 0,0004$	$19,85 \pm 1,13$	$98,77 \pm 2,72$	$5,0 \times 10^6$
8,0	$0,0032 \pm 0,0001$	$4,88 \pm 0,27$	$21,80 \pm 0,34$	$4,5 \times 10^6$

Os valores de K_M diminuem à medida que o pH se eleva. Esse comportamento provavelmente é um resultado da ionização dos dois grupos carboxila do substrato e uma ligação mais forte com a enzima. Também se pode levar em conta a titulação dos resíduos que coordenam o átomo de metal, criando uma ligação mais forte com o magnésio, estabilizando-o mais eficientemente no sítio ativo da enzima. Os valores de k_{cat} e k_{cat}/K_M foram mais elevados na faixa entre os pHs 6,5 e 7,4.

Quando se compara esses resultados com aqueles obtidos por Stanley (2000) [65], que trabalhou com o complexo XylI/E106Q_XylI (em que a enzima homóloga à NahK está ativa e a enzima homóloga à NahL está inativa pela mutação do resíduo 106 de glutamato para glutamina), verifica-se os valores de $K_M = 13 \pm 1 \mu\text{M}$, $k_{\text{cat}} = 67 \pm 2 \text{ s}^{-1}$ e $k_{\text{cat}}/K_M = 5,1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$, podendo-se notar que seus parâmetros de cinética são semelhantes aos da enzima XylI na forma de complexo XylI/E106Q_XylI.

Sendo assim, constata-se que a atividade enzimática de NahK ocorre de forma independente da presença ou interação de NahL.

4.4. Determinação da estrutura cristalográfica de NahK

Duas diferentes amostras de NahK foram utilizadas nos testes de cristalização: a primeira amostra foi obtida após um único passo de IMAC e ainda possuía a cauda de histidinas; a segunda amostra foi obtida após a etapa de clivagem da cauda de histidinas seguida por um segundo passo de purificação por IMAC e SEC. Após alguns dias foram observadas separação de fase ou precipitação na maioria das condições de cristalização. Nucleações foram observadas após uma semana nas gotas de NahK sem cauda de histidinas na condição número 22 do kit Crystal Screen 1 (CS1-22: acetato de sódio trihidratado 0,2 M, tampão TRIS-HCl 0,1 M pH 8,5 e PEG 4000 30% m/v). Cristais no formato de agulhas adequados para difração (Figura 25 a) possuindo até 1 mm de comprimento em seu maior eixo foram crescidos em algumas semanas em condições derivadas da CS1-22 em uma faixa de pH de 8,5 a 10,5 e concentração de precipitante PEG 4000 de 26 a 30%. Cristais no formato de paralelepípedos (Figura 25 b) possuindo um máximo de 0,2 mm de comprimento foram observados após quatro meses nas gotas de 6xHis-NahK na condição número 16 do kit Crystal Screen 1 (CS1-16: tampão HEPES-Na 0,1 M pH 7,5 e sulfato de lítio monohidratado 1,5 M). Cristais da condição CS1-16 e derivação da condição CS1-22 foram utilizados para XRD. A Figura 25 (c) exibe um padrão de difração típico de proteína, neste caso de um cristal de NahK recombinante.

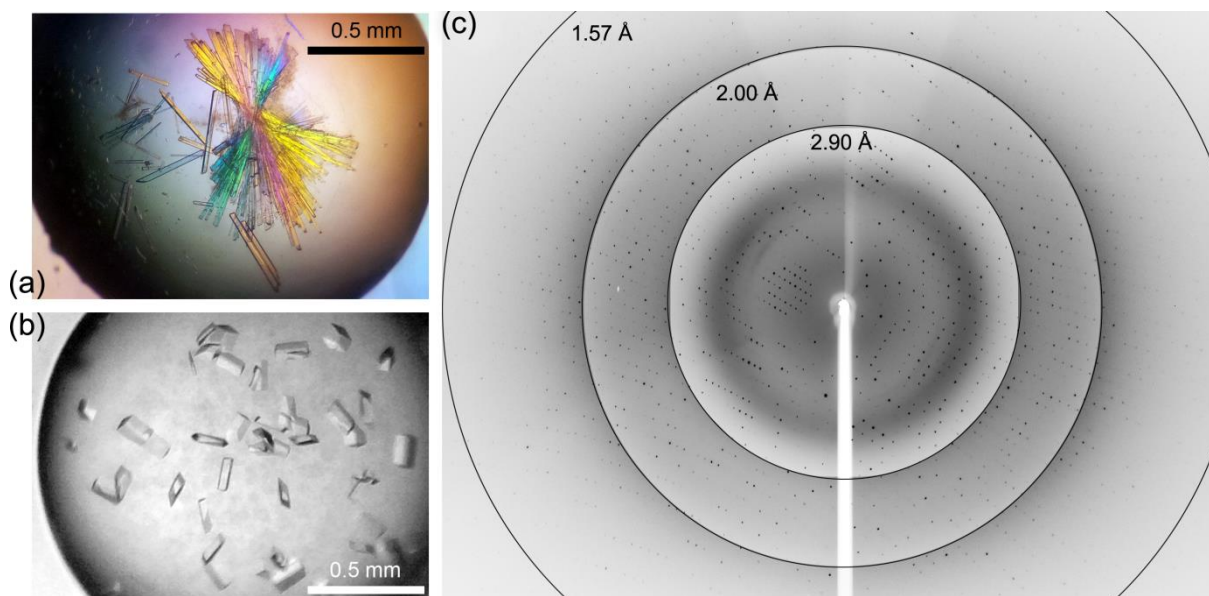


Figura 25. Cristais de NahK e padrão de difração de raios-X.

(a) Cristais em formato de agulha crescidos em condição derivada de CS1-22. **(b)** Cristais em formato retangular crescidos na condição CS1-16. **(c)** Padrão de difração típico de cristais de NahK recombinante. A distância do detector ao cristal foi estabelecida de forma que fosse coletado um conjunto de dados de até 1,57 Å de resolução

Três conjuntos de dados foram coletados com cristais provindos da condição CS1-16 (Tabela 10), que consistem em um cristal de apo-enzima (conjunto 1: modelo *Apo*, 1,74 Å de resolução), um cristal que foi previamente imerso em uma solução contendo cloreto de magnésio (conjunto 2, modelo *Mg*, 1,90 Å de resolução), e um cristal que foi previamente imerso em duas soluções, uma contendo íons de magnésio e outra contendo solução de ácido α -cetoglutárico (conjunto 3: modelo *AKG*, 1,94 Å de resolução). Todos os cristais crescidos nessa condição pertencem ao grupo espacial $P2_12_12_1$ com parâmetros de célula unitária de aproximadamente $a = 43$, $b = 45$, $c = 125$ Å, $\alpha = \beta = \gamma = 90^\circ$ (Figura 26 a). O valor do Coeficiente de Matthews para esses conjuntos sugeriu a presença de apenas uma molécula de proteína por unidade assimétrica com um conteúdo de solvente de aproximadamente 38,11%. A Substituição Molecular para esses três conjuntos de dados foi feita com sucesso utilizando apenas um monômero da proteína HpcG [63], que é a proteína com maior identidade de sequência de aminoácidos com NahK (41%) com estrutura tridimensional já resolvida e depositada no RCSB Protein Data Bank [52]. Ao todo existem 4 cópias da proteína na célula unitária, sendo uma cadeia por unidade assimétrica.

Outros três conjuntos de dados foram coletados com cristais de NahK crescidos a partir de condições melhoradas provenientes da condição CS1-22 (Tabela 11). O primeiro conjunto foi coletado com um cristal nativo (conjunto 4: modelo *ACT*, 1,78 Å de resolução), o segundo conjunto foi coletado com um cristal que, após crescido, foi adicionado em sua gota de cristalização solução contendo cloreto de magnésio e ácido adípico (conjunto 5: modelo *AA*, 1,72 Å de resolução), e por fim o último conjunto de dados foi coletado com um cristal que, também após crescido, foi adicionado em sua gota de cristalização solução contendo íons magnésio e ácido 2-oxoadípico (conjunto 6: modelo *2OA*, 1,57 Å de resolução). Os cristais crescidos a partir dessa condição de cristalização pertencem ao grupo espacial $P2_1$ e possuem parâmetros de célula unitária de aproximadamente $a = 79$, $b = 44$, $c = 81$ Å, $\alpha = 90^\circ$, $\beta = 120^\circ$, $\gamma = 90^\circ$ (Figura 26 b). O valor do Coeficiente de Matthews sugeriu a presença de duas moléculas de proteína por unidade assimétrica com um conteúdo de solvente de aproximadamente 42,3%. A Substituição Molecular utilizando um monômero de HpcG gerou duas soluções de posição da molécula para cada conjunto; ambas as posições foram utilizadas para construir os modelos. Ao todo existem 4 cópias da proteína na célula unitária, entretanto são duas cadeias por unidade assimétrica.

Todos os seis modelos tridimensionais, depois de inicialmente construídos utilizando o programa Buccaneer [86], foram manualmente corrigidos utilizando o programa Coot [88] e refinados no phenix.refine do pacote PHENIX [87]. Os seis modelos finais de NahK foram refinados com uma boa geometria, apresentando um rmsd médio de 0,007 Å para o comprimento das ligações covalentes e um rmsd de 1,14° para seus ângulos. Foi possível manter em média 255 resíduos dos 263 totais de cada cadeia proteica (~96,5%) em regiões favoráveis do gráfico de Ramachandran, e uma média de 8 resíduos por cadeia em regiões permitidas (~3%). Não foi possível determinar a posição do primeiro resíduo (Met1) nos modelos. A cauda de histidinas também não foi visualizada na densidade eletrônica dos modelos provenientes dos cristais ortorrômbicos. Todos os seis modelos estão depositados no PDB [52] sob os códigos 5D2F (*Apo*), 5D2G (*Mg*), 5D2H (*AKG*), 5D2I (*ACT*), 5D2J (*AA*) e 5D2K (*2OA*).

As nove estruturas de NahK (somatório das 6 cadeias 'A' dos seis modelos e 3 cadeias 'B' dos modelos provenientes dos cristais monoclínicos) são bastante semelhantes, apresentando um rmsd de apenas 0,763 Å entre a posição dos C^α de seus 263 resíduos; diminuindo para um rmsd de até 0,497 Å entre a posição de 245 C^α ao desconsiderar os resíduos de número 68 a 81 e 150 a 153, localizados na abertura do sítio ativo e que possuem as maiores diferenças de posição entre as nove cadeias.

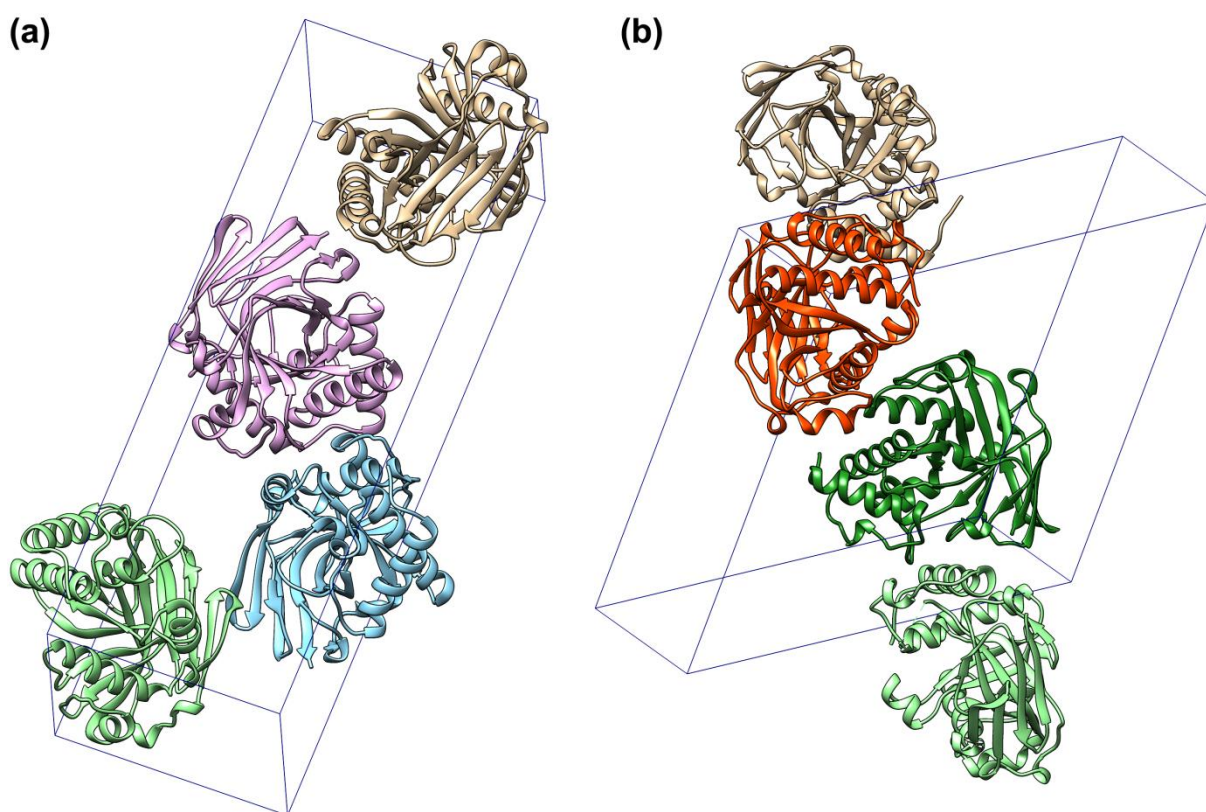


Figura 26. Representação das moléculas de NahK empacotadas nas células unitárias dos cristais provindos das condições CS1-16 (a) e CS1-22 (b).

Tabela 10. Sumário das estatísticas dos conjuntos de dados obtidos a partir dos cristais de NahK crescidos na condição CS1-16.

Os valores entre parênteses são para as zonas de maior resolução.

Conjunto de dados	1 (modelo Apo)	2 (modelo Mg)	3 (modelo AKG)
PDB ID	5D2F	5D2G	5D2H
Linha	D03B-MX1	D03B-MX1	D03B-MX1
Detector	MarCCD 165	MarCCD 165	MarCCD 165
Comprimento de onda (Å)	1,437	1,437	1,437
Distância detector (mm)	60,0	80,0	80,0
Rotação por imagem (°)	1,0	1,0	1,0
Rotação total (°)	180	99	90
Grupo espacial	$P2_12_12_1$ $a = 43,76$	$P2_12_12_1$ $a = 42,77$	$P2_12_12_1$ $a = 43,74$
Parâmetros de célula unitária (Å, °)	$b = 44,94$ $c = 125,14$ $\alpha = \beta = \gamma = 90$	$b = 45,04$ $c = 124,67$ $\alpha = \beta = \gamma = 90$	$b = 44,90$ $c = 125,32$ $\alpha = \beta = \gamma = 90$
Mosaicidade (°)	0,52	0,74	0,95
Resolução (Å)	31,35 – 1,74 (1,78 – 1,74)	40,45 – 1,90 (1,96 – 1,90)	42,27 – 1,94 (2,00 – 1,94)
No. de reflexões totais	180.312	66.421	58.992
No. de reflexões únicas	25.510	18.966	18.138
Redundância	7,1 (6,3)	3,5 (2,7)	3,3 (2,5)
Completeza (%)	97,2 (94,5)	95,7 (72,6)	94,8 (70,7)
[$I / \sigma(I)$]	20,24 (2,82)	10,20 (1,83)	9,87 (1,39)
R_{merge}^a (%)	0,080 (0,516)	0,108 (0,435)	0,104 (0,514)
Coefficiente de Matthews (Å ³ Da ⁻¹)	2,00	1,95	2,00
Conteúdo de Solvente (%)	38,6	37,1	38,6
R fator	0,168	0,153	0,163
R_{free}	0,211	0,204	0,231
No. de átomos de proteína	2.008	1.991	1.999
No. de moléculas de água	398	307	271
B fatores:			
átomos de proteína	16,4	15,7	21,7
geral	18,5	17,0	22,6
rmsd:			
ângulos das ligações	1,17	1,21	1,10
comprimento das ligações	0,007	0,008	0,007

^a $R_{\text{merge}} = \sum_{hkl} \sum_i [I_i(hkl) - \langle I(hkl) \rangle] / \sum_{hkl} \sum_i I_i(hkl)$, onde $I_i(hkl)$ é a intensidade medida da reflexão hkl e $\langle I(hkl) \rangle$ é sua média.

Tabela 11. Sumário das estatísticas dos conjuntos de dados obtidos a partir dos cristais de NahK crescidos em condição melhorada de CS1-16.

Os valores entre parênteses são para as zonas de maior resolução.

Conjunto de dados	4 (modelo ACT)	5 (modelo AA)	6 (modelo 2OA)
PDB ID	5D2I	5D2J	5D2K
Linha	D03B-MX1	W01B-MX2	W01B-MX2
Detector	MarCCD 165	MarMosaic 225	MarMosaic 225
Comprimento de onda (Å)	1,430	1,459	1,461
Distância detector (mm)	70,0	80,0	80,0
Rotação por imagem (°)	1,0	0,7	1,0
Rotação total (°)	300	200	280
Grupo espacial	$P2_1$	$P2_1$	$P2_1$
Parâmetros de célula unitária (Å, °)	$a = 77,49$	$a = 78,77$	$a = 79,33$
	$b = 43,75$	$b = 44,55$	$b = 44,41$
	$c = 79,50$	$c = 81,25$	$c = 81,62$
	$\alpha = \gamma = 90$	$\alpha = \gamma = 90$	$\alpha = \gamma = 90$
	$\beta = 118,2$	$\beta = 118,8$	$\beta = 118,7$
Mosaicidade (°)	0,43	1,03	0,42
Resolução (Å)	26,68 – 1,78 (1,82 – 1,78)	40,59 – 1,72 (1,76 – 1,72)	40,78 – 1,57 (1,60 – 1,57)
No. de reflexões totais	276.928	196.894	385.525
No. de reflexões únicas	45.433	51.584	69.868
Redundância	6,1 (5,9)	3,8 (2,9)	5,5 (4,2)
Completeza (%)	100,0 (100,0)	97,3 (80,6)	99,9 (99,8)
[$I / \sigma(I)$]	29,29 (2,98)	15,36 (3,01)	29,70 (7,97)
R_{merge}^a (%)	0,060 (0,491)	0,075 (0,350)	0,054 (0,177)
Coefficiente de Matthews (Å ³ Da ⁻¹)	2,05	2,16	2,18
Conteúdo de Solvente (%)	40,1	43,1	43,6
R fator	0,146	0,154	0,150
R_{free}	0,186	0,191	0,175
No. de átomos de proteína	3.968	3.996	4.065
No. de moléculas de água	608	525	647
B fatores:			
átomos de proteína	17,4	18,9	15,7
geral	18,9	20,2	17,3
rmsd:			
ângulos das ligações	1,18	1,06	1,15
comprimento das ligações	0,008	0,006	0,008

^a $R_{\text{merge}} = \sum_{hkl} \sum_i [I_i(hkl) - \langle I(hkl) \rangle] / \sum_{hkl} \sum_i I_i(hkl)$, onde $I_i(hkl)$ é a i intensidade medida da reflexão hkl e $\langle I(hkl) \rangle$ é sua média.

4.5. Estrutura geral de NahK

NahK apresenta três domínios principais. O domínio N-terminal é compreendido pelos 58 resíduos iniciais que se enovelam em três hélices conectadas por pequenas voltas, sendo duas hélices- α longas e uma hélice- 3_{10} curta (Figura 27 a, hélices A, B e C). O segundo e maior domínio é composto por um barril- β incompleto formado por uma folha- β bastante torcida feita por 8 fitas paralelas e antiparalelas interligadas por voltas e pequenas hélices (Figura 27 a, folha A). Há um sítio de ligação a metal formado por três resíduos de glutamato (Glu109 e Glu111 da fita 5, e Glu142 da fita 6) localizados no topo da estrutura em barril- β (Figura 27 b) e estão localizados no fundo do sítio ativo dessa enzima. O domínio C-terminal é constituído principalmente por uma estrutura em folha- β mais plana que contém 5 fitas (Figura 27 a, folha B). A cadeia polipeptídica passa cinco vezes entre as folhas A e B, evidenciando a íntima relação e interação entre o domínio do barril e a folha C-terminal.

Uma busca dentro do banco de dados de proteínas não redundantes no servidor BLAST [61] utilizando a sequência de resíduos de aminoácidos de NahK identificou essa proteína como um membro da superfamília FAH, assim como previamente reportado para Xyll [95]. Apesar de baixa identidade de sequências baseada nas estruturas de NahK e FAH (apenas 9%), o enovelamento dessas duas enzimas é similar. Entretanto somente os domínios formados pelo barril- β incompleto e pela folha- β C-terminal de NahK possuem semelhanças com os respectivos domínios de FAH (Figura 9 b); os domínios N-terminal de NahK e FAH são distintos.

O domínio C-terminal de FAH, composto pelo barril- β incompleto juntamente da segunda folha- β , é a principal característica estrutural compartilhada pelos membros da superfamília FAH, sendo que nem todas as enzimas dessa superfamília, que já possuem estruturas tridimensionais conhecidas, possuem o domínio N-terminal semelhantes.

A porcentagem de estruturas secundárias na estrutura depositada de NahK (5D2G), segundo o PDB, é de 28% de hélices- α e 28% de fitas- β . Esse resultado é próximo das proporções de estruturas secundárias obtidas através da deconvolução dos dados no experimento de dicroísmo circular (Figura 20 b).

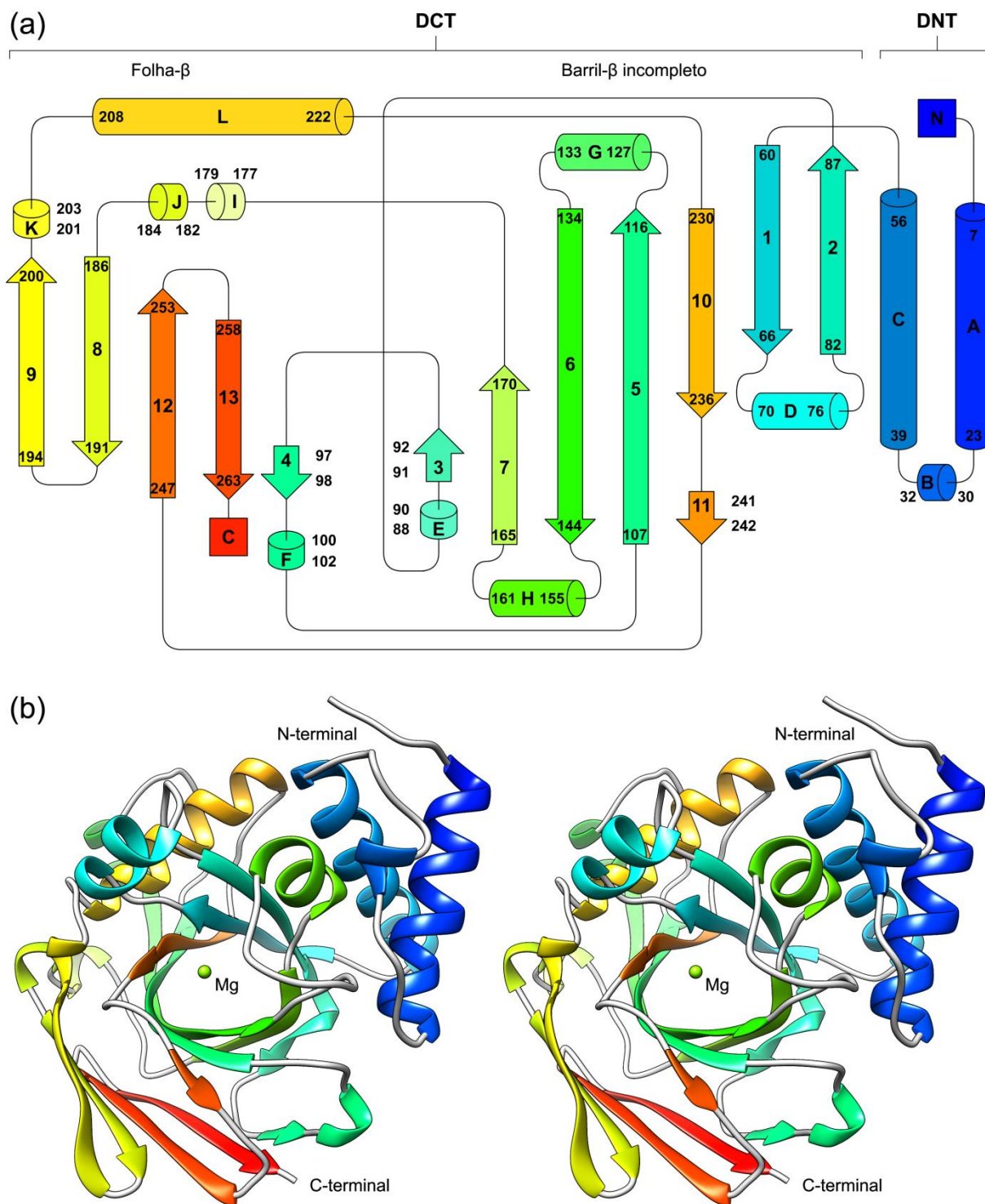


Figura 27. Representação das estruturas secundárias e terciária de NahK.

(a) Diagrama topológico de NahK: hélices são exibidas como tubos e nomeadas alfabeticamente; fitas- β são exibidas como setas e nomeadas numericamente. A transição de cores de azul a vermelho representa o progresso da cadeia de N a C-terminal. (b) Representação em estéreo (*cross-eyed*) da cadeia principal de NahK ligada ao magnésio, a representação de cores é a mesma de (a). O íon magnésio (esfera verde) ligado à proteína se encontra no topo da estrutura em barril- β . O diagrama topológico foi gerado com ajuda da plataforma on-line Pro-origami [96].

Uma busca através do servidor DALI [97] dentro do PDB [52] utilizando o arquivo contendo as coordenadas atômicas da estrutura de NahK identificou várias proteínas com enovelamento similar ao dessa enzima e que também pertencem à superfamília FAH (Tabela 12). Somente duas das proteínas listadas possuem similaridade estrutural com NahK ao longo de toda sua cadeia polipeptídica: HpcG e MhpD, que compartilham a posição de, em média, 260 dos 264 resíduos de NahK com rmsd médio de 1,4 Å. Um alinhamento de sequências e de estruturas de NahK, HpcG e MhpD pode ser observado na Figura 28. As outras proteínas da superfamília FAH compartilham em média apenas 184 resíduos em posições semelhantes dos 264 resíduos de NahK, com um rmsd variando de 2,4 a 3,0 Å.

Tabela 12. Proteínas com estrutura conhecida pertencentes à superfamília FAH e similaridades com NahK.

PDB ID	Função	Organismo	n° aa	pares C ^α	rmsd	%id
2EB4						
2EB5	hidratase (HpcG)	<i>Escherichia coli</i>	267	261	1,2	39%
2EB6						
1SV6	hidratase (MhpD)	<i>Escherichia coli</i>	270	259	1,6	31%
2WQT						
3QDF	isomerase	<i>Mycobacterium marinum</i>	268	181	2,6	15%
3RR6	função hipotética	<i>Mycobacterium abscessus</i>	265	181	2,6	15%
4DBF						
4DBH	descarboxilase (Cg1458)	<i>Corynebacterium glutamicum</i>	288	182	2,6	12%
1WZO	isomerase	<i>Thermus thermophilus</i>	246	183	2,5	19%
2DFU	isomerase	<i>Thermus thermophilus</i>	264	179	2,4	12%
4PFZ	isomerase	<i>Mycobacterium smegmatis</i>	259	183	2,8	15%
1GTT	descarboxilase / isomerase	<i>Escherichia coli</i>	429	185	2,7	14%
117O	(HpcE)					
1SAW	função hipotética	<i>Homo sapiens</i>	225	184	2,8	15%
1NR9	função hipotética	<i>Escherichia coli</i>	223	184	2,6	17%
3S52	função hipotética	<i>Yersinia pestis</i>	221	188	2,8	16%
3V77	função hipotética	<i>Oleispira antarctica</i>	224	188	2,9	18%
4MAQ	função hipotética	<i>Burkholderia cenocepacia</i>	240	188	3,3	16%
3R6O	isomerase	<i>Mycobacterium abscessus</i>	329	179	2,5	17%
2Q18						
2Q19						
2Q1A	desidratase (KdaD)	<i>Sulfolobus solfataricus</i>	293	178	2,7	15%
2Q1C						
2Q1D						
3BQB						
1NKQ	função hipotética	<i>Saccharomyces cerevisiae</i>	259	184	2,8	15%
3LZK	função hipotética	<i>Sinorhizobium meliloti</i>	359	188	2,9	14%
4QKU	função hipotética	<i>Burkholderia cenocepacia</i>	441	188	2,9	14%
1HYO						
1QCN						
1QCO	hidrolase (FAH)	<i>Mus musculus</i>	423	190	3,0	9%
1QQJ						
2HZY						

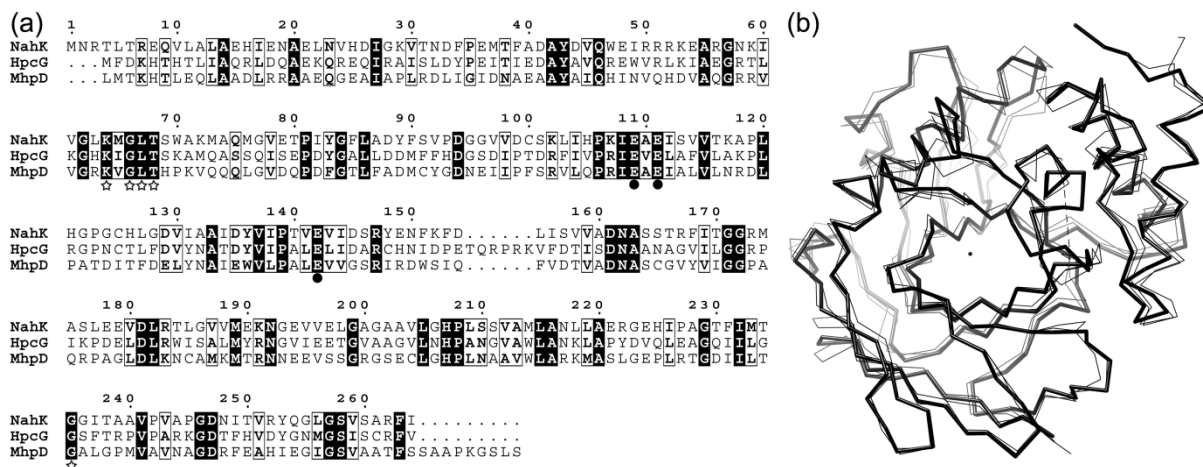


Figura 28: Alinhamento de seqüências de aminoácidos (a) e das estruturas terciárias (b) de NahK, HpcG e MhpD.

(a) Os resíduos de glutamato que coordenam o metal estão indicados por círculos, enquanto que demais resíduos conservados do sítio ativo estão indicados por estrelas. (b) Superposição do esqueleto C^α de NahK (fio mais grosso) com HpcG e MhpD (fios mais finos); o ponto no centro da estrutura indica a posição do magnésio.

Uma análise mais profunda nas estruturas dessas proteínas mais distintas revela que esses 184 resíduos fazem parte dos elementos estruturais que compõe o domínio do barril- β incompleto e a folha- β C-terminal. De fato, a maior diferença entre NahK, HpcG e MhpD com relação aos demais membros da superfamília FAH ocorre no domínio N-terminal dessas proteínas, e não no domínio do barril.

A enzima FAH de camundongo (PDB ID: 1QCN, [66]) juntamente com uma hidrolase putativa de *Burkholderia cenocepacia* (PDB ID: 4QKU, não publicado) possuem um grande domínio N-terminal formado por cerca de 120 e 135 resíduos de aminoácidos, respectivamente, que se enovelam num sanduíche α/β composto por 5 fitas- β antiparalelas intercaladas por diversas hélices- α , e são representados pela estrutura em azul exibida na Figura 29 (a) gerada a partir da estrutura de FAH. A enzima desidratase KdaD de *Sulfolobus solfataricus* (PDB ID: 2Q18, [68]) possui uma versão mais simples desse domínio α/β N-terminal, sendo este composto por 5 fitas- β antiparalelas e apenas 2 hélices- α ao longo de 73 resíduos, representados pela estrutura em laranja na Figura 29 (b).

As enzimas 2-hidroxihepta-2,4-dieno-1,7-dioato isomerase de *Mycobacterium marinum* (PDB ID: 3QDF, [98]) e de *Thermus thermophilus* (PDB ID: 2DFU e 1WZO, não publicado), uma proteína de função hipotética de *Mycobacterium abscessus* (PDB ID: 3RR6, [98]), a oxaloacetato descarboxilase Cg1458 de *Corynebacterium glutamicum* (PDB ID: 4DBF,

[69]), juntamente com a 5-carboximetil-2-hidroxiuconato isomerase de *Mycobacterium smegmatis* (PDB ID: 4PFZ, não publicado), possuem uma variante ainda mais simples do domínio N-terminal de FAH, sendo que seus respectivos domínio N-terminal são compostos quase exclusivamente por apenas um dobramento de 5 fitas- β antiparalelas compostas por cerca de 42 a 59 resíduos de aminoácidos, representados pela estrutura em verde na Figura 29 (c) gerada a partir da estrutura 3QDF.

Já as estruturas de 2-hidroxihepta-2,4-dieno-1,7-dioato isomerase de *Mycobacterium abscessus* (PDB ID: 3R6O, [98]), juntamente com uma proteína de função hipotética de *Sinorhizobium meliloti* (PDB ID: 3LZK, não publicado) possuem apenas 3 fitas- β antiparalelas seguidas por uma hélice- α em seus respectivos domínio N-terminal, constituído por cerca de 79 resíduos de aminoácidos, representados pela estrutura rosa na Figura 29 (d) gerada a partir da estrutura 3LZK.

A estrutura de uma proteína de função hipotética relacionada ao metabolismo de aminoácidos aromáticos em humanos (PDB ID: 1SAW, [67]), juntamente com proteínas de função hipotética de *Escherichia coli* (PDB ID: 1NR9, não publicado), *Yersinia pestis* (PDB ID: 3S52, não publicado), *Oleispira antarctica* (PDB ID: 3V77, [99]), *Burkholderia cenocepacia* (PDB ID: 4MAQ, não publicado) e de *Saccharomyces cerevisiae* (PDB ID: 1NKQ, não publicado) possuem domínios N-terminal extremamente curtos sendo constituídos por pequenos elementos de estrutura secundária em hélice- α , fita- β ou voltas ao longo de cerca de apenas 16 resíduos de aminoácidos, representados pela estrutura roxa na Figura 29 (e) gerada a partir da estrutura 3S52.

Outra característica comum dentre a maioria dos membros da superfamília FAH é a formação de dímeros, sendo que a primeira subunidade de cada dímero está representada na Figura 29 pelas estruturas de cadeia principal em cinza juntamente com o domínio N-terminal em cores diversas, sendo que a segunda subunidade está representada em cinza juntamente com o domínio N-terminal com a cor correspondente. Nessa estrutura quaternária a porção que compreende as fitas opostas da folha- β do barril- β incompleto da primeira subunidade está posicionada juntamente com a região correspondente do barril- β incompleto da segunda subunidade, de forma que os dois barris interagem e se estabilizam mutuamente.

A enzima bifuncional HpcE de *E. coli* (PDB ID: 1GTT, [60]) não possui qualquer domínio N-terminal, sendo constituída apenas por dois domínios barril- β incompleto com suas respectivas folhas- β 'C-terminais'. Nessa enzima, o primeiro domínio barril está unido ao segundo por uma volta com cerca de 16 resíduos de aminoácidos representados pela estrutura azul na Figura 29 (f). Assim, apesar de possuir apenas uma única cadeia polipeptídica, HpcE mantém a estrutura da montagem em dímero característica das proteínas da superfamília FAH ao interagir seus dois domínios barril da mesma maneira que as demais proteínas diméricas dessa superfamília.

NahK de *Pseudomonas putida* G7, juntamente com HpcG (PDB ID: 2EB4, [63]) e MhpD (PDB ID: 2WQT, [62]) de *Escherichia coli* possuem um domínio N-terminal formado por 3 hélices ao longo de em média 56 resíduos de aminoácidos, representados pela estrutura em vermelho na Figura 29 (g) gerada pela estrutura 2EB4. Ao contrário das demais proteínas da superfamília FAH, essas enzimas não possuem uma estrutura quaternária na forma de dímero, e, além disso, seu domínio N-terminal se posiciona exatamente sobre a região correspondente do barril- β responsável pela dimerização nas outras proteínas da superfamília. Espera-se que NahL também tenha essa característica.

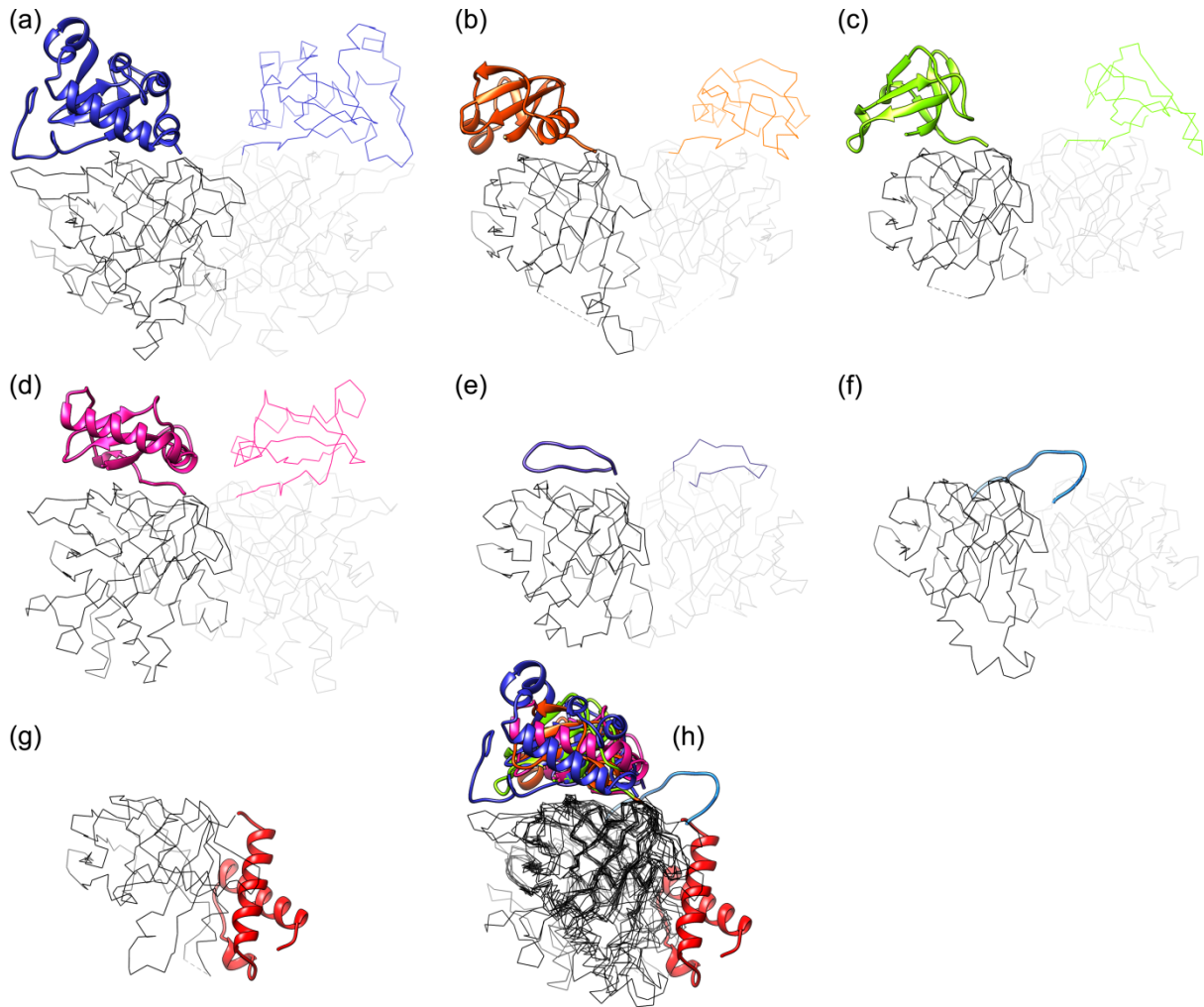


Figura 29. Comparação dos diferentes domínios N-terminal das enzimas da superfamília FAH.

(a) Representação da estrutura terciária e quaternária da enzima FAH de *Mus musculus* (1QCN) e de uma hidrolase putativa de *B. cenocepacia* (4QKU). **(b)** Representação da estrutura terciária e quaternária de KdaD de *S. solfataricus* (2Q18). **(c)** Representação da estrutura terciária e quaternária de enzimas de *M. marinum* (3QDF), *T. thermophilus* (2DFU e 1WZO), *M. abscessos* (3RR6), *C. glutamicum* (4DBF), e de *M. smegmatis* (4PFZ). **(d)** Representação da estrutura terciária e quaternária de enzimas de *M. abscessos* (3R6O) e *S. meliloti* (3LZK). **(e)** Representação da estrutura terciária e quaternária de enzimas de *Homo sapiens* (1SAW), *E. coli* (1NR9), *Y. pestis* (3S52), *O. antarctica* (3V77), *B. cenocepacia* (4MAQ) e de *S. cerevisiae* (1NKQ). **(f)** Representação da estrutura terciária da enzima HpcE de *E. coli* (1GTT). **(g)** Representação da estrutura terciária de NahK de *P. putida*, HpcG (2EB4) e MhpD (2WQT) de *E. coli*. **(h)** Superposição dos monômeros das sete estruturas anteriores, evidenciando o dobramento comum do domínio do barril- β e do posicionamento dos diferentes domínios N-terminais.

Além dessa diferença, HpcG e MhpD formam pentâmeros através da interação de diversos elementos estruturais na superfície do domínio barril (Figura 30 a), e, como um passo a mais na formação de sua estrutura quaternária, HpcG une dois desses pentâmeros em um homodecâmero através de interações entre as folhas- β do domínio C-terminal (Figura 30 b).

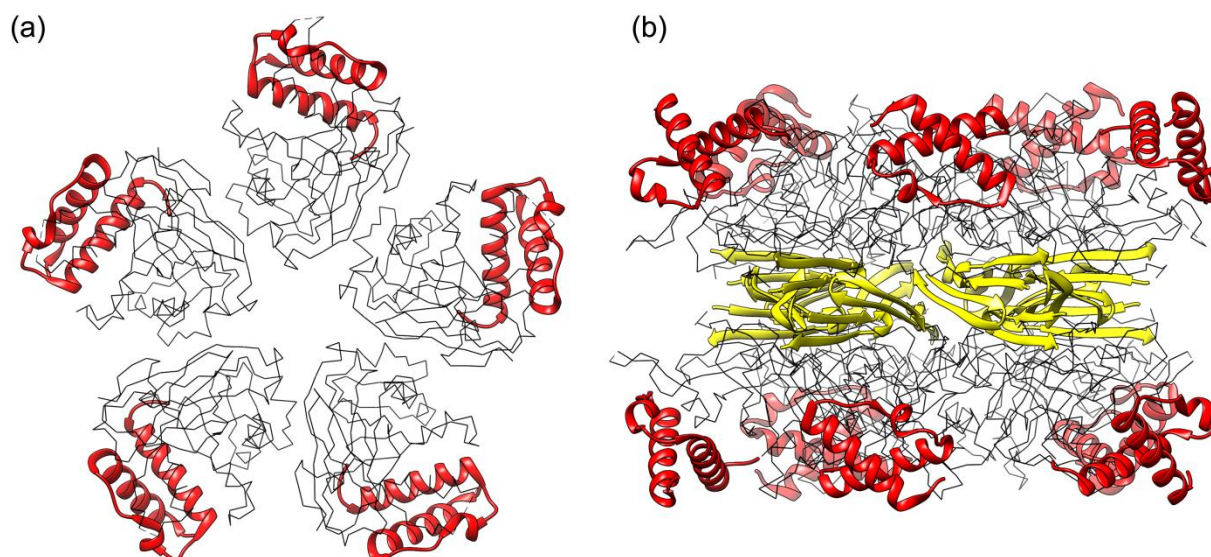


Figura 30. Estrutura quaternária de HpcG.

(a) Anel pentamérico formado por cinco monômeros de HpcG; o domínio N-terminal exibido em vermelho não participa das interações entre os monômeros. **(b)** Estrutura quaternária em homodecâmero de HpcG formada pela união de dois anéis pentaméricos; o domínio C-terminal formado pela folha- β mais planar exibida em amarelo se encontra na interface entre os dois pentâmeros.

Izumi, ao descrever a estrutura de HpcG, destaca que uma volta formada entre os resíduos Asp142 e Ala167 possui diversas interações com o domínio N-terminal, e que a deleção dessa volta, incluindo o domínio N-terminal, não desestabilizaria a proteína ou comprometeria a atividade catalítica [63]. Essa sugestão foi feita uma vez que Izumi observou que o domínio N-terminal de HpcG e MhpD não está envolvido na oligomerização e interage com o solvente sem nenhum contato com as outras subunidades (Figura 30), e que, além disso, em HpcG o barril de cada subunidade interage com outros dois barris do pentâmero oposto, e as demais interações que formam o homodecâmero foram propostas como suficientes para estabilizar a estrutura.

As observações feitas com base na estrutura de NahK mostram que essa sugestão de deleção do loop na respectiva posição (resíduos Asp145 a Ser164) juntamente com o domínio N-terminal poderia desestabilizar o barril- β incompleto e influenciar a atividade enzimática, uma vez que a cadeia lateral do resíduo Ser164 parece ser responsável por auxiliar no correto posicionamento do substrato, enquanto os outros resíduos desta volta formam parte da cavidade do sítio ativo. Essa observação, provavelmente, também poderia ser aplicada à estrutura de HpcG.

Como pode ser observado na Figura 31, o domínio barril de NahK não possui as ligações de hidrogênio entre as fitas opostas da folha- β que forma o barril- β incompleto (fitas 2 e 7), e, adicionalmente, NahK não possui a volta β tipo IV presente em FAH para fechar o barril. Entretanto, o domínio N-terminal em NahK, como mencionado anteriormente também para HpcG e MhpD, se posiciona exatamente sobre essa região do barril, sugerindo uma função de estabilização para o fechamento do barril.

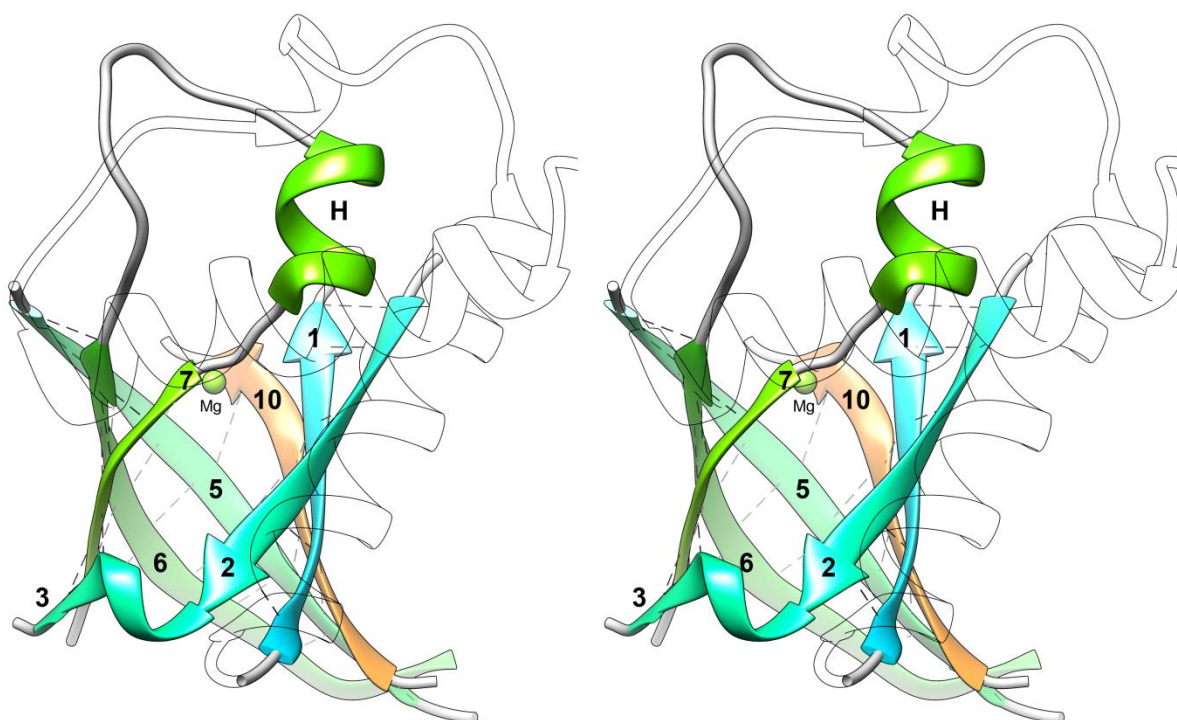


Figura 31: Representação em estéreo (*cross-eyed*) da cadeia principal dos domínios N-terminal e barril- β de NahK.

A parte do domínio barril formado pelas fitas- β estão representadas pelas estruturas em cores, enquanto que o domínio N-terminal encontra-se apenas delineado. O íon magnésio (esfera verde) ligado à proteína se encontra no topo da estrutura em barril- β . As ligações de hidrogênio entre as fitas- β do barril estão representadas por linhas tracejadas.

A hélice C do domínio N-terminal de NahK faz diversas interações hidrofóbicas e ligações de hidrogênio com a fita 2 do barril (Figura 32 a), enquanto a volta compreendida entre os resíduos Asp145 e Ser164, que é uma continuação da fita 7 e também compreende a hélice H, interage com diversos resíduos do domínio N-terminal através de uma interface predominantemente hidrofóbica, possuindo também algumas interações por ligações de hidrogênio (Figura 32 b). Sendo assim, sugere-se que o domínio N-terminal em NahK mantém as fitas 2 e 7 justapostas e o barril fechado mesmo sem grandes interações entre essas fitas.

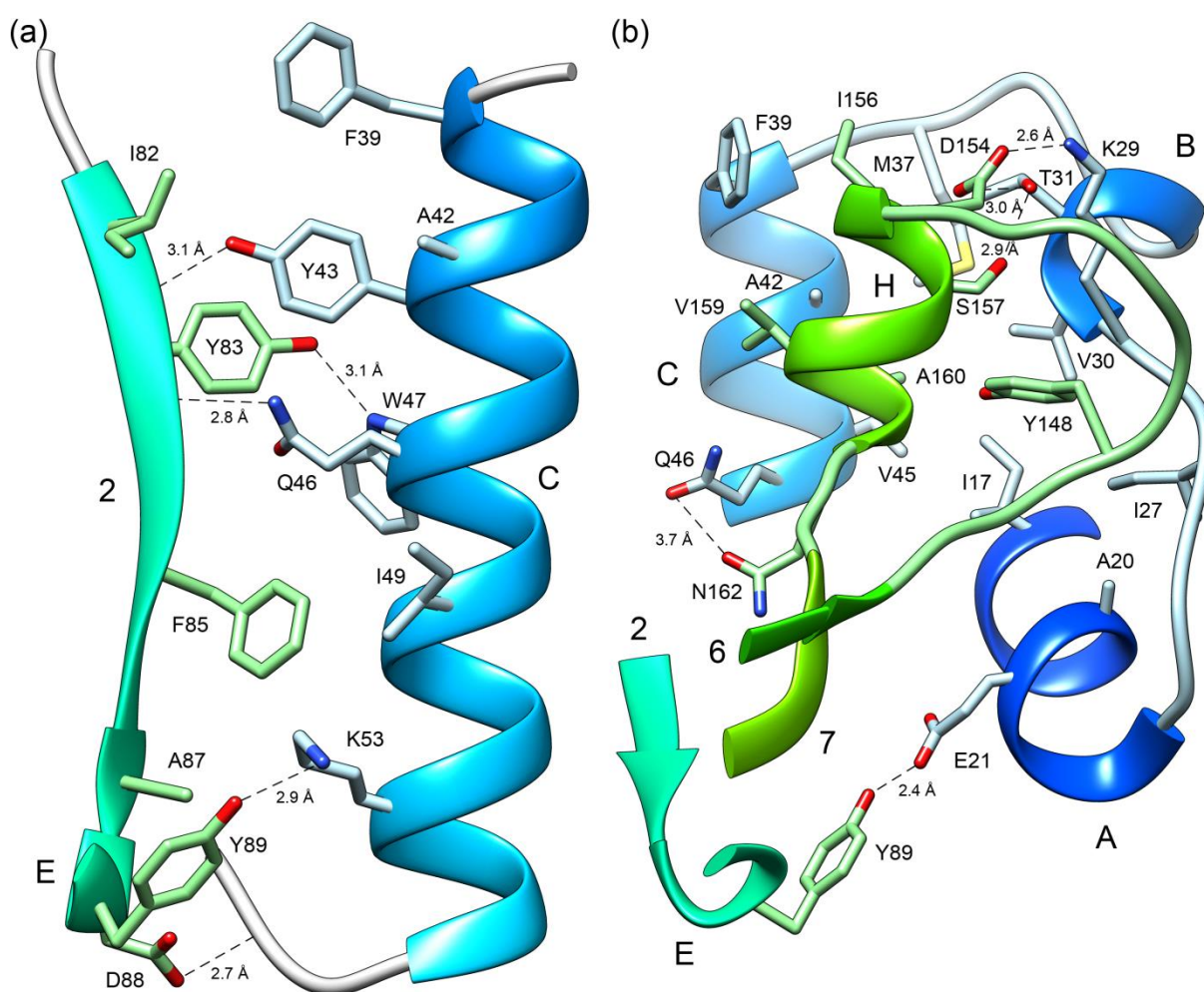


Figura 32: Representação de alguns dos resíduos responsáveis pelas interações entre a fita 2, hélices E e H e uma volta do domínio barril com resíduos do domínio N-terminal em NahK.

(a) Interações diversas da hélice C e fita 2. **(b)** Interações diversas das demais hélices e voltas do domínio N-terminal com a volta compreendida entre os resíduos Asp145 a Ser164, hélice H e E do domínio barril. Resíduos do domínio N-terminal estão coloridos de azul, enquanto resíduos do barril estão coloridos de verde; ligações de hidrogênio estão representadas por linhas tracejadas.

Além da função de estabilização dessa abertura lateral do barril, esse dobramento do domínio N-terminal, que até então parece ser exclusivo de NahK, HpcG e MhpD (e como poderá ser visto adiante, também está presente em NahL), excluiria a necessidade de uma segunda subunidade para exercer a provável função de estabilização, assim como ocorre em todos os demais membros da superfamília FAH.

Segundo a ferramenta PISA [100], que analisa interfaces entre macromoléculas, a área de interação entre as duas subunidades de FAH é de 2778 Å². Entretanto FAH é uma enzima que possui diversos elementos estruturais anexados ao domínio barril quando comparado com a maioria dos membros de sua superfamília, o que aumenta a área de interface entre suas subunidades. Os outros membros que também apresentam a montagem quaternária característica em dímero, mas que possuem menor número de aminoácidos formando o domínio barril, possuem áreas de interface menores que FAH como pode ser observado na Tabela 13.

Tabela 13: Relação entre número de resíduos de aminoácidos que formam o domínio barril e a área de superfície enterrada entre os domínios barril dos monômeros.

PDB ID / Proteína	n° aa do domínio do barril	Área de interface Å ²	%id	PDB ID / Proteína	n° aa do domínio do barril	Área de interface Å ²	%id
NahK*	~206	1325	100	1SAW	~207	1319	15
NahL*	~206	1156	38	1NR9	~204	1389	17
HpcG*	~213	1251	39	3S52	~203	1456	16
MhpD*	~208	1058	31	3V77	~210	1222	18
3QDF	~204	1126	15	4MAQ	~207	2176	16
3RR6	~200	1103	15	3R6O	~221	1453	17
4DBF	~208	1211	12	2Q18	~219	1892	15
1WZO	~203	1375	19	1NKQ	~248	1911	15
2DFU	~217	1511	12	3LZK	~260	2590	14
4PFZ	~201	1195	15	4QKU	~296	2706	14
1GTT	~206	1557	14	FAH	~296	2778	9

* a área de interface corresponde à área enterrada entre o domínio N-terminal e o domínio barril.

Quinze das dezoito proteínas da superfamília FAH, que possuem uma média de 210 resíduos formando o barril, apresentam uma média de 1459,7 Å² de área entre suas duas subunidades, enquanto que as três restantes (3LZK, 4QKU e FAH), que possuem uma média de 284 resíduos no domínio barril, apresentam uma média de 2691,3 Å² de área na interface entre as duas subunidades. Se considerarmos o domínio N-terminal de NahK (resíduos de 1 a 58) como uma cadeia polipeptídica separada do domínio barril, a área de interface entre esses domínios seria de 1325 Å². Ao se aplicar o mesmo raciocínio para as proteínas HpcG,

MhpD e NahL, a área do domínio barril enterrada pelo domínio N-terminal seria de em média 1152 \AA^2 , valor semelhante a área enterrada pelas proteínas que possuem montagem em dímero com número de resíduos de aminoácidos semelhante formando o domínio barril; o que corrobora a função do domínio N-terminal de NahK (como também de HpcG, MhpD e NahL) como substituinte de um segundo monômero para estabilização do barril.

Existe uma diminuição na densidade eletrônica entre os resíduos 150 e 153 (Asp-Phe-Lys-Phe) na cadeia B do modelo *ACT*. Essa porção é visível nos outros modelos e compreende resíduos que se posicionam na lateral da abertura do sítio ativo. Além disso, em todos os modelos de NahK, os resíduos da hélice D e da volta compreendida entre os resíduos 149 e 155 possuem os maiores valores de b-fator dentre os outros resíduos da proteína. Coincidentemente, os resíduos das respectivas regiões em HpcG e MhpD também possuem altos valores de b-fator ou não foram resolvidos na estrutura, sugerindo uma grande flexibilidade dessas regiões nessas proteínas.

Uma vez que o sequenciamento do gene *nahK*, subclonado no vetor pET28a-TEV, ficou incompleto, não foi possível saber da existência de prováveis mutações ocorridas durante sua amplificação por PCR. Assim, somente durante a etapa de refinamento das 6 estruturas de NahK foi possível observar, através de interpretações da nuvem eletrônica do modelo, a mutação de um resíduo de leucina para prolina na posição 155. Apesar dessa mutação, o volume e hidrofobicidade do resíduo substituído não foram significativamente alterados, de forma que praticamente não houve impacto no enovelamento e atividade de NahK. Como será discutido a frente, a estrutura de L155P-NahK enquanto isolada de NahL e de NahK enquanto complexo com NahL se mantém a mesma.

4.6. O sítio ativo de NahK

A enzima Xyll de *P. putida* mt-2 (homóloga de NahK) é uma enzima dependente de metal que requer magnésio para sua atividade [64]. Johnson *et al.* em 2004 identificou resíduos conservados e potencialmente relacionados à catalise através de análises de alinhamento de sequências de aminoácidos de Xyll e homólogas, e construiu mutantes dessa enzima que foram expressos e purificados complexados com XylJ (homóloga de NahL) [95].

Os mutantes E109Q_XylI/XylJ e E111Q_XylI/XylJ perderam a atividade de descarboxilação de 4-OD mas mantiveram a atividade hidratase de VPH. Foi sugerido que os resíduos mutados em XylI (Glu109 e Glu111) poderiam estar envolvidos na ligação ao metal. A estrutura cristalográfica de NahK provê a base estrutural para comprovação dessas observações e confirmação dos papéis desses resíduos.

Nos modelos *Apo* e *Mg* de NahK o sítio ativo foi identificado como uma grande cavidade com aproximadamente 15 Å de profundidade e 19 Å de largura em sua abertura na superfície da molécula (Figura 33), sendo o fundo dessa cavidade localizado no topo da estrutura do barril. De acordo com análises feitas utilizando o servidor CASTp [101], essa cavidade possui um volume de aproximadamente 985 Å³.

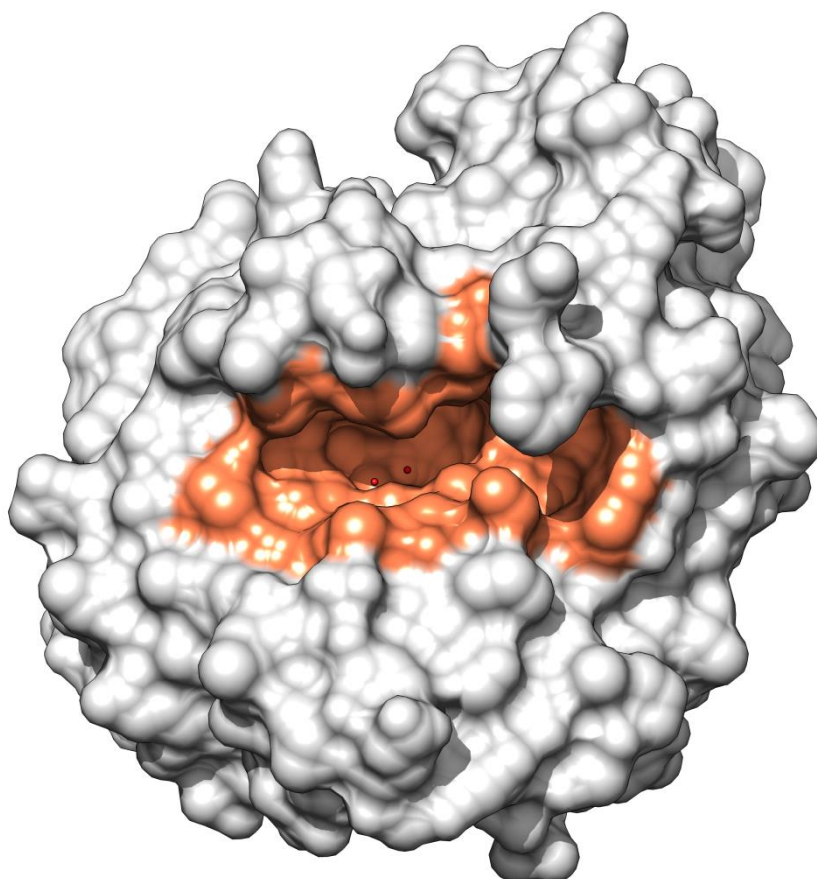


Figura 33: Modelo de superfície de NahK exibindo em destaque a grande cavidade que forma o sítio ativo.

Análises do mapa de densidade eletrônica do fundo dessa cavidade no modelo *Apo* mostra a presença de três moléculas de água compartilhando ligações de hidrogênio com as cadeias laterais de três resíduos de glutamato, não há densidade eletrônica que sugira um

metal ligado à estrutura no sítio ativo deste modelo (Figura 34 a). A Figura 35 (a) exhibe detalhes do sítio ativo do modelo *Apo*, evidenciando a posição dos resíduos Glu109, Glu111 e Glu142 juntamente com a Lys64 e suas interações com as moléculas de água presentes no fundo do sítio ativo; também são exibidas as posições dos resíduos de Lys72 e Met76 da hélice D em uma conformação 'aberta', e do resíduo de Phe153 para comparação com modelo 2OA. O sítio ativo desse modelo possui um volume de aproximadamente 1090 Å³.

Entretanto, no modelo *Mg*, há uma densidade eletrônica que evidencia a clara incorporação de um átomo de metal no sítio ativo (Figura 34 b). Três moléculas de água continuam presentes no fundo da cavidade, sendo que no modelo *Mg* elas coordenam o metal juntamente com os resíduos Glu109, Glu111 e Glu142 de forma octaédrica com distância média de ligações de 1,98 Å. Os detalhes dessas interações podem ser observados na Figura 35 (b). O modelo *Mg* confirma os papéis dos resíduos de glutamato na ligação ao metal. Nota-se que os resíduos Lys72 e Met76 da hélice D, e Phe153 mantêm seu posicionamento 'aberto', e a cavidade do sítio ativo desse modelo possui aproximadamente 880 Å³. Uma das três moléculas de água que coordenam o metal exibidas no modelo *Mg* está conservada no sítio ativo de todos os outros modelos que possui metal ligado à enzima, e encontra-se enterrada entre as cadeias laterais da proteína. Além de coordenar o metal, essa molécula de água também é estabilizada por ligações de hidrogênio com o oxigênio do grupo carbonila da Met65 e com um dos átomos O^ε do Glu111. As outras duas moléculas de água que coordenam o metal são eventualmente deslocadas para permitir o posicionamento dos ligantes dentro do sítio ativo dos outros modelos.

A cadeia lateral do resíduo Lys64 também é conservada e aponta para o sítio ativo em todos os modelos. O átomo N^ζ desse resíduo forma uma ponte salina com um dos átomos O^ε do Glu142 e uma ligação de hidrogênio com uma das águas que coordenam o metal ou, na presença de ligantes, com um dos oxigênios do grupo carboxila dos ligantes (no caso do ligante α-cetoglutarato, acetato, adipato e 2-oxoadipato) ou também com o oxigênio da cetona (2-oxoadipato). Esse resíduo é conservado na mesma posição nos sítios ativos de HpcG e MhpD (Lys61 e Lys64, respectivamente), e também está presente no sítio ativo de NahL (Lys61), sugerindo que esse resíduo conservado nessas enzimas pode ter um papel na catálise ou na ligação ao substrato.

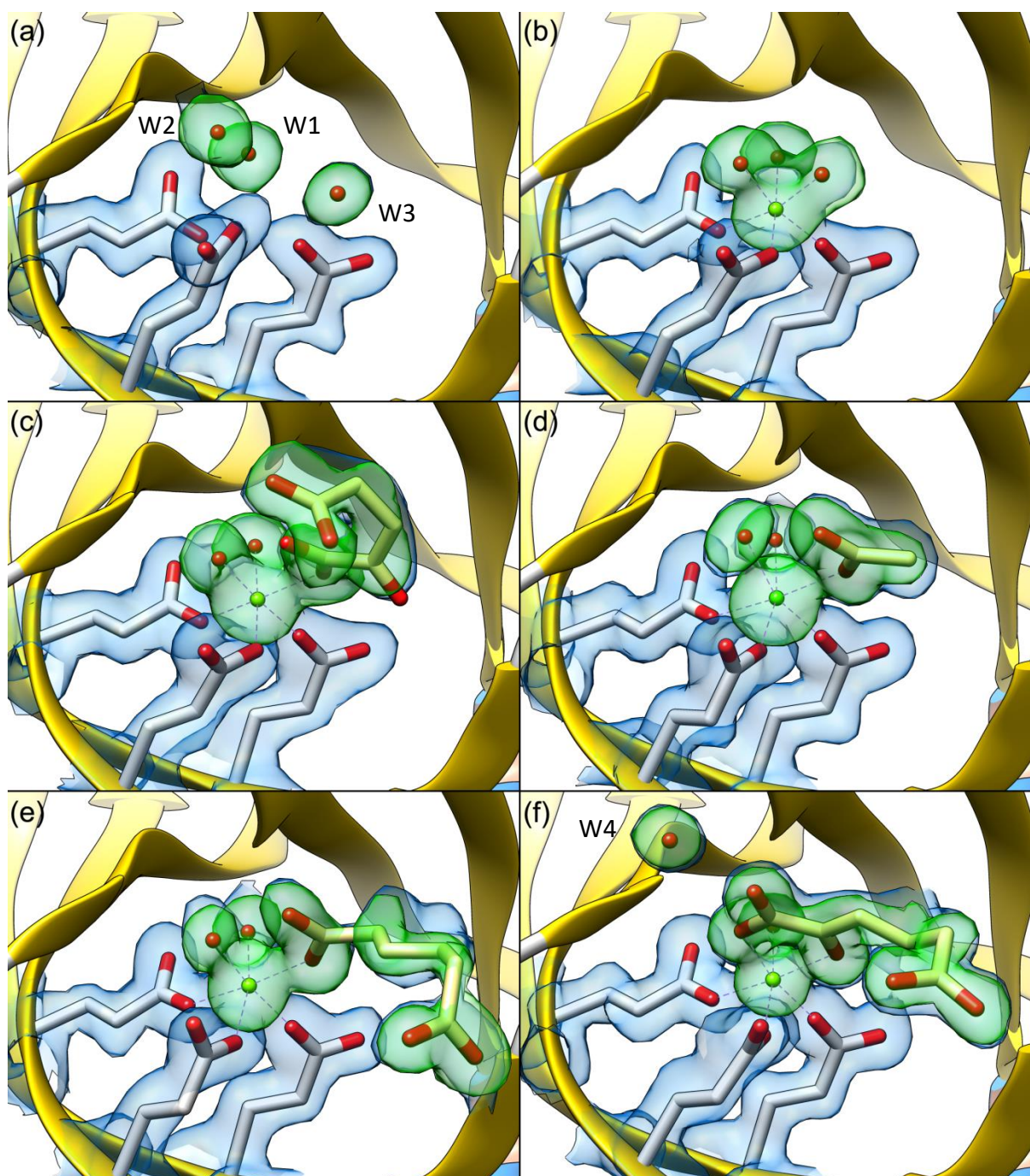


Figura 34: Mapas de densidade eletrônica ilustrando a posição do metal, moléculas de água e ligantes no sítio ativo dos modelos de NahK.

O mapa azul ($2F_o-F_c$) gerado através da utilização das fases do modelo completo, e verde (F_o-F_c) gerado pela omissão dos ligantes estão contornados respectivamente a 1,0 e 4,0 σ . **(a)** Modelo Apo: sítio ativo com três moléculas de água. **(b)** Modelo Mg: sítio ativo com um íon magnésio, incorporado a partir da solução de *soaking*, sendo coordenado por três resíduos de glutamato e três moléculas de água. **(c)** Modelo AKG: além da presença de um íon de magnésio, a densidade eletrônica acima do metal sugere a presença de uma molécula de 2-cetoglutarato coordenando o metal juntamente com duas moléculas de água. **(d)** Modelo ACT: densidade eletrônica no sítio ativo sugere a presença de um íon metálico e uma molécula de acetato. **(e)** Modelo AA: densidade eletrônica sugere a presença de um íon de magnésio e uma molécula de adipato juntamente com duas moléculas de água no sítio ativo. **(f)** Modelo 2OA: a densidade eletrônica de uma molécula de 2-oxoadipato é observada sobre a densidade eletrônica de um íon de magnésio; apenas uma molécula de água está coordenando o metal, uma segunda molécula de água presa dentro do sítio ativo também é exibida.

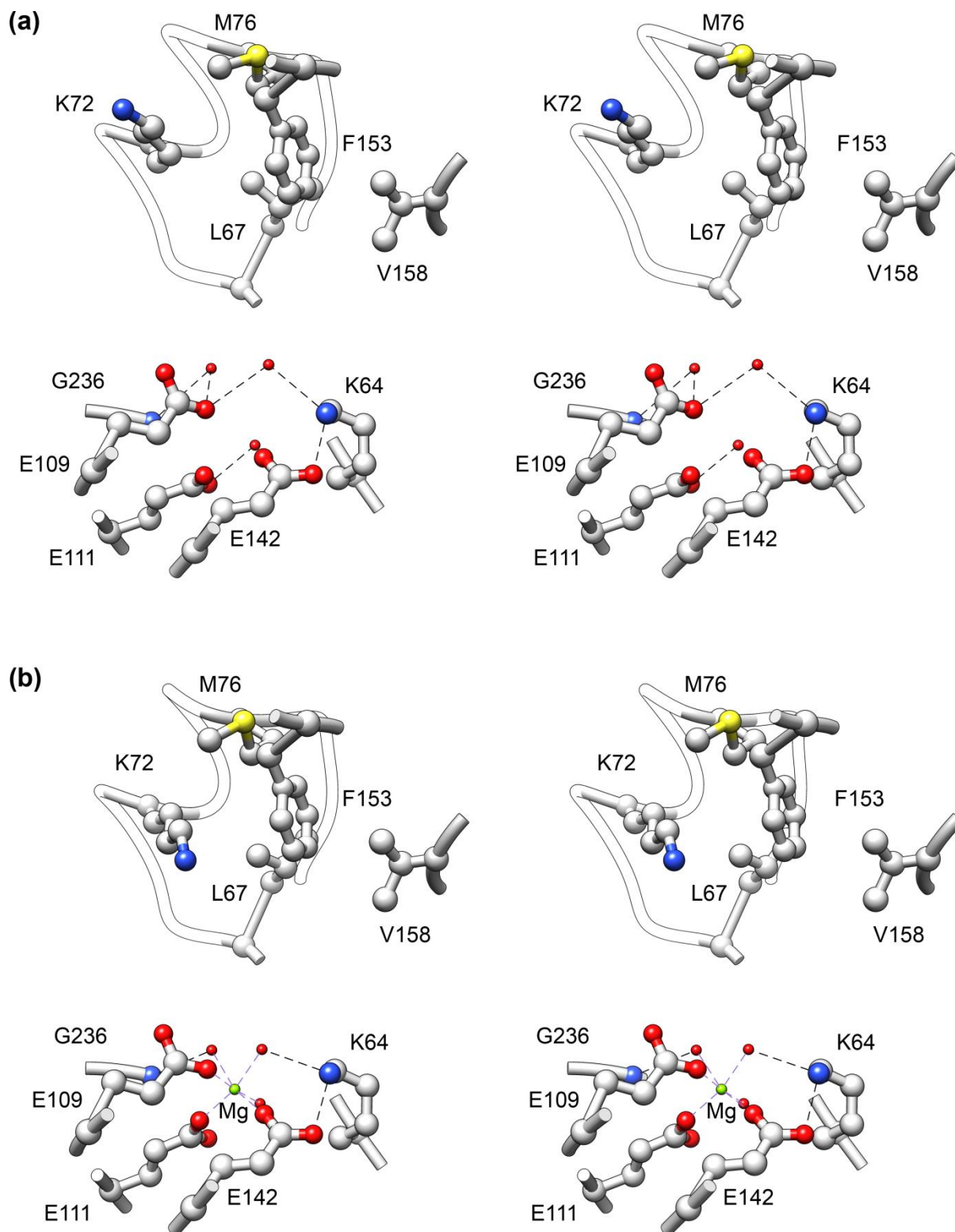


Figura 35: Representação em estéreo (*cross-eyed*) de alguns resíduos no sítio ativo de NahK dos modelos *Apo* (a) e *Mg* (b).

Resíduos da proteína estão representados pelo modelo 'bolas e varetas'. Ligações de hidrogênio ou ligações salinas estão representadas por linhas tracejadas pretas, ligações de coordenação do metal estão representadas por linhas tracejadas roxas.

Os modelos *AKG*, *ACT*, *AA* e *2OA* possuem os ligantes α -cetogluturato, acetato, adipato e 2-oxoadipato nos respectivos sítios ativos desses modelos (Figura 34 c, d, e, f). Todos esses ligantes foram capazes de deslocar ao menos uma molécula de água que coordenava o metal para interagir com este através de um átomo de oxigênio proveniente de seus grupos carboxila ou cetona. Adicionalmente, como mencionado, também há uma interação por ligação de hidrogênio entre este oxigênio do ligante e o átomo N⁷ da Lys64.

No modelo *AKG*, um dos oxigênios do primeiro grupo carboxila (C-1) interage com o metal de forma monodentada (a 2,4 Å de distância), e também com a cadeia lateral da Lys64 por ligação de hidrogênio (2,8 Å), enquanto o outro oxigênio desse grupo interage com uma das moléculas de água que coordenam o magnésio (2,5 Å, Figura 36 a). O oxigênio do grupo cetona desse ligante (C-2) interage por uma ligação de hidrogênio com o grupo carbonila do resíduo Ala163 (2,8 Å). Um dos oxigênios do segundo grupo carboxila do ligante (C-5) interage com o nitrogênio da cadeia principal do resíduo Leu67 através de uma única ligação de hidrogênio (3,0 Å). Nesse modelo a média das distâncias das ligações da coordenação do metal é de 2,06 Å, enquanto que a hélice D continua posicionada de forma 'aberta', contribuindo para a formação da cavidade do sítio ativo com um volume de aproximadamente 923 Å³. Apesar de possuir um grupo cetona posicionado adjacente a um grupo carboxila, esse ligante não foi observado interagindo com o metal de forma bidentada como o ligante 2-oxoadipato no modelo *2OA*.

No modelo *ACT*, que possui duas cadeias proteicas na unidade assimétrica, ambas as moléculas de acetato (uma em cada sítio ativo) interagem com o metal de forma monodentada através de um dos oxigênios de seu único grupo carboxila (distância média de 2,3 Å, Figura 36 b). Esse mesmo oxigênio também interage por ligação de hidrogênio com a Lys64 (2,9 Å). O segundo oxigênio do grupo carboxila interage com uma das águas que coordenam o metal através de uma ligação de hidrogênio (2,4 Å). A volta formada pelos resíduos de 150 a 153 na cadeia B (Asn150, Phe151, Lys152 e Phe153) está ausente nesse modelo devido a uma alta flexibilidade dessa região, mas encontra-se resolvida na cadeia A, e nessa cadeia o volume da cavidade do sítio ativo é de aproximadamente 742 Å³. Apesar de não ter havido um *soaking* para o cristal que gerou esse modelo, há a clara presença de um metal sendo coordenado de forma octaédrica pelos três resíduos de glutamato, por duas moléculas de água e pelo acetato. Como as proteínas da superfamília FAH foram reportadas

como enzimas que se ligam a Ca^{2+} , Mg^{2+} e Mn^{2+} , o posicionamento de um átomo de cálcio justificou bem a densidade eletrônica na posição do metal (o posicionamento de um átomo de magnésio subestimava a densidade eletrônica no local). Além disso, com o refinamento da estrutura, pôde-se observar que a distância média entre o metal e os átomos de oxigênio que o coordenam é de aproximadamente 2,28 Å, enquanto que no modelo *Mg*, com um átomo de magnésio ligado à proteína, essa distância é de aproximadamente 1,98 Å. Essas distâncias estão de acordo com o esperado para as distâncias de coordenação octaédrica entre Mg^{2+} e seis oxigênios (aproximadamente 2,0 Å), e entre Ca^{2+} e seis oxigênios (aproximadamente 2,3 Å) [102]. A incorporação do átomo de cálcio deu-se provavelmente durante a etapa de expressão em *E. coli*, enquanto que a incorporação do ligante acetato deu-se a partir da condição de cristalização que era rica em acetato de sódio. Apesar da condição de cristalização também ser rica em sódio, é menos provável que o metal no sítio ativo no modelo *ACT* seja um Na^+ , pois, além desse metal ter uma nuvem eletrônica menos densa que o cálcio, o sódio adotaria uma simetria octaédrica distorcida e com distâncias bastante variáveis (de 2,3 a 2,7 Å) entre o Na^+ e os seis oxigênios que o coordenariam [103].

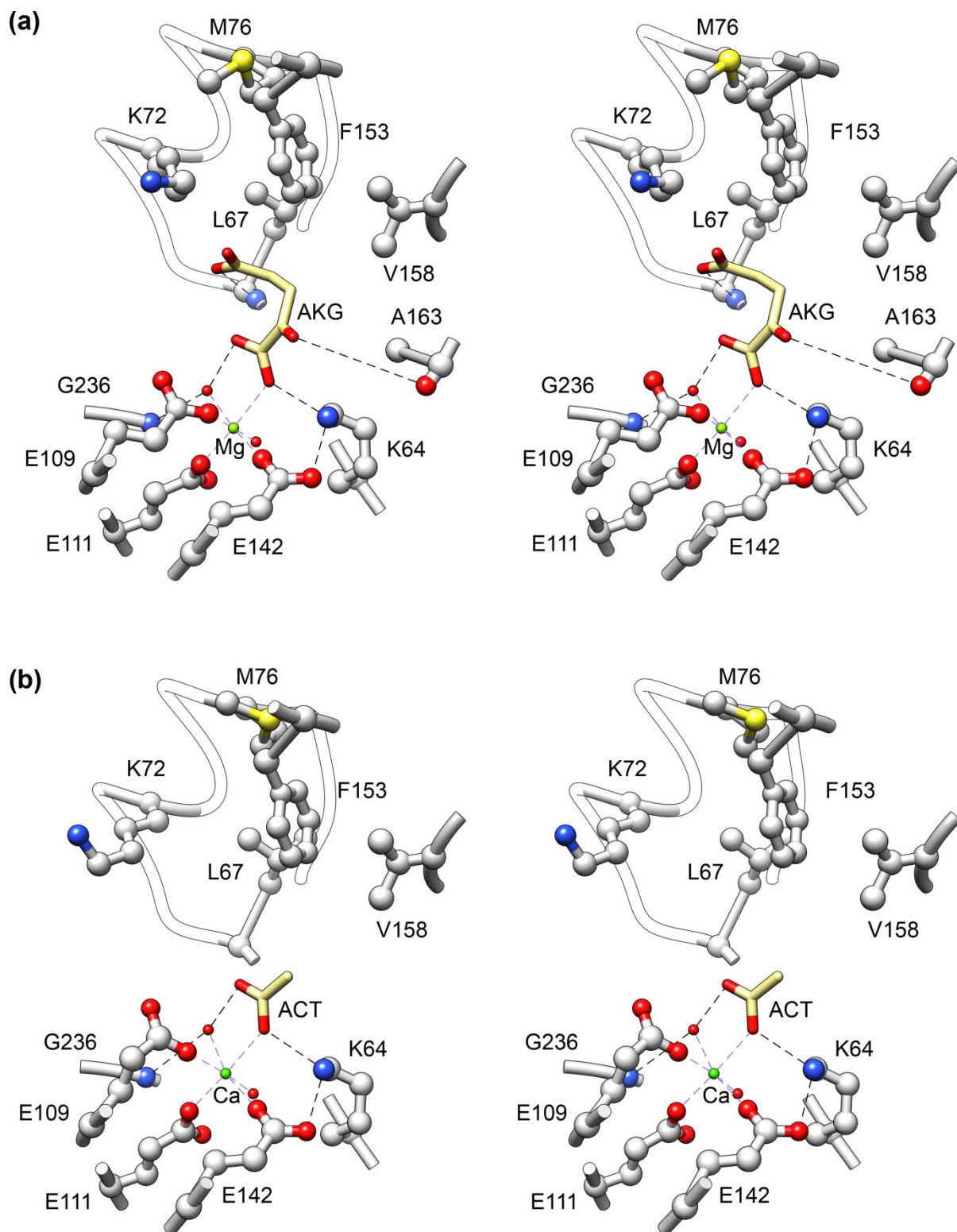


Figura 36: Representação em estéreo (*cross-eyed*) de alguns resíduos e dos ligantes no sítio ativo de NahK, modelos AKG (a) e ACT (b).

Resíduos da proteína estão representados pelo modelo ‘bolas e varetas’, enquanto que o ligante está representado pelo modelo de ‘varetas’. Ligações de hidrogênio ou ligações salinas estão representadas por linhas tracejadas pretas, ligações de coordenação do metal estão representadas por linhas tracejadas roxas.

O modelo AA também possui duas cadeias de proteína por unidade assimétrica, entretanto o ligante adipato foi observado interagindo com o metal de forma monodentada somente no sítio ativo da cadeia A, enquanto que na cadeia B há a presença de um acetato interagindo com o metal no sítio ativo de forma monodentada assim como no modelo ACT. Na cadeia A o átomo de magnésio possui ligações de coordenação com distâncias de aproximadamente 2,15 Å, enquanto que na cadeia B, a média das distâncias é de 2,06 Å. Um dos oxigênios do primeiro grupo carboxila do ligante adipato (C-1) que interage com o metal (2,2 Å) também faz uma ligação de hidrogênio com a Lys64 (2,9 Å), enquanto que o segundo oxigênio desse grupo faz uma ligação de hidrogênio com uma das águas que coordenam o metal (2,6 Å) (Figura 37 a). Os átomos de oxigênio do segundo grupo carboxila desse ligante (C6) fazem uma interação por ligação de hidrogênio com a cadeia lateral da Ser164 (2,7 Å) e uma interação por uma ligação de hidrogênio fraca com a Lys72 (3,6 Å). Enquanto a cavidade do sítio ativo da cadeia B (que contém o acetato) é de aproximadamente 798 Å³, a cavidade do sítio ativo da cadeia A (que contém o adipato) é ligeiramente menos volumosa comparada com as cavidades dos modelos Apo, Mg, AKG e ACT, apresentando um volume de aproximadamente 690 Å³.

É importante notar que, devido ao posicionamento das cadeias laterais dos resíduos hidrofóbicos Leu67, Met76, Phe153 e Val158 formando parte da cavidade do sítio ativo e criando um impedimento estérico para o posicionamento linear desse ligante nessa cavidade, juntamente com as interações polares dos resíduos Lys72 e Ser164 que estabilizam o segundo grupo carboxila do ligante, a porção distal do adipato (simbolizada pelo plano vermelho na Figura 37 a) fica posicionada com uma geometria 'quebrada' com um ângulo de aproximadamente 69° com relação ao restante do plano do ligante (simbolizado pelo plano alaranjado na Figura 37 a). A importância dessa geometria será discutida adiante.

O modelo 2OA também possui duas cadeias de proteína em sua unidade assimétrica. Na cadeia A observa-se o ligante 2-oxoadipato interagindo com o metal de forma bidentada através de um dos oxigênios do primeiro grupo carboxila (C-1) (2,1 Å) e o oxigênio do grupo cetona (C-2) (2,2 Å), que também interage com a Lys64 através de uma ligação de hidrogênio (2,7 Å) (Figura 37 b). O segundo oxigênio do primeiro grupo carboxila interage com uma água aprisionada dentro da cavidade do sítio ativo através de uma ligação de hidrogênio (2,7

Å). O segundo grupo carboxila do ligante (C-6) é estabilizado através de ligações de hidrogênio com a cadeia lateral da Lys72 (2,7 Å) e Ser164 (2,8 Å). A cadeia lateral da Lys72 também faz uma ligação de hidrogênio com o Glu109 (2,5 Å).

Assim como observado para o ligante adipato, o ligante 2-oxoadipato também não se posiciona linearmente no sítio ativo, sendo impedido estericamente pelas cadeias laterais dos resíduos hidrofóbicos Leu67, Met76, Phe153 e Val158. Esse impedimento, juntamente com as interações polares do segundo grupo carboxila do ligante com os resíduos Lys72 e Ser164, também fazem com que essa porção distal do ligante (representada pelo plano vermelho na Figura 37 b) fique com uma geometria 'quebrada' com um ângulo de 70° com relação ao plano de todo o restante da molécula (plano alaranjado na Figura 37 a). Essa geometria pode ser importante no mecanismo de descarboxilação de NahK.

Ainda na cadeia A do modelo 2OA, quando comparado com os outros modelos, a hélice D faz um grande movimento em direção ao ligante, diminuindo consideravelmente o volume do sítio ativo para apenas 247 Å³ e posicionando as cadeias laterais dos resíduos Lys72 e Met76 justapostos ao ligante. Nessa conformação 'fechada' da hélice D, a cavidade do sítio ativo é separada do solvente que banha a enzima, e fica complementar em forma e carga com a molécula 2-oxoadipato. É importante ressaltar que o ligante 2-oxoadipato é extremamente similar em forma e carga ao substrato natural da enzima NahK, com a única diferença sendo a adição de dois átomos de hidrogênio devido a ausência da ligação dupla entre os carbonos 3 e 4 (Figura 12, moléculas 1 e 2). Sendo assim, podemos sugerir que o substrato natural de NahK se comportaria dentro do sítio ativo da enzima de forma semelhante ao ligante 2-oxoadipato, induzindo, provavelmente, a mesma mudança na posição da hélice D e adotando a mesma geometria em que o ligante ficou capturado na cavidade do sítio ativo da cadeia A.

Os magnésios, tanto na cadeia A como na cadeia B do modelo 2OA exibem ligações de coordenação com simetria octaédrica e distâncias médias de 2,09 Å.

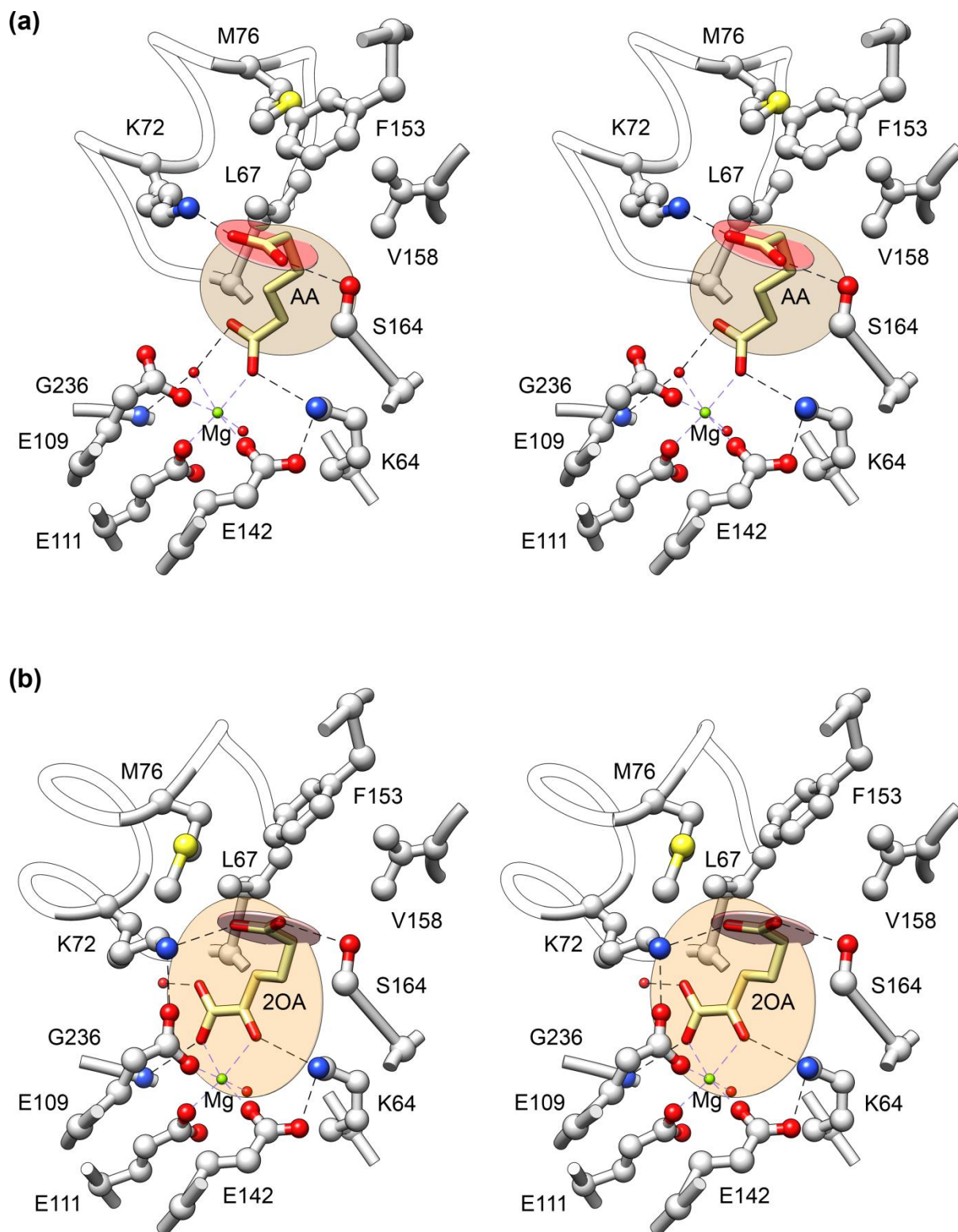


Figura 37: Representação em estéreo (*cross-eyed*) de alguns resíduos e dos ligantes no sítio ativo de NahK, cadeia A do modelo AA **(a)** e cadeia A do modelo 2OA **(b)**.

Resíduos da proteína estão representados pelo modelo ‘bolas e varetas’, enquanto que o ligante está representado pelo modelo de ‘varetas’. Ligações de hidrogênio ou ligações salinas estão representadas por linhas tracejadas pretas, ligações de coordenação do metal estão representadas por linhas tracejadas roxas. Os círculos alaranjados representam o plano formado pelos carbonos de 1 a 5 dos ligantes, os círculos vermelhos representam os planos do segundo grupo carboxila dos ligantes juntamente com o carbono 5.

Na cadeia B do modelo 20A o ligante é visualizado em uma conformação dupla (Figura 38). Nessa cadeia a hélice D permanece em sua conformação ‘aberta’, e a cavidade do sítio ativo possui um volume de aproximadamente 890 Å³. Os posicionamentos do ligante nessa cadeia sugerem os prováveis passos de interação do substrato natural ao entrar no sítio ativo: primeiro (conformação do ligante colorida em verde) o ligante deslocaria uma molécula de água que coordena o metal a fim de criar uma interação monodentada com esse (assim como os ligantes α -cetoglutarato, acetato e adipato também fizeram), enquanto a Ser164 já inicia o posicionamento do segundo grupo carboxila do ligante; posteriormente (conformação do ligante colorida em amarelo) o ligante deslocaria uma segunda molécula de água que coordena o metal e interagiria com esse de uma forma bidentada (como ocorre na cadeia A do modelo 20A), a hélice D então se move, posicionando-se sobre o ligante e fechando a cavidade do sítio ativo para que ocorra a reação.

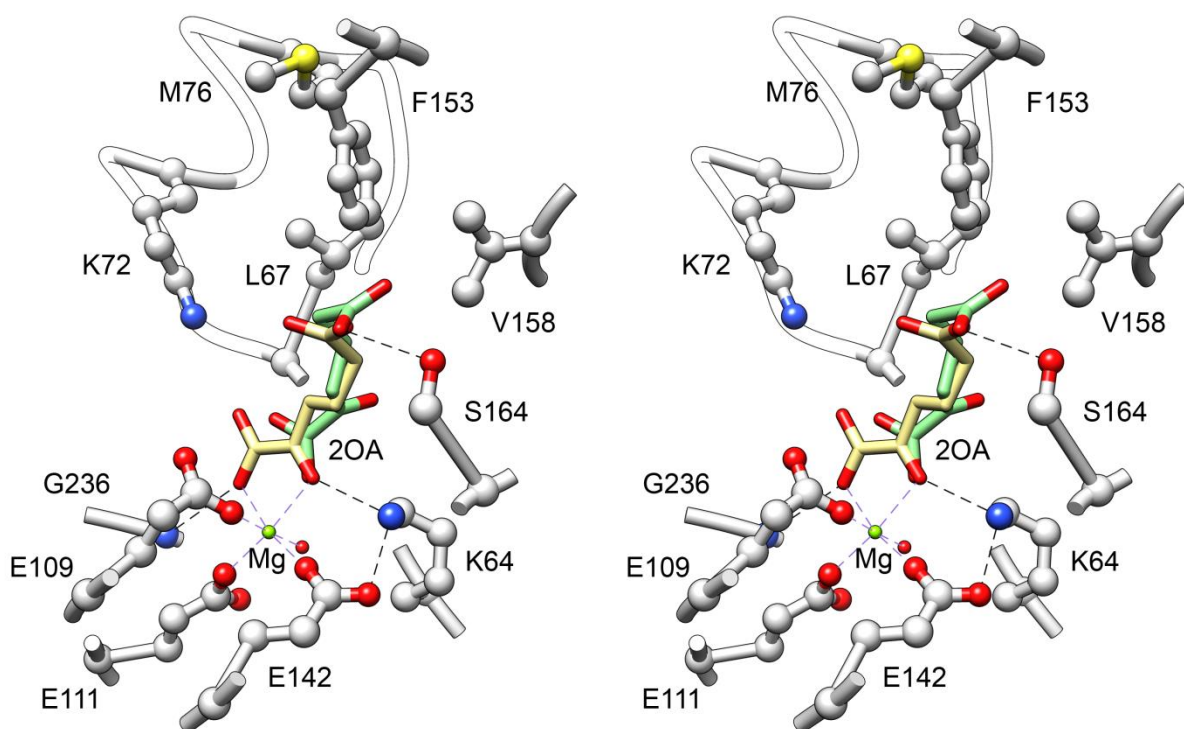


Figura 38. Representação em estéreo (*cross-eyed*) de alguns resíduos e dos ligantes no sítio ativo de NahK, cadeia B do modelo 20A.

Resíduos da proteína estão representados pelo modelo ‘bolas e varetas’, enquanto que o ligante está representado pelo modelo de ‘varetas’. Ligações de hidrogênio ou ligações salinas estão representadas por linhas tracejadas pretas, ligações de coordenação do metal estão representadas por linhas tracejadas roxas.

4.7. NahK possui um domínio ‘tampa’

Ao analisar as estruturas de NahK, nota-se que a hélice D manteve-se numa posição que chamamos de ‘aberta’ nos modelos *Apo*, *Mg*, *AKG*, *ACT*, *AA* e cadeia B do modelo *2OA*, contribuindo para a formação da cavidade do sítio ativo com um grande volume. Entretanto, como já apresentado, na cadeia A do modelo *2OA*, com a ligação da molécula 2-oxoadipato, a hélice D fez um grande movimento em direção ao ligante, preenchendo a maior parte da cavidade do sítio ativo e protegendo o ligante do solvente. Chamamos essa conformação de ‘fechada’. A Figura 39 representa as cadeias A dos modelos *Mg* e *2OA*, e evidencia, através da comparação das posições dos C $^{\alpha}$ de cada resíduo, a movimentação da hélice D e da volta formada entre os resíduos de 149 a 154 devido a ligação do 2-oxoadipato. Em sua maior diferença, a variação chegou a um rmsd de 9,5 Å para o resíduo Gln75. Não coincidentemente, essas mesmas regiões também possuem os maiores valores de b-fator comparado com os demais resíduos de superfície da proteína.

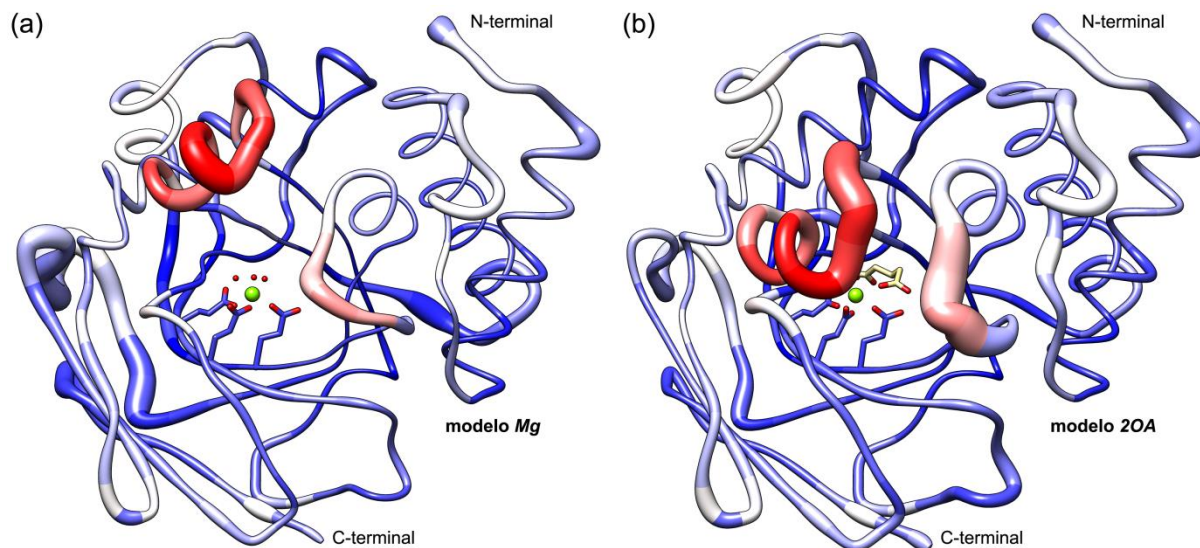


Figura 39. Representação em ‘worm’ do modelo *Mg* e da cadeia A do modelo *2OA*.

As tonalidades de cores representam o rmsd dos pares de C $^{\alpha}$ entre as duas cadeias, indo de azul (rmsd entre 0.088 e 0.9 Å), passando por branco (rmsd = 1.0 Å), até vermelho (rmsd entre 1.1 e 9.504 Å), evidenciando as diferenças de posições entre as cadeias carbônicas dos dois modelos. A espessura do tubo representa o b-fator médio do resíduo, evidenciando regiões com maior mobilidade térmica na proteína.

A região correspondente à hélice D de NahK não foi resolvida em algumas estruturas cristalinas de outras proteínas da superfamília FAH (PDB IDs: 1NR9, 4MAQ, 4PFZ, 4QKU, 1SAW [67], 1GTT [60], 3QDF e 3R6O [98], 1WZO e 2DFU), sugerindo que essa é uma região flexível e conservada na superfamília. Ao mesmo tempo em que, para duas outras enzimas da superfamília FAH em que há estruturas cristalinas ligadas a análogos de substratos ou inibidores (KdaD [68] e Cg1458 [69]), essa região, além de ter sido resolvida, é observada fechando o topo da cavidade do sítio ativo, bloqueando o ligante do solvente e estabilizando-o dentro do sítio catalítico. De fato, observa-se que a maior parte das estruturas cristalinas em que falta esse domínio tampa são estruturas nativas sem substratos, produtos ou inibidores ligados as enzimas, evidenciando que esse domínio está presente em outros membros da superfamília FAH e pode ser outra característica em comum além do barril- β incompleto [68,69].

4.8. Mecanismo de descarboxilação de NahK

A via de *meta*-clivagem do catecol (intermediário para o qual a via de degradação do naftaleno converge, e compreende a maior parte da via inferior do plasmídeo NAH7) é estudada há mais de 46 anos em diversas espécies de *Pseudomonas* [104-106]. Entretanto, no final dos anos 60, ainda restavam incertezas quanto aos substratos/produtos intermediários dessa via (Figura 40).

Inicialmente, em 1969, achava-se que o produto da descarboxilação de 4-oxalocrotonato pela enzima 4-OD era a molécula 2-oxo-4-pentenoato (Figura 40 5c) [105]. Em 1973 foi estabelecido que o produto da enzima VPH trata-se da molécula L-(S)-4-hidroxi-2-oxopentanoato (6c), e não seu D-enantiômero (6b) ou um de seus isômeros DL-2,4-dihidroxi-2-pentenoato (6a) [104]. Em 1989, Harayama e Whitman identificaram a molécula 2-oxo-3-hexeno-1,6-dioato (4b) como o produto da enzima 4-OT, ao invés de outro provável isômero (4a), e que essa forma cetônica é o substrato da enzima 4-OD. Além disso, Harayama, através de estudos de absorvância de luz ultravioleta, sugeriu no mesmo trabalho que a molécula 2-hidroxi-2,4-pentadienoato (5a) era o produto mais provável de 4-OD e substrato de VPH ao invés de duas possíveis formas cetônicas (5b e 5c) [64], mas ainda

assim o produto de 4-OD permanecia incerto. Somente em 2000, com a expressão de 4-OD complexada com VPH inativada (XylI/E106Q_XylJ), ambas as enzimas providas do plasmídeo TOL pWW0, Stanley e Whitman confirmaram o produto de 4-OD através de estudos de ressonância nuclear magnética (Figura 40 5a) [65].

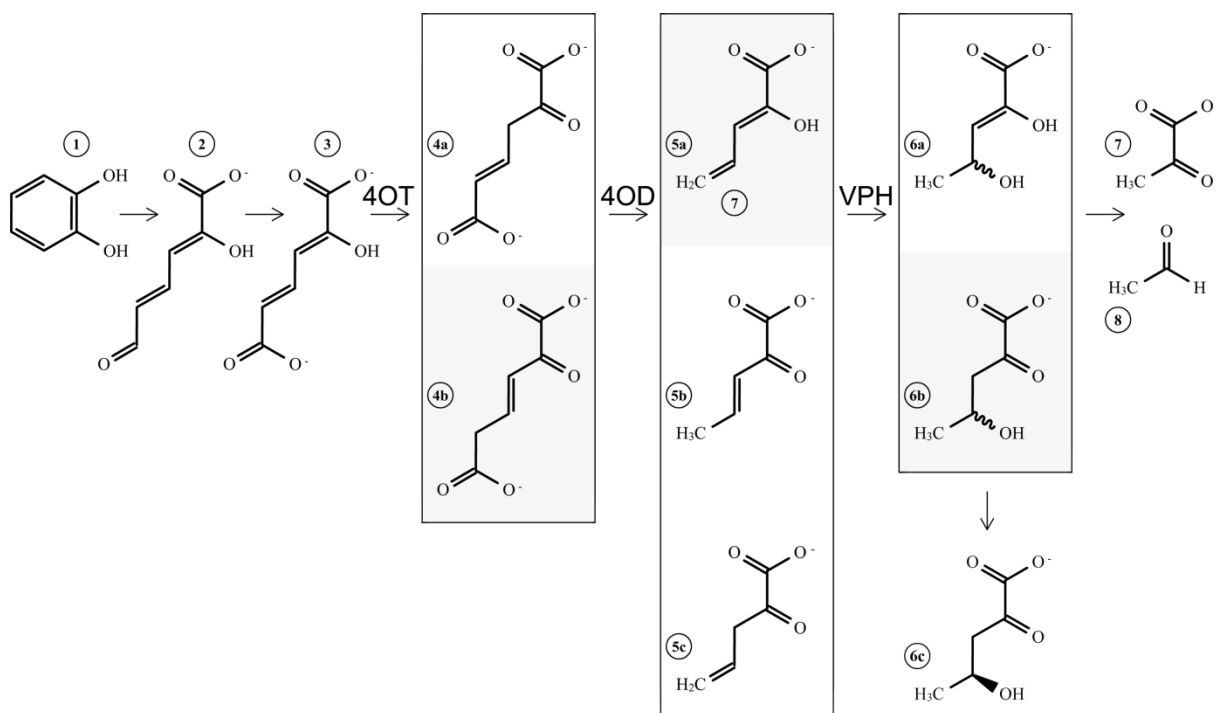


Figura 40: Via de degradação por metaclivagem do catecol e intermediários incertos até meados de 1970. **(1)** catecol; **(2)** 2-hidroximuconato semialdeído; **(3)** 2-hidroximuconato; **(4a)** 2-oxo-4-hexeno-1,6-dioato; **(4b)** 2-oxo-3-hexeno-1,6-dioato; **(5a)** 2-hidroxi-2,4-pentadienoato; **(5b)** 2-oxo-3-pentenoato; **(5c)** 2-oxo-4-pentenoato; **(6a)** DL-2,4-dihidroxi-2-pentenoato; **(6b)** DL-4-hidroxi-2-oxopentanoato; **(6c)** L-4-hidroxi-2-oxopentanoato; **(7)** piruvato; **(8)** acetaldeído. Os intermediários destacados pelo quadro cinza são os corretos da via.

Até a presente data, existem três descarboxilases pertencentes à superfamília FAH: 4-oxalocrotonato descarboxilase (NahK, XylI e outras homólogas), 5-oxopent-3-eno-1,2,5-tricarboxilato descarboxilase (HpcE) [60], e uma oxaloacetato descarboxilase Cg1458 [69]. Ran e colaboradores propuseram que a enzima Cg1458 utiliza um mecanismo envolvendo uma tríade glutamato-histidina-água para descarboxilar o oxaloacetato. Apesar de NahK e Cg1458 pertencerem a superfamília FAH e utilizarem Mg^{2+} como cofator, NahK não possui o resíduo de histidina na posição equivalente em seu sítio catalítico. Além disso, nossos dados estruturais sugerem que as únicas moléculas de água presentes no sítio catalítico após o fechamento do domínio tampa são as águas W1 e W4, e ambas estão distantes dos átomos de carbono C5-6 do substrato para interagir com os mesmos e iniciar um ataque nucleofílico.

Como apresentado, HpcE e NahK possuem substratos bastante semelhantes, sendo que a única diferença é a presença de um grupo carboximetil ligado ao carbono C-5 ao invés de um hidrogênio. Entretanto, apesar da estrutura de HpcE já estar resolvida e descrita na literatura, não foi proposto nenhum mecanismo catalítico para essa enzima. Além disso, foi sugerido que o metal em seu sítio ativo serviria apenas para manter os resíduos catalíticos no lugar, e não participaria no mecanismo de descarboxilação [60].

Vários mecanismos de descarboxilação têm sido descritos para diferentes enzimas. Como NahK não utiliza outro cofator além de um simples íon metálico, mecanismos que são dependentes de piridoxal-5'-fosfato, biotina, flavina, NAD⁺/NADP⁺ e outros cofatores são descartados [107,108]. Mecanismos baseados em uma base de Schiff também foram descartados [109]. Com o produto de 4-OD finalmente estabelecidos por Stanley em 2000 [65], foi sugerido que o mecanismo de descarboxilação de 4-OD poderia ocorrer de forma semelhante à descarboxilação da enzima oxaloacetato descarboxilase (enzima diferente de Cg1458).

O oxaloacetato é descarboxilado não enzimaticamente em solução aquosa em pH neutro, mas a reação é acelerada por íons metálicos. O produto imediato da descarboxilação é um enolato ligado ao metal, que, posteriormente, captura um próton para transformar-se no produto final (Figura 41) [107].

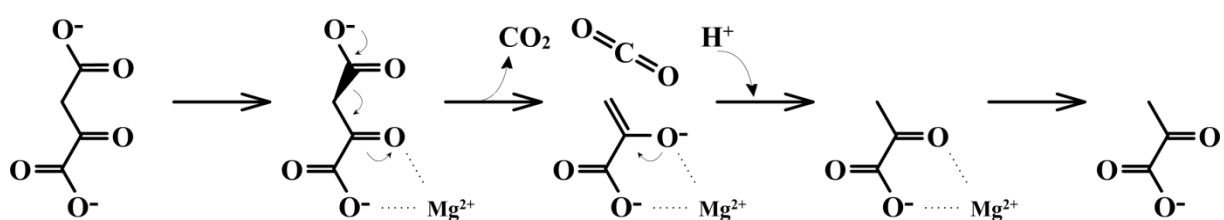


Figura 41. Mecanismo de descarboxilação do oxaloacetato catalisada por metal.

No mecanismo de ação da enzima oxaloacetato descarboxilase, assim como no mecanismo não enzimático catalisado por metal, o íon metálico se liga ao grupamento α -cetoácido do substrato antes da descarboxilação, que ocorre pelo deslocamento da nuvem eletrônica do substrato em direção ao metal, enfraquecendo e rompendo a ligação entre os carbonos C3 e C4. No último passo, enquanto na reação não enzimática um próton é adicionado no C3 de forma não estereoespecífica, gerando um produto racêmico, na reação

enzimática a cadeia lateral de uma lisina é responsável pela doação do próton, gerando um produto estereoespecífico. A reação enzimática é mais rápida que a não enzimática catalisada por metal por um fator de aproximadamente 10^3 [107,110]. Assim como nessa oxaloacetato descarboxilase, a utilização de um íon metálico na estrutura da proteína para a formação e estabilização do enolato é provavelmente a estratégia catalítica adotada por NahK.

Sabe-se que a geometria do estado de transição do substrato é um aspecto importante dentre os fatores que as enzimas podem controlar para facilitar a descarboxilação. O grupo CO_2 que será adicionado ou removido da molécula deve se posicionar acima ou abaixo do plano do restante do substrato, ao invés de se posicionar juntamente a esse plano [107]. A Figura 42, adaptada de O'Leary, exhibe uma molécula cujo plano principal é formado por três átomos de carbono e pelo átomo de oxigênio do grupo cetona; enquanto que o grupo CO_2^- que será removido se posiciona acima do plano dessa molécula. O controle enzimático dessa conformação pode, por si só, aumentar a velocidade da reação em duas ordens de grandeza.

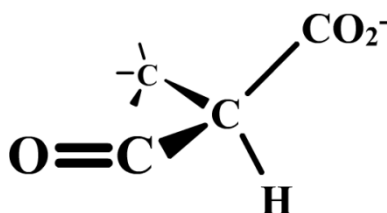


Figura 42: Conformação favorável para descarboxilação.

Também se sabe que a interconversão de grupo carboxila (R-CO_2^- , polar e carregado) a CO_2 (não polar e não carregado) é sensível a solvatação. Um ambiente polar favorecerá a carboxilação, enquanto que um ambiente apolar favorecerá a descarboxilação [107,108].

Outro aspecto importante para a proposição do mecanismo de descarboxilação em NahK é o fato do substrato possuir uma ligação dupla entre os carbonos C-3 e C-4 (**1**, Figura 12), caracterizando um grupo vinil, e de essa ligação dupla estar conjugada com o grupo cetona no carbono 2. A vinilogia, ou reatividade análoga a vinil, reconhece a possibilidade de que o efeito indutivo de algum grupo funcional possa ser transmitido através de uma ligação dupla ou de um sistema conjugado de ligações duplas [111]. Um exemplo simples da vinilogia pode ser observado na Figura 43, onde a deslocação de uma carga negativa em um

ânion carboxilato (a) ocorre também de forma correspondente em seu análogo vinílico (b), no qual os átomos de oxigênio que formam os grupos carbonil estão separados por um grupo vinil, se comportando e exibindo as mesmas propriedades de um grupo carboxilato.

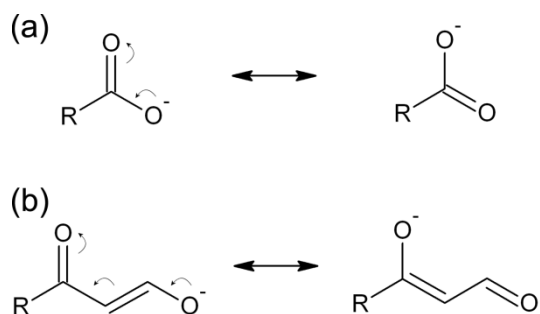


Figura 43: Deslocação de cargas negativas em um ânion carboxilato genérico (a) e seu respectivo ânion carboxilato vinílico (b).

Dessa forma, percebe-se que a molécula 4-oxalocrotonato nada mais é que um vinílico correspondente do oxaloacetato, sendo que o grupo carboxilato que é descarboxilado se encontra separado do grupamento α -cetoácido por um grupo vinil. Assim, o efeito da deslocação da nuvem eletrônica em direção ao metal, que enfraquece e rompe a ligação entre os carbonos C-3 e C-4 na descarboxilação do oxaloacetato, pode ser transmitido através do grupo vinil na molécula 4-oxalocrotonato e, da mesma maneira, enfraquecer e romper a ligação entre os carbonos C-5 e C-6.

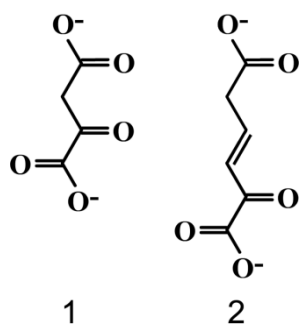


Figura 44: Oxaloacetato (1) e seu correspondente vinílico 4-oxalocrotonato (2).

Baseando-se nesses conhecimentos prévios, juntamente com os modelos de estrutura de NahK em sua forma apo e complexadas, foi possível propor a forma de interação da proteína com o substrato, a estratégia e o mecanismo de descarboxilação dessa enzima.

Assim como observado para uma das conformações do ligante 2-oxoadipato na cadeia B do modelo 2OA, e em todos os outros ligantes dos modelos *AKG*, *ACT* e *AA*, ao entrar no sítio ativo o substrato 4-oxalocrotonato desloca uma molécula de água que estava coordenando o metal a fim de interagir de forma monodentada com o magnésio através de um dos oxigênios de seu primeiro grupo carboxila, e também com a Lys64 (Figura 45 a). A cadeia lateral do resíduo Ser164 inicia o posicionamento do segundo grupo carboxila enquanto o substrato desloca uma segunda molécula de água para interagir com o metal de forma bidentada (b) (assim como o ligante 2-oxoadípico na cadeia A do modelo 2OA); a cadeia lateral da Lys72 também interage com o segundo grupo carboxila mesmo sem o fechamento da hélice D, já o posicionando num plano fora do plano do restante da molécula, deixando-o numa geometria favorável à descarboxilação. Com o fechamento da hélice D, o segundo grupo carboxila é posicionado contra as cadeias laterais hidrofóbicas dos resíduos Met76, Phe151 e Phe153, se encontrando agora num ambiente apolar, o que também facilita a descarboxilação (c). Com a interação direta entre o metal e o grupo carbonila do substrato (c) é favorecida a forma de dienolato da molécula, criando uma carga negativa no oxigênio da cetona e translocando a ligação dupla entre os carbonos C-3 e C-4 para os carbonos C-2 e C-3 (d), enfraquecendo assim a ligação entre os carbonos C-5 e C-6. Pela somatória desses três fatores (geometria favorável, ambiente apolar, e translocação da nuvem eletrônica em direção ao metal), ocorre a quebra da ligação entre os carbonos C-5 e C-6, e o grupo carboxila se transforma numa molécula de CO₂ (d). O intermediário dienolato recebe um próton da cadeia lateral do resíduo Lys64 (d) se transformando em um dienol (e) que é o produto liberado pela enzima (f) após o desligamento do substrato do magnésio e abertura da hélice D. O sítio ativo se regeneraria através da captura de um próton do meio para restaurar a carga positiva da Lys64 e permitir um novo ciclo de catálise.

Dessa forma, observa-se que NahK não utiliza nenhum resíduo para se ligar covalentemente ao substrato ou a intermediários, como é comum em diversos outros mecanismos de descarboxilação [108]. Ao invés disso, NahK estabiliza o substrato numa geometria favorável para a reação, provê um íon metálico fixado em sua estrutura para

auxiliar no deslocamento da nuvem eletrônica do substrato e estabilizar a forma de dienolato, e também provê um ambiente hidrofóbico livre de solvente para facilitar a conversão de ácido carboxílico para CO_2 . Controlando todos esses fatores a reação catalisada por essa enzima chega a valores elevados de eficiência próximos do limite de difusão das moléculas.

As descobertas aqui feitas para NahK foram submetidas para publicação sob a forma de um artigo científico exibido no Anexo I.

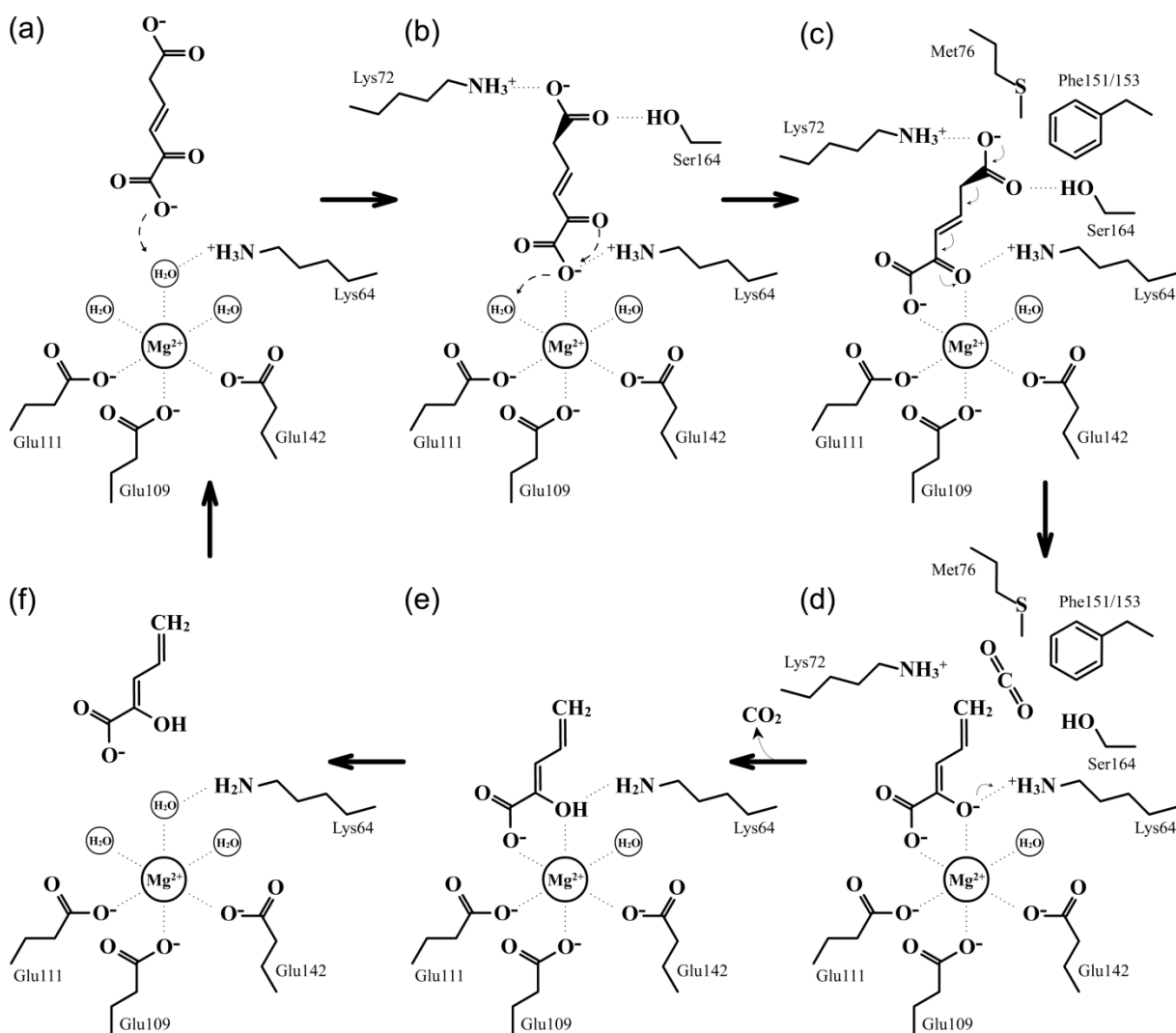


Figura 45. Mecanismo proposto para a descarboxilação de 4-oxalocrotonato por NahK.

4.9. Produção de NahL

O gene *nahL* de *P. putida* G7 (aproximadamente 800 pb incluindo as sequências das enzimas de restrição presente nos iniciadores) foi amplificado por PCR utilizando os iniciadores nahL-R-pET e nahL-F-pET (Figura 46), purificado e ligado ao vetor de clonagem pGEM-T (Promega) e inserido em *E. coli* DH5 α para propagação. A amplificação do gene ocorreu de forma mais eficiente e sem bandas inespecíficas utilizando T_m entre 51 e 54 °C. Alguns clones obtidos foram analisados por PCR de colônias e um clone foi selecionado para sequenciamento, digestão e subclonagem do gene no vetor pET28a-TEV. O vetor pET28a-TEV foi digerido pelas enzimas *Nde*I e *Xho*I, ligado ao inserto contendo o gene *nahL* e inserido em *E. coli* BL21(DE3) para expressão.

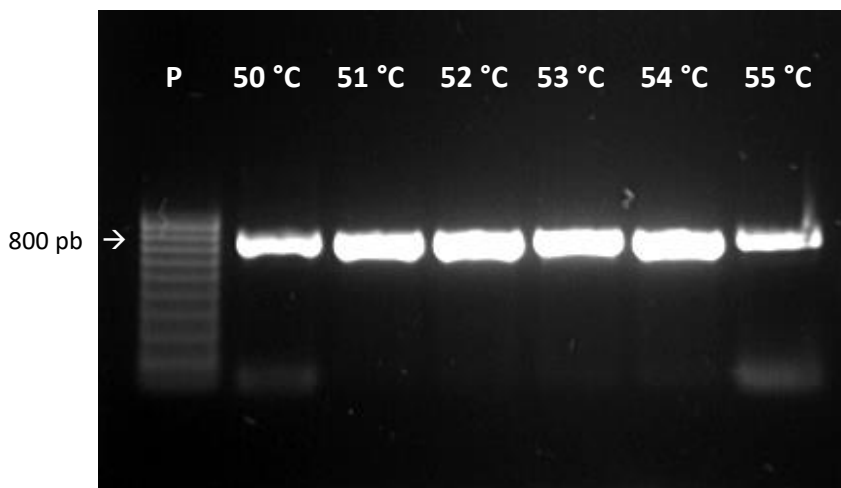


Figura 46. Amplificação do gene *nahL* de *P. putida* G7 para clonagem no vetor pET28a-TEV. P – padrão 100 pb EZ Load Molecular Ruler (Bio-Rad); demais canaletas – amplificações do gene *nahL* utilizando diferentes temperaturas de anelamento.

De acordo com o frame de leitura do vetor pET28a-TEV e os iniciadores utilizados para a amplificação do gene *nahL* a seguinte sequência de resíduos de aminoácidos foi expressa:

```
MGHHHHHENLYFQ/GHMDQILINELGDELYQAMLNREAIAPLTERGFAISIDDAYHISLRMLERRLAAGEKIIGKKIGVTSK  
AVQNMLGVYQPDFGYLTDAMVFNSGEAVPISQRLMQPKAEGEIAFILKDDLIGPGVTNADVLAATECVMPCFEIVDSRIRD  
WKIKIEDTVADNASCGLFVLGDQAVSPRQVDLVTGMLVEKNGQLLSTGAGAAALGSPVNCVAWLANTLGRFGIPLKAGE  
VILSGSLVPLEPVKAGDFMRVEIGGIGSASVRFT
```

A cauda de histidinas está demarcada em azul, o sítio de clivagem da protease TEV está destacado em verde. Nessa construção, após uma possível clivagem com a protease, restariam apenas dois resíduos de aminoácidos no N-terminal além da sequência nativa de NahL.

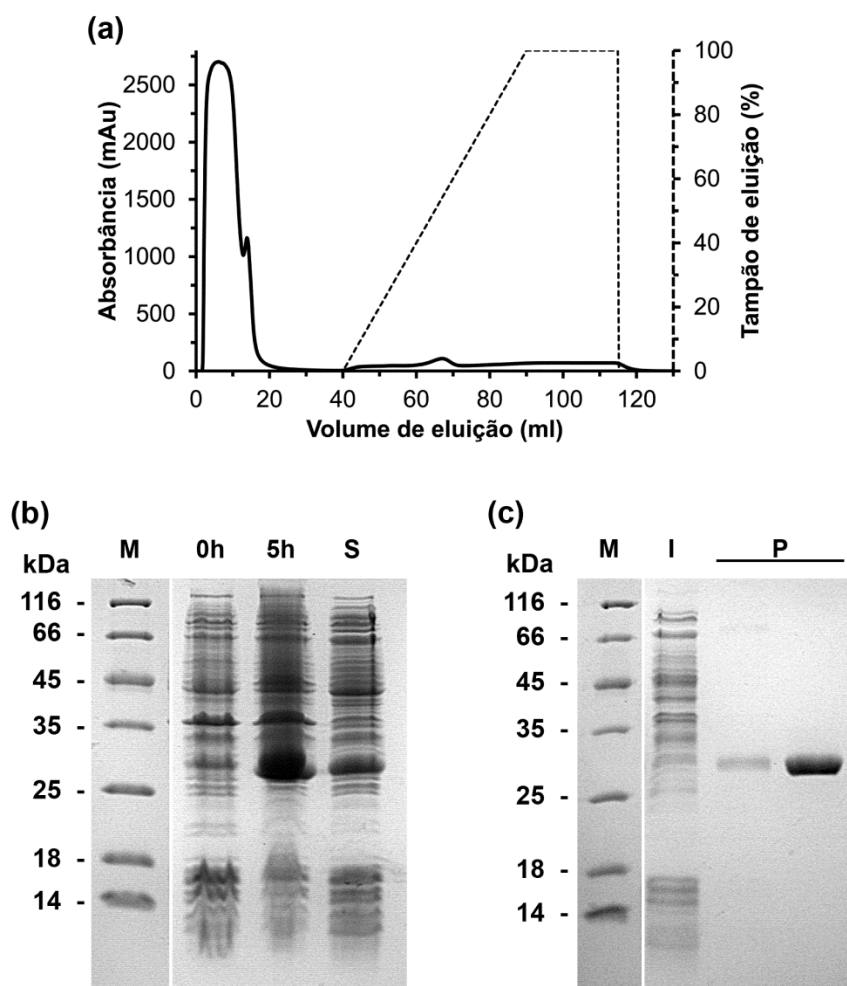


Figura 47. Cromatograma e análise por SDS-PAGE da purificação de 6xHis-NahL por IMAC.

(a) – Cromatograma da purificação de 6xHis-NahL por IMAC. A absorção foi medida a 280 nm.

(b) – SDS-PAGE (15%) da expressão de NahL. Canaleta **M**: Padrão de massa molecular (Fermentas); Canaleta **0h**: extrato proteico antes da indução por IPTG; Canaleta **5h**: extrato proteico após 5h de indução por IPTG 0,5 mM; Canaleta **S**: fração solúvel após lise e centrifugação das células.

(c) – SDS-PAGE (15%) da purificação de NahL por Cromatografia de afinidade ao níquel. Canaleta **M**: Padrão de massa molecular (Fermentas); Canaleta **I**: fração que não ficou retida na coluna durante a aplicação da amostra. Canaleta **P**: NahL recombinante obtida após eluição da coluna.

É possível observar uma banda de superexpressão de aproximadamente 28 kDa na canaleta correspondente ao tempo final de expressão (Figura 47 b, 5h) e que há uma banda de pequena intensidade na mesma altura na canaleta correspondente a fração solúvel (S)

indicando que NahL é pouco solúvel quando expressa a 30°C. A fração solúvel da expressão ainda assim foi purificada por uma etapa de IMAC e observou-se um pequeno pico durante a eluição entre os volumes 60 e 70 mL correspondente a fração da canaleta P (Figura 47 a). A concentração proteica dessa fração foi calculada e chegou-se a um valor correspondente a pouco mais de 1 mg total de proteína por litro de expressão.

Levando-se em consideração a perda de proteína em processos posteriores de clivagem, dessalinização, purificação e concentração de amostra, o pequeno rendimento obtido na expressão solúvel de NahL tornariam os experimentos mais demorados e onerosos devido a necessidade de expressão de vários litros de cultura e um baixo rendimento final. Além disso, foi reportado que a enzima XylJ (homóloga de NahL, Figura 48) também foi expressa predominantemente em forma insolúvel na ausência de Xyll, sendo possível sua expressão em forma solúvel em maior quantidade somente na forma de complexo 4-OD/VPH [65,95]. Portanto a estratégia de expressar NahL sozinha foi abandonada, e esforços foram direcionados para a expressão e o estudo do complexo NahK/NahL.

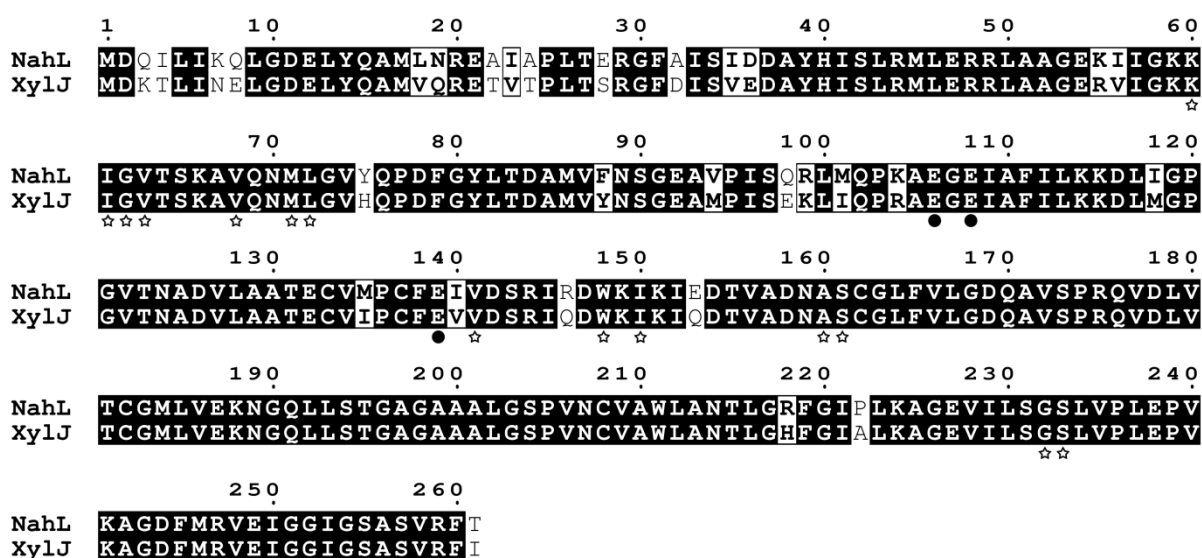


Figura 48: Alinhamento das sequências de resíduos de aminoácidos de NahL e XylJ. Os resíduos que coordenam o íon metálico no sítio ativo estão destacados por um círculo, enquanto que os outros resíduos que também compõem o sítio ativo estão destacados por estrelas.

4.10. Produção do complexo NahK/NahL

Para a construção do vetor para a coexpressão de NahK e NahL o gene *nahK* foi amplificado para clonagem no vetor pETDuet-1 utilizando os iniciadores nahK-F-pETd e nahK-R-pETd (Tabela 3), enquanto o gene *nahL* foi amplificado utilizando os iniciadores nahL-F-pET e nahL-R-pET (Figura 49). Após ligação do inserto de *nahL* no vetor pGEM-T, esse vetor foi digerido com as enzimas *NdeI* e *XhoI* e o inserto subclonado no segundo sítio múltiplo de clonagem do vetor pETDuet-1 previamente digerido com as mesmas enzimas. O vetor foi sequenciado para confirmar a correta inserção do gene *nahL* no respectivo sítio múltiplo de clonagem e para confirmar a correta sequência do gene. Após a confirmação da inserção de *nahL* neste vetor, o inserto de *nahK*, previamente ligado no vetor pGEM-T, foi removido desse vetor e inserido no primeiro sítio múltiplo de clonagem do vetor pETDuet-1 já contendo o gene *nahL*. Para isso o vetor pETDuet-1-nahL foi digerido utilizando as enzimas *BamHI* e *HindIII* e ligado ao inserto contendo o gene *nahK*.

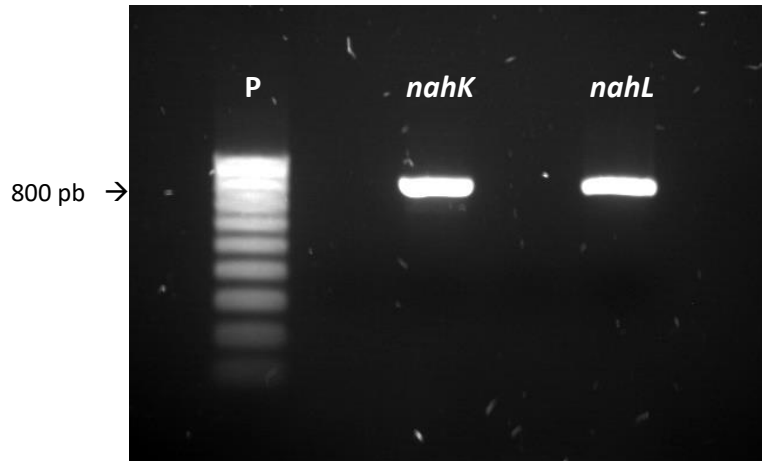


Figura 49: Amplificação do gene *nahK* e *nahL* de *P. putida* G7 para clonagem no vetor pETDuet-1. P – padrão 100 pb EZ Load Molecular Ruler (Bio-Rad); demais canaletas – amplificações do gene *nahK* e *nahL*.

Nessa construção, a expressão de NahK a partir do primeiro sítio múltiplo de clonagem do vetor gerou a seguinte sequência de aminoácidos:

MGSSHHHHHSQDPENLYFQ/GMNRTLREQVLALAEHIENAELNVHDIGKVTNDFPEMTFADAYDVQWEIRRRKEARG
NKIVGLKMGLTSWAKMAQMGVETPIYGFADYFVDPDGGVVDSCSLIHPKIEAEISVVTKAPLHGPGCHLGDVIAAIDYVIPT
VEIDSRYENFKFDLISVVADNASSTRFITGGRMASLEEVDLRTLGVVMEKNGEVVELGAGAAVLGHPLSSVAMLNLLAER
GEHIPAGTFIMTGGITAAPVAPGDNITVRYQGLGVSARFI

É importante destacar que a sequência correspondente à cauda de histidinas (destacada em azul) foi incorporada à proteína pelo vetor, enquanto que a sequência reconhecida e clivada pela protease TEV (destacada em verde) foi incorporada à sequência proteica através do iniciador direto utilizado para a amplificação do gene *nahK*, uma vez que o vetor pETDuet-1 não continha nenhuma sequência incorporada que permitisse a remoção da cauda de histidinas após expressão. Nessa construção, após a clivagem da cauda pela protease TEV, resta apenas um resíduo de glicina na porção N-terminal da proteína, deixando-a quase completamente nativa.

A expressão de NahL nessa construção a partir do segundo sítio múltiplo de clonagem gera uma sequência de aminoácidos totalmente nativa:

**MDQILINELGDELYQAMLNREAIAPLTERGFAISIDDAYHISLRMLERRLAAGEKIIGKKIGVTSKAVQNMLGVYQPDFGYLTD
AMVFNSGEAVPISQRLMQPKAEGEIAFILKKDLIGPGVTNADVLAATECVMPCFEIVDSRIRDWKIKIEDTVADNASCGLFVL
GDQAVSPRQVDLVTGMLVEKNGQLLSTGAGAAAALGSPVNCVAWLANTLGRFGIPLKAGEVILSGSLVPLEPVKAGDFMR
VEIGGIGSASVRFT**

Cada um dos dois sítios múltiplos de clonagem do vetor pETDuet-1 possui seu próprio promotor T7 para transcrição do gene para RNA mensageiro e seu próprio operador *lac* para controle dessa transcrição. Portanto, apesar dos dois genes estarem no mesmo vetor, a expressão de cada um ocorre de forma independente.

A expressão do complexo NahK/NahL foi realizada com sucesso utilizando o mesmo protocolo da expressão de NahK a 30 °C em *E. coli* BL21 (DE3) (Figura 50). A lise foi realizada em aparelho homogeneizador. A estratégia de purificação também foi a mesma de NahK, consistindo de dois passos de IMAC, intercaladas pela clivagem da cauda de histidinas pela adição de protease TEV à amostra. Entretanto, a estratégia adotada para a coexpressão e copurificação do complexo foi de produzir apenas NahK com a cauda de histidinas enquanto NahL é produzida em sua sequência nativa. Dessa forma, se NahL interagir com NahK ela também ficaria retida na coluna de cromatografia de afinidade.

Na Figura 50 é possível observar que, após a indução da expressão (canaleta 5h), há o aparecimento de duas bandas de superexpressão bastante próximas uma da outra (NahK com cauda tem uma massa molecular estimada de 30,9 kDa, enquanto NahL possui 27,9 kDa). O complexo também foi expresso de forma parcialmente solúvel.

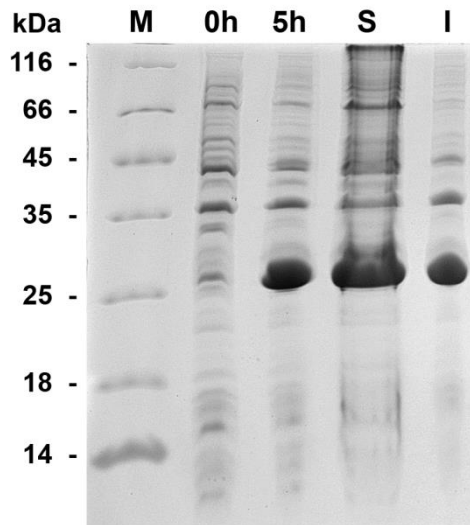


Figura 50: Análise por SDS-PAGE da coexpressão de NahK e NahL.

M – Padrão de massa molecular (Fermentas); **0h** – extrato proteico antes da indução por IPTG; **5h** – extrato proteico após 5h de indução por IPTG 0,5 mM; **S** e **I** – frações solúvel e insolúvel, respectivamente, após lise e centrifugação das células.

A cauda de histidinas presente em NahK permitiu a purificação parcial do complexo por IMAC em coluna HisTrap HP. O perfil cromatográfico típico dessa primeira etapa de purificação pode ser observado na Figura 51 (a). Amostras foram coletadas durante esta etapa de purificação e analisadas por SDS-PAGE (Figura 51 b). Durante a injeção na coluna da fração solúvel da expressão do complexo, proteínas que não se ligaram à resina passaram diretamente pela coluna e são observadas como um grande platô na curva de absorção (linha contínua) entre os volumes de eluição 0 e 65 mL (fração I). Após a lavagem para a remoção do restante das proteínas que não se ligaram à resina, que ocorreu entre os volumes de eluição 65 e 100 mL, há um gradiente crescente com tampão de eluição que contém imidazol em maior concentração (linha tracejada). Houve dois picos durante a eluição das proteínas da coluna. O primeiro pico observado entre os volumes de eluição 100 e 115 mL é em grande parte constituído por contaminantes que se ligaram à coluna durante a fase de injeção (Figura 51 b, fração P1). O segundo pico obtido entre os volumes de eluição 115 e 130 mL é constituído principalmente por duas bandas proteicas com as massas moleculares estimadas para NahK e NahL (Figura 51 b, fração P2).

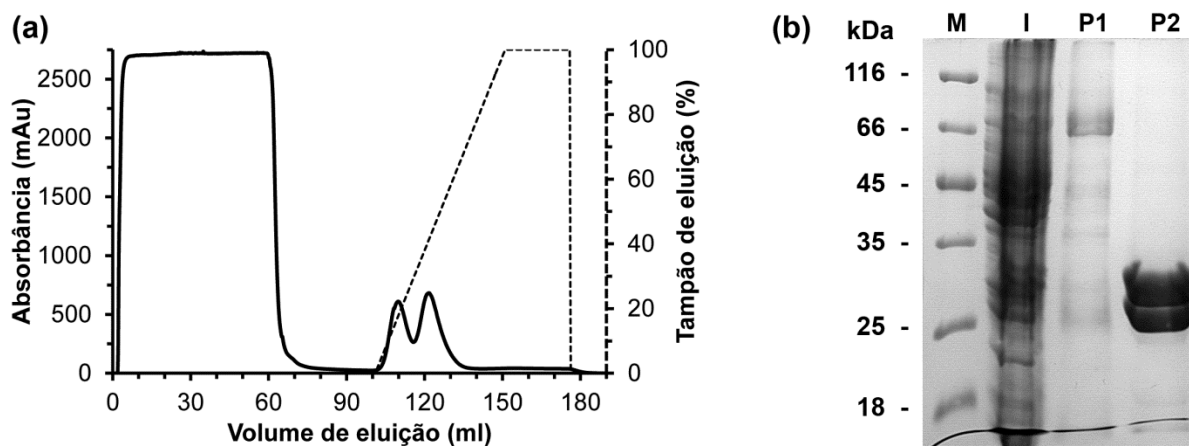


Figura 51: Cromatograma e análise por SDS-PAGE da primeira etapa de purificação do complexo NahK/NahL. **(a)** Cromatograma da primeira etapa de purificação do complexo por IMAC. A absorção foi medida a 280 nm. **(b)** SDS-PAGE da purificação do complexo. **M** – Padrão de massa molecular (Fermentas); **I** – Fração de proteínas que não se ligaram à coluna durante a injeção da amostra; **P1** – fração constituinte do primeiro pico obtido durante a eluição; **P2** – fração do segundo pico obtido durante a eluição.

Após a clivagem da cauda de histidinas pela protease TEV a amostra foi purificada novamente por IMAC. O perfil cromatográfico típico dessa segunda etapa de purificação do complexo pode ser observado na Figura 52 (a). Durante a injeção da amostra, entre os volumes de eluição 0 e 30 mL observou-se a eluição de um platô constituído principalmente por uma banda entre os marcadores de 25 e 30 kDa. Essa banda (Figura 52 b, fração C) contém NahK sem a cauda de histidinas (massa estimada de 28,5 kDa) e NahL (27,9 kDa). Nesse gel não foi possível separar as duas bandas devido à proximidade de suas massas moleculares e a quantidade de proteína presente na fração analisada. Para efeito de comparação, é possível observar a diferença de migração de NahK não clivada juntamente com NahL (Figura 52 b, fração ãC) e NahK clivada com NahL (Figura 52 b, fração C). Os pequenos picos obtidos durante o gradiente de eluição entre os volumes 80 e 110 mL são provavelmente de contaminantes e NahK não clivada, enquanto que o grande pico de eluição obtido entre os volumes 120 e 130 é provavelmente de protease TEV.

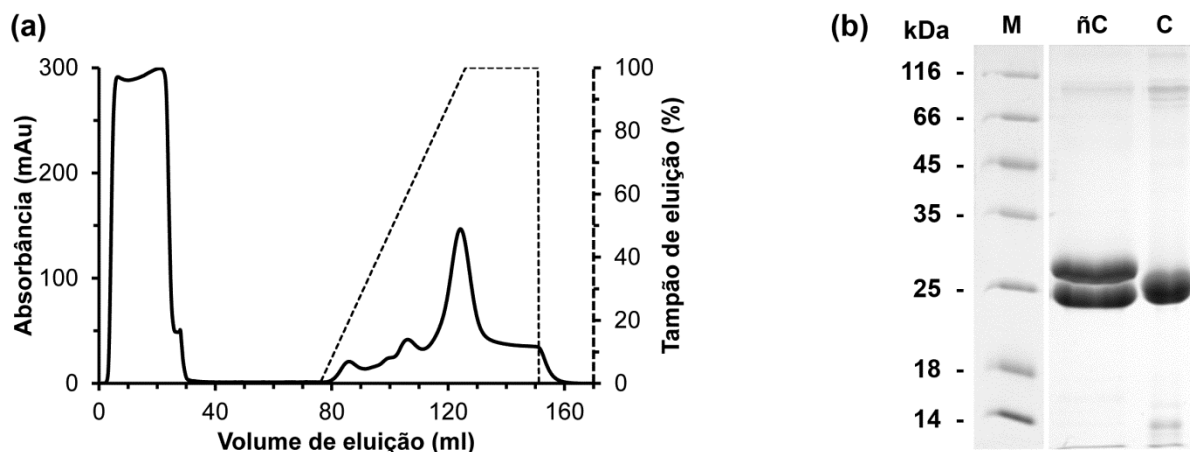


Figura 52: Cromatograma e análise por SDS-PAGE da segunda etapa de purificação do complexo por IMAC. **(a)** Cromatograma da segunda etapa de purificação do complexo NahK/NahL por IMAC. A absorção foi medida a 280 nm. **(b)** SDS-PAGE da purificação do complexo clivado pela protease TEV. **M** – Padrão de massa molecular (Fermentas); **ñC** – fração de complexo não clivado; **C** – fração obtida durante a injeção do complexo clivado na coluna (volumes de eluição entre 0 e 30 mL).

Após a segunda etapa de purificação do complexo NahK/NahL por IMAC, uma etapa final para polimento da purificação foi realizada por SEC. A Figura 53 (a) exibe o perfil cromatográfico típico obtido durante esta etapa de purificação. As frações constituintes dos dois picos foram analisadas por SDS-PAGE (Figura 53 b).

O primeiro e maior pico, obtido entre os volumes de eluição 70 e 80 mL, possui as duas proteínas constituintes do complexo, como pode ser observado pelas duas bandas na canaleta P1 da Figura 53 (b). Além disso, de acordo com o volume de eluição, estima-se que o complexo tenha uma massa molecular de aproximadamente 200 kDa. O segundo pico, obtido entre os volumes de eluição de 90 e 100 mL, contém apenas uma única banda proteica com massa estimada de 25 a 30 kDa, podendo ser um excesso de NahL de acordo com a altura dessa banda com relação a banda do complexo (NahK clivada ainda possui massa molecular maior que NahL). A separação do complexo dos monômeros de NahL foi importante para a etapa de cristalização, uma vez que não obtiveram-se cristais com a amostra purificada apenas por IMAC. Um litro de cultura rendeu aproximadamente 10 mg do complexo após todas etapas de purificação.

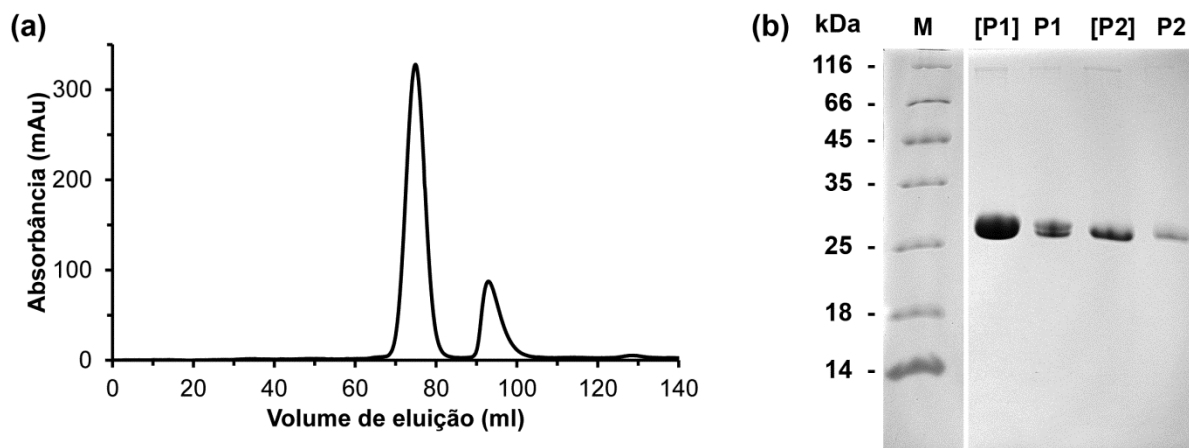


Figura 53: Cromatograma e análise por SDS-PAGE da etapa de purificação do complexo por SEC.

(a) Cromatograma da última etapa de purificação do complexo NahK/NahL por SEC. A absorção foi medida a 280 nm. **(b)** SDS-PAGE da purificação do complexo por SEC. **M** – Padrão de massa molecular (Fermentas); **[P1]** e **P1** – fração obtida durante a eluição do primeiro pico entre 70 e 80 mL aplicada no gel em duas concentrações diferentes, constituída principalmente pelo complexo; **[P2]** e **P2** – fração obtida durante a eluição do pico entre 90 e 100 mL, constituída provavelmente por um excesso de NahL e que não participa do complexo.

Como complexos homólogos de NahK/NahL já haviam sido superexpressos e purificados [64,65,95,109], era de se esperar que NahK interagiria com NahL, permitindo assim sua copurificação. Entretanto, nenhum dos protocolos de expressão previamente reportados utilizou uma cauda de histidinas como agente facilitador na etapa de purificação. Apesar disso, observamos que a cauda de histidinas, inserida no N-terminal de NahK, não foi um empecilho para a solubilidade desta proteína e nem atrapalhou na formação do complexo.

4.11. Estado oligomérico e estruturas secundárias do complexo

Através da técnica de DLS foi possível averiguar o estado de agregação do complexo NahK/NahL em solução. A Figura 54, obtida através de ensaio de DLS para o complexo após todas as etapas de purificação, apresenta a distribuição de tamanhos de partícula pela porcentagem de massa (linha contínua) e pela intensidade de espalhamento da luz (linha tracejada). Tanto a curva de massa como a de intensidade de espalhamento indicam que a amostra do complexo era composta por uma população de partículas com cerca de 10,4 nm de diâmetro e aproximadamente 16,6% de polidispersidade, indicando que a amostra estava

monodispersa e com uma alta pureza. Considerando esta partícula como uma proteína globular, a massa estimada é de aproximadamente 190 kDa. Este resultado indica que o complexo formado por NahK e NahL possuiria ao menos seis subunidades proteicas, já que cada monômero tanto de NahK quanto NahL possui aproximadamente 28 kDa. Entretanto a massa estimada continha um erro associado de ± 31 kDa, não sendo uma estimativa de massa tão precisa.

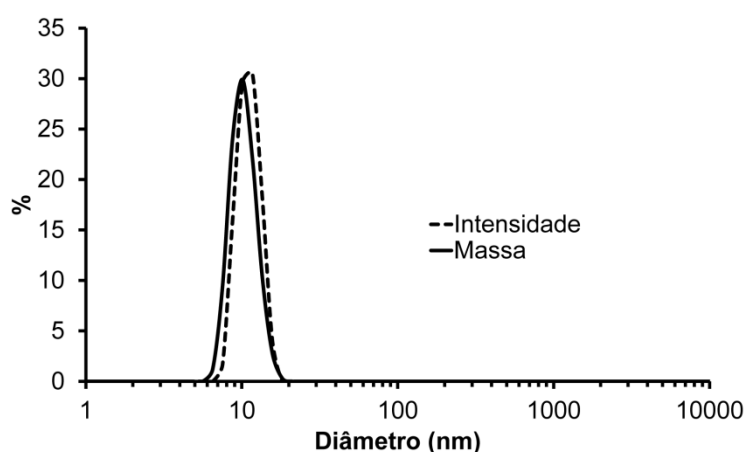


Figura 54: Distribuição de tamanhos de partícula do complexo recombinante NahK/NahL obtido por DLS.

A estrutura secundária do complexo e sua estabilidade térmica foram caracterizadas através de CD. O espectro de CD do complexo NahK/NahL obtido a 20 °C foi similar ao espectro de NahK, com duas bandas negativas em 222 nm e 210 nm e uma banda positiva em 195 nm (Figura 55 a). A deconvolução dos dados através do programa CDNN estima 29% de estruturas em hélices- α e 36% de estruturas em fitas- β (Figura 55 b). O complexo proteico mostrou-se bastante estável durante o aumento de temperatura entre 20 e 60 °C, quase não variando a porcentagem entre as estruturas ou alterando significativamente a forma do espectro. Entre 60 e 70°C ocorre uma transição abrupta, indicando o desenovelamento do complexo através de perda de estruturas em hélices- α e ganho de estruturas em fitas- β e estruturas aleatórias.

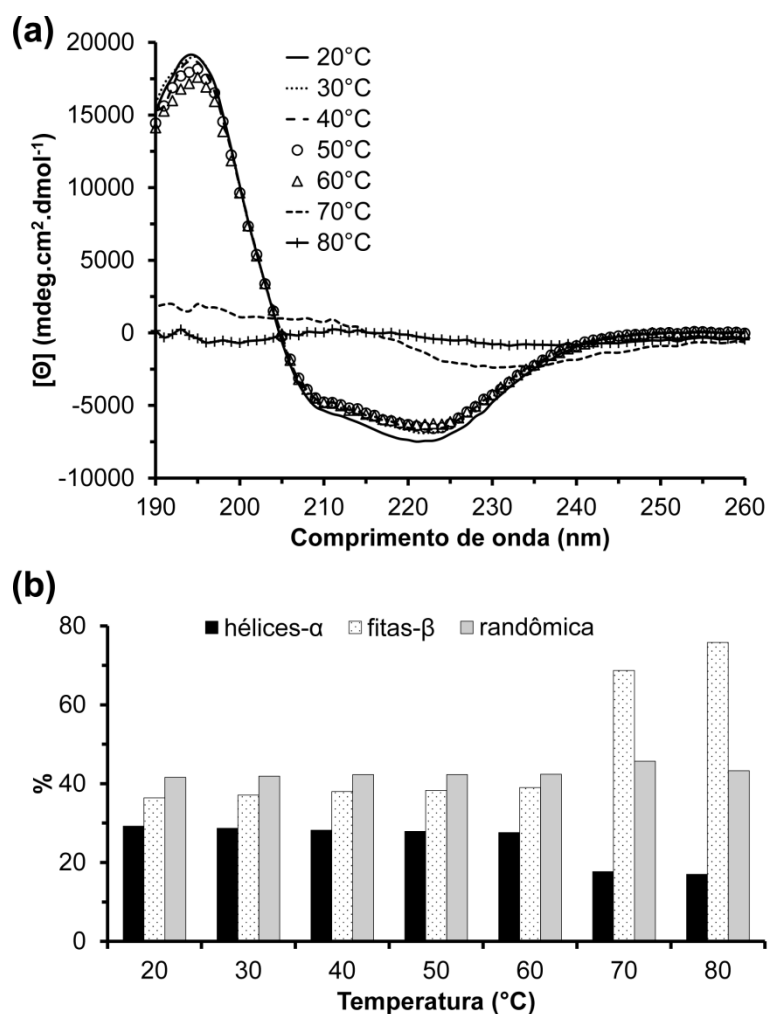


Figura 55: Dicroísmo circular do complexo NahK/NahL.

(a) Espectro de dicroísmo circular do complexo em diferentes temperaturas. **(b)** Porcentagem de estruturas secundárias de acordo com a temperatura.

4.12. Espalhamento de Raios-X a baixo ângulo

Experimentos de SAXS foram realizados a fim de calcular com maior precisão o tamanho do complexo formado por NahK e NahL, estimar sua massa molecular e o número de subunidades que o compõe. Essa técnica é bastante similar à XRD: uma amostra também é posicionada na frente de um feixe de raios-X e um detector é posicionado à frente da amostra. Entretanto o material analisado é uma amostra em solução (e não um cristal) e os dados coletados não são as intensidades de pontos de difrações, mas sim a intensidade de espalhamento dos raios-X em função de um vetor q . Partículas maiores tendem a espalhar

mais os raios-X, enquanto partículas menores tendem a espalhar menos. A curva de intensidade também sofre variações de acordo com o formato da partícula e a concentração desta na solução. Análises dessa curva de espalhamento permitem identificar o tamanho da partícula e até mesmo obter informações a respeito do formato da mesma e gerar um modelo de baixa resolução [81].

As curvas de SAXS obtidas para as diferentes concentrações do complexo NahK/NahL tiveram praticamente o mesmo formato. Entretanto, nas curvas de concentrações mais baixas há mais ruído nos valores mais altos de q , o que interfere na resolução final do modelo. Para a concentração de 6 mg/mL há menos ruído nessa porção dos dados, entretanto há uma diminuição na intensidade inicial nos valores mais baixos de q , o que interfere no valor calculado do raio de giro (Figura 56 a). Portanto, a curva obtida a 1,5 mg/mL foi normalizada com a curva a 3 mg/mL, e uma nova curva foi gerada a partir de suas médias. Essa curva foi denominada de *baixa resolução*. As quatro curvas obtidas para a concentração de 6 mg/mL foram praticamente idênticas, portanto foram normalizadas e unidas em uma única curva média para essa concentração, sendo essa curva denominada de *alta resolução*. A curva final foi gerada pela normalização e união dos pontos iniciais da curva *de baixa resolução* com os pontos finais da curva *de alta resolução*. Essa curva final (Figura 56 b) possibilitou um cálculo mais preciso do raio de giro, resultando no valor de aproximadamente 42 Å, juntamente com dados de maior resolução para construção do modelo tridimensional. Não foram observados danos pela radiação, nem nos tempos de exposição mais elevados, como de 10 minutos. A curva $P(r)$ é uma curva de distribuições de pares de distâncias calculada a partir da curva de intensidades de espalhamento, e, para o complexo NahK/NahL, essa curva exhibe distâncias máximas de aproximadamente 120 Å, valor aproximado do diâmetro estimado pelos experimentos de DLS (10,4 nm).

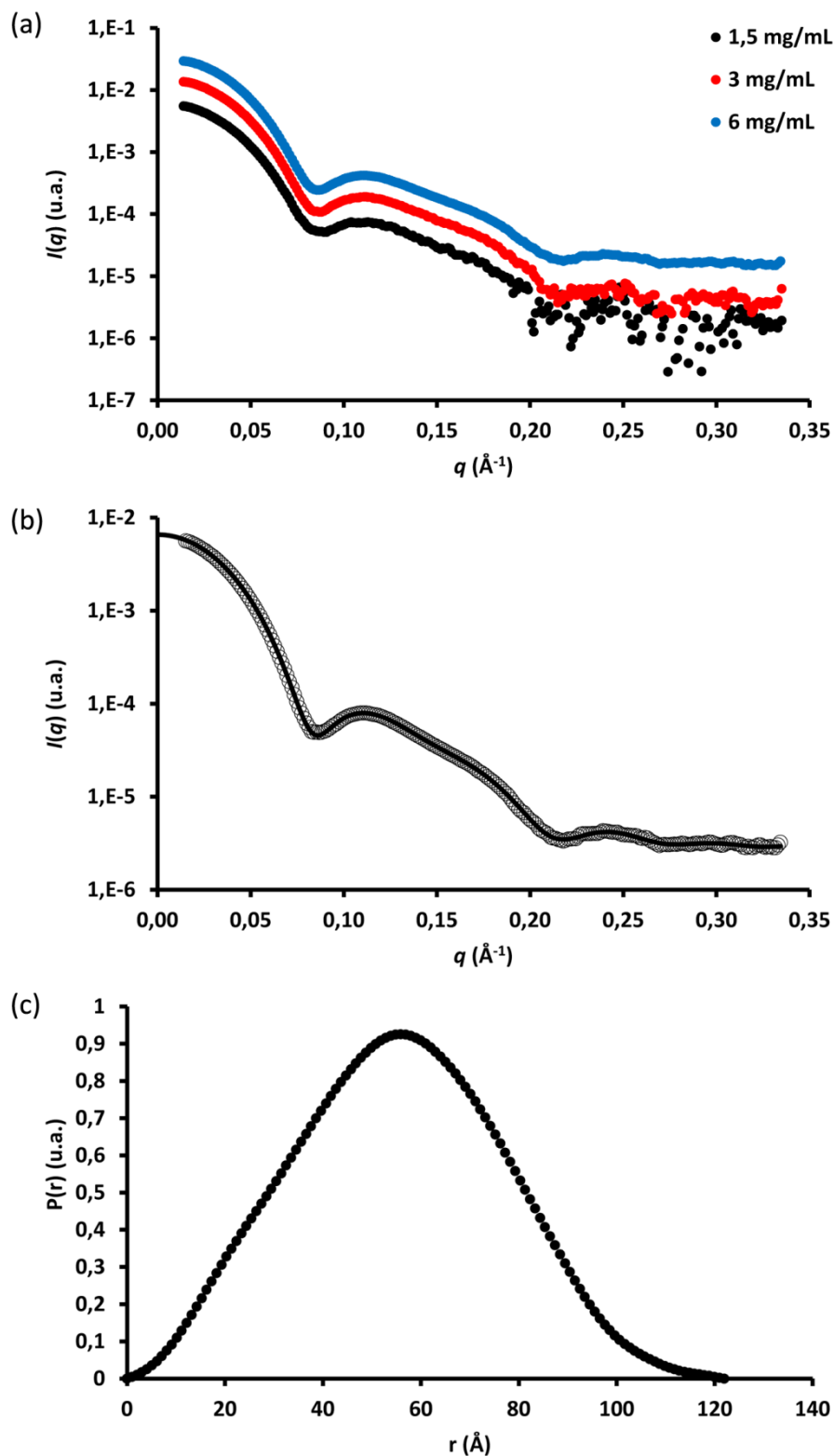


Figura 56: Curvas de SAXS e processamento de dados.

(a) curvas de espalhamento obtidas para diferentes concentrações do complexo NahK/NahL. **(b)** curva final utilizando dados de alta resolução da curva de concentração maior, e de baixa resolução da curva de concentração menor. **(c)** curva de distribuição de distâncias $P(r)$ calculada a partir da curva de espalhamento (b), em que se observa que a máxima distância do oligômero é em torno de 120 \AA .

A estimativa de massa molecular do complexo NahK/NahL utilizando a ferramenta SAXS MoW foi de 275 kDa \pm 2%. Considerando que as duas enzimas possuem uma massa aproximada de 28 kDa, sugere-se que o complexo formado por essas duas enzimas é constituído por 10 subunidades, possivelmente 5 de NahK e 5 de NahL.

Esse modelo de aproximadamente 280 kDa formado por um heterodecâmero de 4-OD e VPH nunca foi proposto para esse complexo proteico, entretanto, resultados que poderiam propor esse modelo foram primeiramente obtidos por Collinsworth *et al.* em 1973 [104]. Collinsworth e colegas estudaram um extrato proteico de uma espécie de *Pseudomonas putida* e isolaram um complexo nativo com atividade de vinilpiruvato hidratase. Através de estudos de centrifugação analítica foi calculado que esse complexo possuía 287 kDa de massa molecular, e, apesar de terem observado duas bandas proteicas em análises de SDS-PAGE (27,6 e 28,0 kDa), eles acreditaram que se tratavam de bandas derivadas da mesma enzima, sugerindo que seu complexo era composto por 10 subunidades.

Em 1989, Harayama e colegas estudaram um complexo proteico equivalente de *Pseudomonas putida* mt-2 formado pelas enzimas XylI e XylJ, que também apresentava atividade VPH, e que foi testado e confirmado para atividade de 4-OD. Também foi verificado por SDS-PAGE que esse complexo era formado por duas bandas proteicas (27 e 28 kDa). Embora já estivesse claro na época de que esse complexo era o mesmo estudado por Collinsworth, a informação obtida através da centrifugação analítica jamais foi explorada, e um valor de apenas 130 kDa foi sugerido por Harayama através de resultados de SEC [64]. No nosso trabalho observamos que os resultados obtidos por SEC e DLS subestimaram a massa molecular do complexo (200 e 190 kDa, respectivamente). Sendo assim, é provável que a sugestão de Harayama de 130 kDa para o complexo 4-OD/VPH também estivesse equivocada, e que o complexo de 287 kDa estudado por Collinsworth também possuía a enzima 4-OD, apenas não foi verificada a atividade de descarboxilação.

Para obter-se um modelo de baixa resolução do complexo foi utilizado o programa DAMMIN. Inicialmente foram gerados 12 modelos aleatórios sem imposição de simetria. Esses modelos foram processados no pacote de programas DAMAVER e foi gerado um modelo final exibido na Figura 57. Esse modelo possui um formato irregular, achatado e com uma cavidade central.

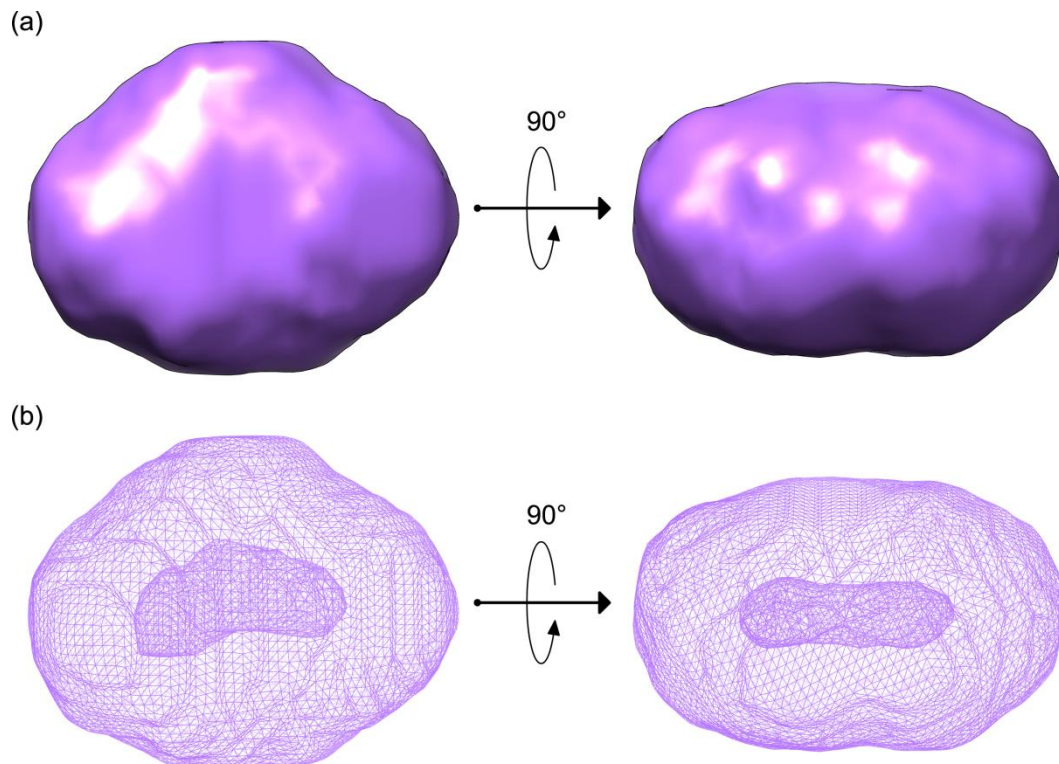


Figura 57: Modelo de baixa resolução obtido por SAXS sem imposição de simetria. **(a)** Modelo de superfície sólida. **(b)** Modelo de superfície em grade evidenciando a cavidade central.

Como foi aqui proposto que o complexo NahK/NahL possui 10 subunidades, e que ele poderia se oligomerizar assim como o complexo de HpcG, foi imposta no programa DAMMIN uma simetria 'P5', na qual existe um eixo de simetria com 5 'fatias' na partícula, sendo cada fatia idêntica às demais. Dessa forma, foram gerados e analisados 16 novos modelos. O modelo final gerado pelo pacote DAMAVER está apresentado na Figura 58. É possível observar que este modelo possui um canal central que atravessa a partícula do topo para a base no sentido do eixo de simetria. Também são visíveis cinco 'lóbulos' bem definidos se projetando da partícula na parte superior, que possui uma forma extremamente similar ao anel pentamérico de MhpD. A parte inferior está menos definida, mas é possível observar que a forma geral do complexo, segundo este modelo, é bastante similar ao homodecâmero de HpcG (Figura 30).

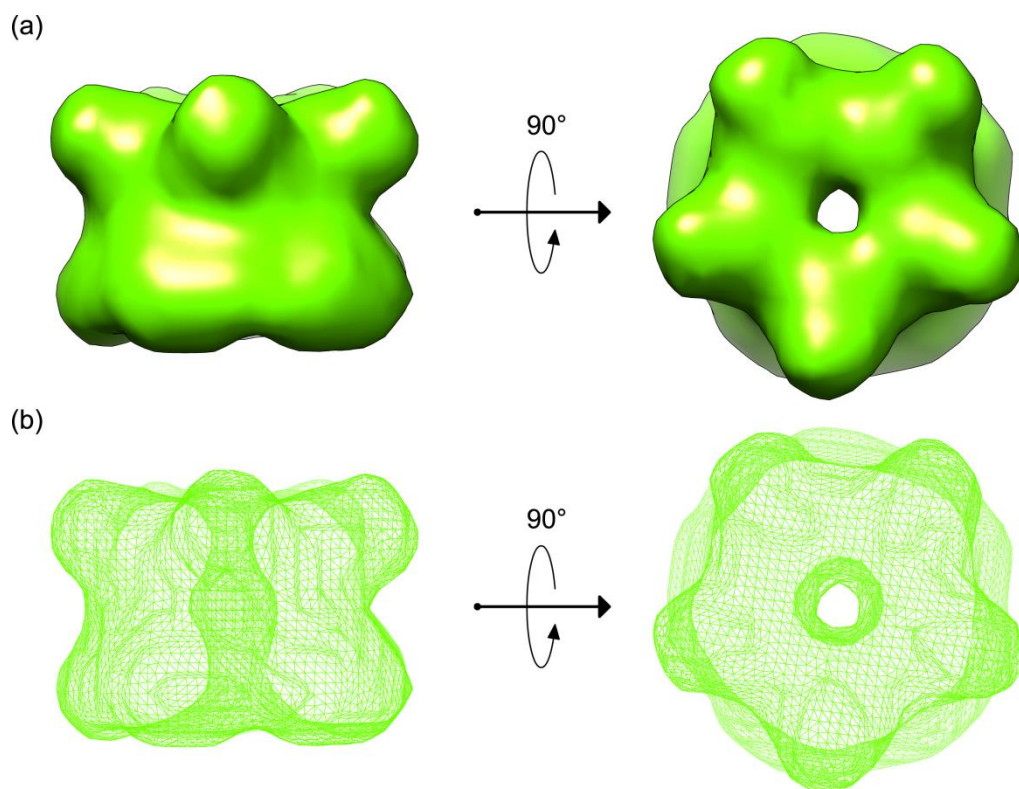


Figura 58: Modelo de baixa resolução obtido por SAXS com simetria imposta 'P5'.
(a) Modelo de superfície sólida. **(b)** Modelo de superfície em grade evidenciando o canal central.

Devido ao resultado promissor impondo a simetria 'P5', foram gerados outros 20 modelos com uma simetria 'P52', na qual existem dois eixos de simetria, um de 5 lados, e um de dois lados perpendicular ao primeiro eixo. O modelo final está representado na Figura 59. Nota-se que, devido à imposição de simetria, os cinco lóbulos superiores nesse modelo são iguais aos cinco lóbulos inferiores. Ainda há a presença de um canal atravessando o centro da estrutura ao longo do primeiro eixo de simetria.

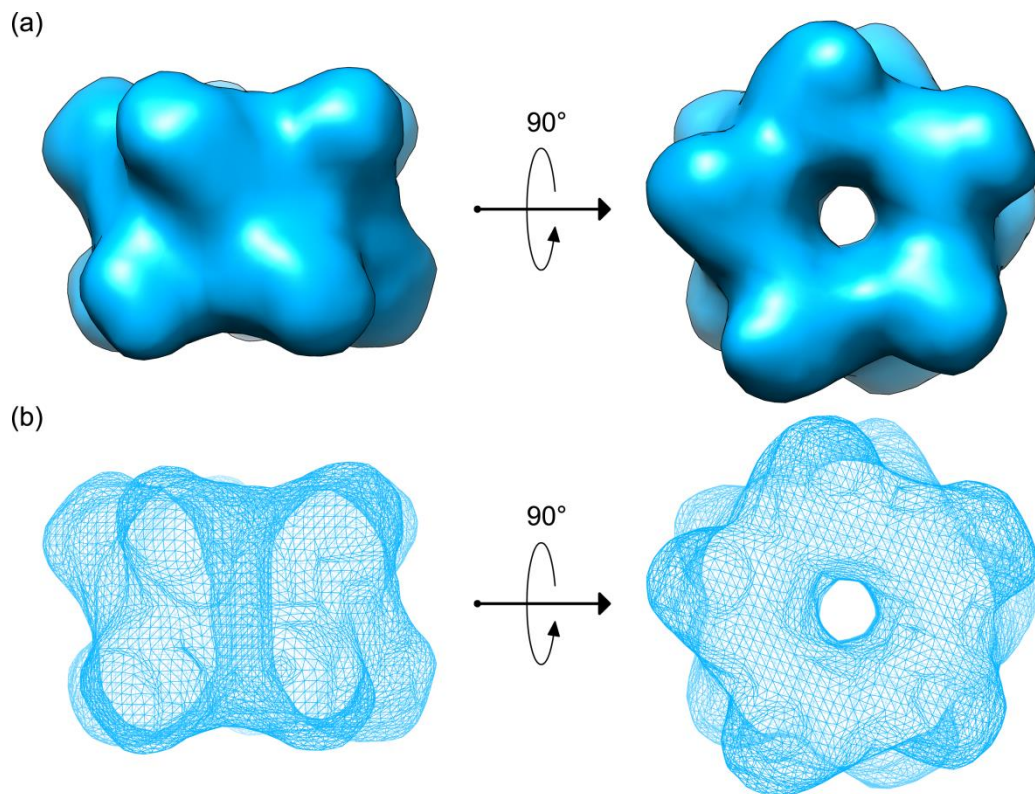


Figura 59: Modelo de baixa resolução obtido por SAXS com simetria imposta 'P52'.
(a) Modelo de superfície sólida. **(b)** Modelo de superfície em grade evidenciando o canal central.

Com esse resultado foi possível fazer a perfeita superposição do homodécâmero cristalográfico de HpcG com o modelo de baixa resolução do complexo NahK/NahL obtido por SAXS. A superposição está exibida na Figura 60. Dessa forma, propõe-se que NahK e NahL de fato se oligomerizam da mesma maneira que os monômeros de HpcG, formando um complexo de 10 subunidades.

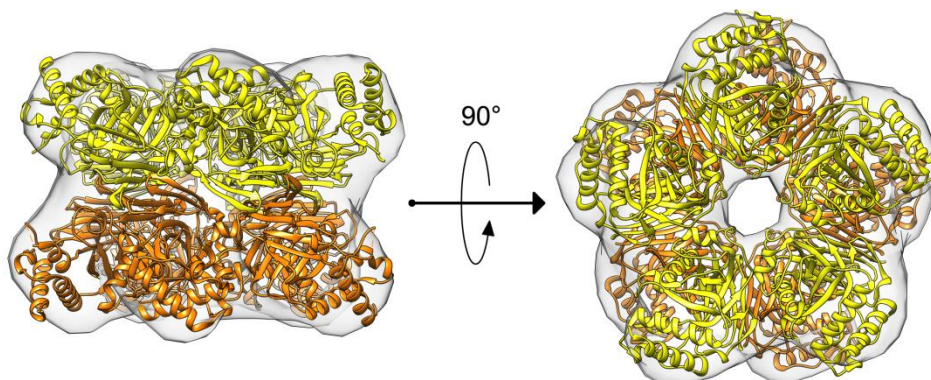


Figura 60: Superposição do modelo cristalográfico do homodécâmero de HpcG e o modelo de baixa resolução do complexo NahK/NahL. O primeiro anel pentamérico está colorido em amarelo, enquanto o segundo anel está em alaranjado. O modelo de superfície do complexo NahK/NahL está em cinza.

4.13. Atividade enzimática de NahK na forma de complexo

Os parâmetros cinéticos de NahK na forma de complexo foram obtidos seguindo o mesmo protocolo para NahK em sua forma isolada, a diferença é que foi acompanhado a diminuição da concentração de substrato a 236 nm ao invés do acúmulo de produto a 265 nm. O substrato 4-oxalocrotonato foi previamente preparado pela mistura de 2HM e a enzima 4-OT. A Figura 61 exibe o comportamento da curva de absorção ao longo da reação de descarboxilação de 4-oxalocrotonato pela enzima NahK na forma de complexo com NahL. É possível observar que há o decréscimo da absorbância a 236 nm por consequência do consumo do substrato, entretanto, não há o surgimento de absorbância a 265 nm pelo acúmulo do produto como na reação de NahK na ausência de NahL (Figura 22). Isso se explica pelo fato de que na forma de complexo NahL consome rapidamente o produto de NahK, impossibilitando o seu acúmulo.

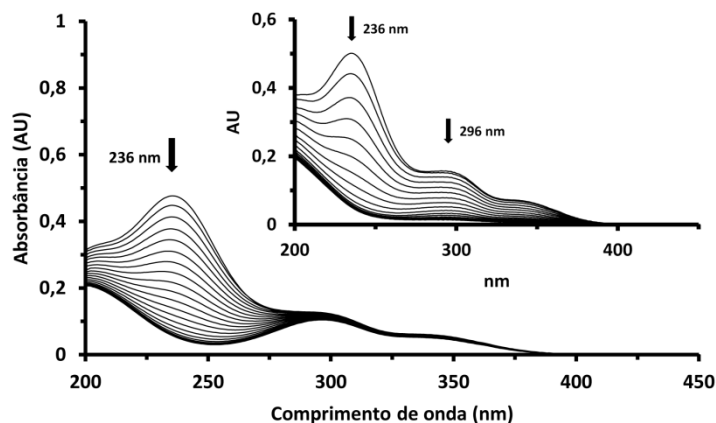


Figura 61: Curvas de absorvância da reação da enzima NahK ao descarboxilar 4-oxalocrotonato na presença de NahL.

A figura inserida na parte superior é para a mesma reação, mas na presença de quantidades superiores da enzima 4-OT.

Como a concentração final de 4-OT na reação é bastante baixa, ainda é possível observar o pico residual de 2HM a 296 nm. Entretanto, se for adicionada a enzima 4-OT em quantidade suficiente na reação para que os 20% restantes de 2HM sejam convertidos rapidamente à 4-oxalocrotonato a medida que este é consumido por NahK, observa-se a diminuição da absorvância a 296 nm juntamente com a diminuição da absorvância a 236 nm (Figura 61, gráfico inserido na parte superior direita). Como o produto de NahL não possui absorvância significativa acima de 250 nm, a curva vai se achatando.

A atividade de NahK na forma de complexo (a 0,3 µg/mL) foi avaliada inicialmente em pH 7,4 a temperatura ambiente em três diferentes concentrações de cloreto de magnésio. A Figura 62 mostra o gráfico obtido a partir dos valores da velocidade inicial da reação em função da concentração de substrato para diferentes concentrações de magnésio. A absorvância foi medida a 236 nm de acordo com o consumo do substrato. A Tabela 7 exhibe os dados de cinética obtidos desse experimento.

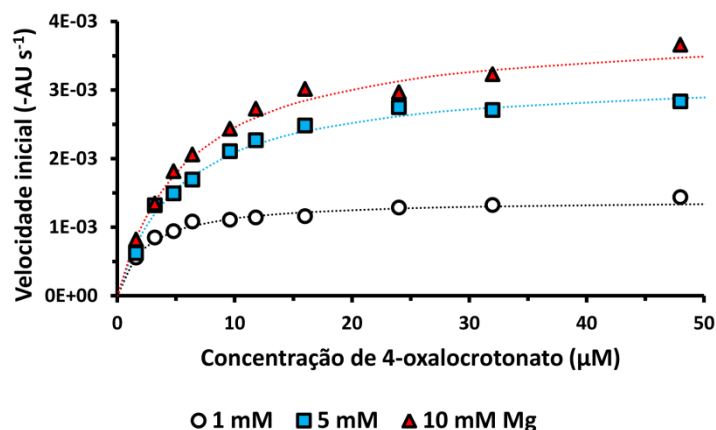


Figura 62: Velocidades iniciais da reação de NahK na presença de NahL em diferentes concentrações de substrato e de Mg^{2+} .

Tabela 14: Parâmetros de cinética para NahK na presença de NahL utilizando-se diferentes concentrações de magnésio.

[Mg^{2+}] (mM)	V_{max} (-AU s ⁻¹)	K_M (μM)	k_{cat} (s ⁻¹)	k_{cat}/K_M (M ⁻¹ s ⁻¹)
1	0,0014 ± 0,00003	2,42 ± 0,26	40,00 ± 0,86	1,6 × 10 ⁷
5	0,0032 ± 0,0001	5,35 ± 0,42	92,29 ± 2,02	1,7 × 10 ⁷
10	0,0039 ± 0,0001	5,93 ± 0,49	111,41 ± 2,86	1,9 × 10 ⁷

Observamos que, para a atividade de NahK na presença de NahL, a concentração de magnésio também interfere na velocidade máxima da reação e na afinidade da enzima pelo substrato, mas não altera significativamente a eficiência da reação, já que todos os três valores de k_{cat}/K_M foram bastante semelhantes. Em comparação com o mesmo ensaio para a atividade de NahK na ausência de NahL (Figura 23 e Tabela 8), percebemos que, devido à presença de NahL e a possível estabilização que esta fornece à NahK por estarem complexadas em sua forma natural, há um aumento de uma ordem de grandeza na eficiência da reação devido um aumento considerável em k_{cat} . Os valores de K_M permaneceram semelhantes. Dessa forma, para efeitos de melhor comparação com NahK na ausência de NahL, adotou-se a concentração de Mg^{2+} 5 mM para realização dos demais ensaios de cinética.

A atividade da enzima NahK na presença de NahL (a 0,3 μg/mL) também foi testada em valores de pH variando entre 6 e 8 em temperatura ambiente (Figura 63). Os parâmetros de cinética podem ser visualizados na Tabela 15.

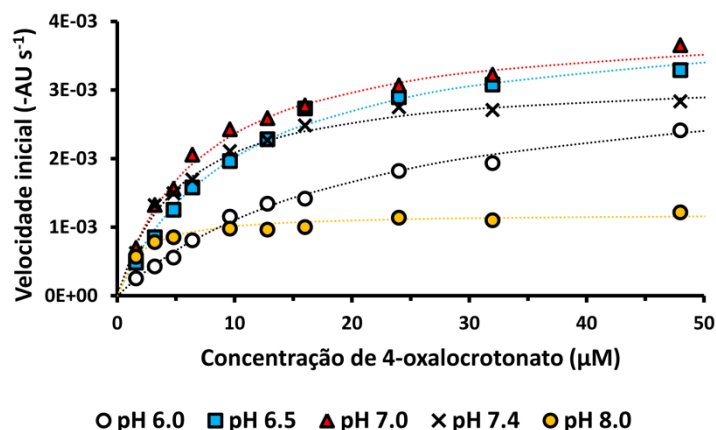


Figura 63: Influência do pH na cinética de NahK na presença de NahL.

Tabela 15: Parâmetros de cinética para NahK na ausência de NahL utilizando-se diferentes valores de pHs em tampão fosfato de sódio.

pH	V_{\max} (-AU s ⁻¹)	K_M (μM)	k_{cat} (s ⁻¹)	k_{cat}/K_M (M ⁻¹ s ⁻¹)
6,0	0,0034 ± 0,0002	20,79 ± 1,99	97,13 ± 5,71	4,7 × 10 ⁶
6,5	0,0041 ± 0,0001	10,33 ± 0,91	117,13 ± 2,86	1,1 × 10 ⁷
7,0	0,0040 ± 0,0001	6,92 ± 0,49	114,27 ± 2,57	1,7 × 10 ⁷
7,4	0,0032 ± 0,0001	5,35 ± 0,42	91,42 ± 2,00	1,7 × 10 ⁷
8,0	0,0012 ± 0,00002	1,84 ± 0,23	34,28 ± 0,57	1,9 × 10 ⁷

Assim como no experimento em diferentes concentrações de magnésio, provavelmente devido ao fato de NahK estar em sua forma natural complexada com NahL, percebe-se um incremento de uma ordem de grandeza na eficiência da reação em todos os valores de pH quando comparado com os resultados de NahK sozinha sob as mesmas circunstâncias (Figura 24 e Tabela 9). No geral, percebe-se uma eficiência mais limitada em pH 6,0 e um aumento dessa eficiência em pH 6,5. Os maiores valores de eficiência foram em torno do pH 7,4. Quando comparado com o experimento de NahK sozinha, os valores de k_{cat} são bastante semelhantes (exceto para pH 6,0), entretanto há uma considerável diminuição nos valores de K_M , evidenciando que a enzima aumenta sua afinidade pelo substrato, elevando assim sua eficiência.

Quando observamos o resultado cinético de XylI complexada com XylJ em pH 7,4 obtido por Stanley em 2000 [65], verificamos os seguintes parâmetros cinéticos: $K_M = 6,6 \pm 0,6 \mu\text{M}$, $k_{\text{cat}} = 68 \pm 2 \text{ s}^{-1}$, $k_{\text{cat}}/K_M = 1,0 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$. Portanto, NahK complexada com NahL apresenta parâmetros extremamente similares aos de 4-OD/VPH do plasmídeo pWWO,

pode-se sugerir ainda que a eficiência da reação do complexo NahK/NahL é ligeiramente maior que o do complexo XylI/XylJ devido um valor ligeiramente mais baixo de K_M e um valor mais elevado de k_{cat} .

4.14. Atividade enzimática de NahL na forma de complexo

A atividade de NahL na forma de complexo (a 0,3 $\mu\text{g/mL}$) foi medida de forma semelhante à NahK, entretanto, para isso foi utilizado o substrato natural de NahL, produzido no laboratório do prof. Dr. Christian P. Whitman, e observada a diminuição da absorbância do substrato a 265 nm. Um esquema geral do comportamento das curvas de absorção da reação de NahL na presença de NahK está representado na Figura 64. Como pode ser observado, há um decréscimo na absorbância a 265 nm, que representa a diminuição da concentração do substrato. Como o produto de NahL não possui absorbância significativa acima de 250 nm, o que se nota é um achatamento da curva ao longo do tempo. Como a concentração de magnésio de 5 mM foi adotada em todos os experimentos prévios, e como a concentração de 10 mM não alterou significativamente a eficiência de NahK, esse valor foi adotado para os ensaios com NahL.

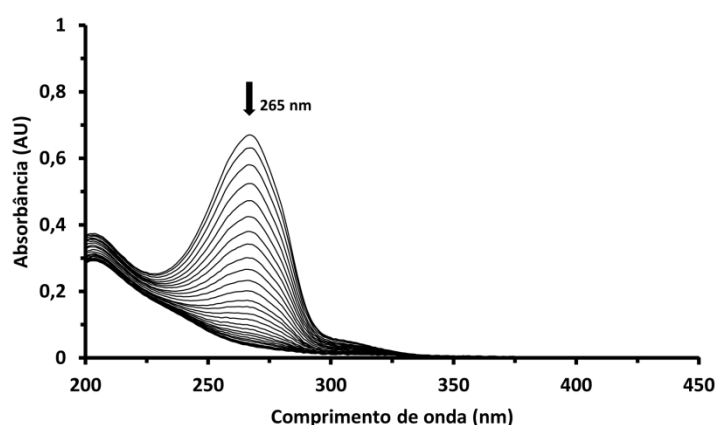


Figura 64: Curvas de absorbância da reação da enzima NahL ao hidratar o substrato 2-hidroxi-2,4-pentadienoato.

A atividade da enzima NahL na presença de NahK em pHs diferentes está exibida na Figura 65. Os parâmetros de cinética podem ser visualizados na Tabela 16.

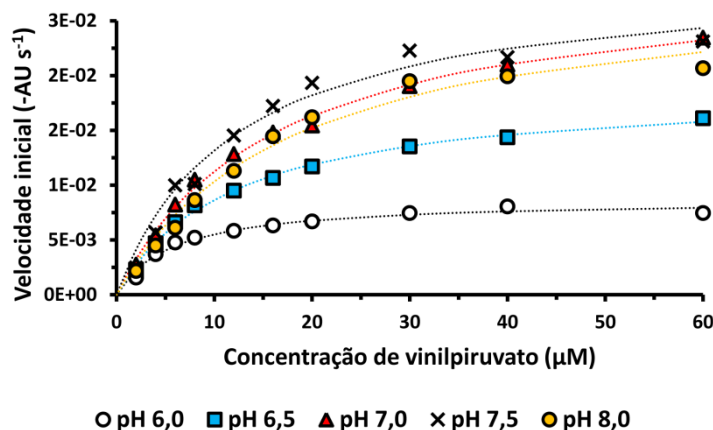


Figura 65: Influência do pH na cinética de NahL na presença de NahK.

Tabela 16: Parâmetros de cinética para NahL na presença de NahK utilizando-se diferentes valores de pHs em tampão fosfato de sódio.

pH	V_{\max} (-AU s ⁻¹)	K_M (μM)	k_{cat} (s ⁻¹)	k_{cat}/K_M (M ⁻¹ s ⁻¹)
6,0	0,0087 ± 0,0003	5,78 ± 0,77	170,76 ± 5,89	3,0 × 10 ⁷
6,5	0,0189 ± 0,0004	11,84 ± 0,69	370,96 ± 7,86	3,1 × 10 ⁷
7,0	0,0295 ± 0,0008	16,18 ± 1,08	579,01 ± 15,70	3,6 × 10 ⁷
7,4	0,0293 ± 0,0018	12,20 ± 2,01	575,01 ± 35,33	4,7 × 10 ⁷
8,0	0,0287 ± 0,0018	17,71 ± 2,67	563,31 ± 35,33	3,2 × 10 ⁷

Como podemos observar, o valor de pH altera muito pouco a eficiência de NahL, sendo que todos os valores de k_{cat}/K_M foram bastante aproximados e da ordem de 10⁷ M⁻¹ s⁻¹. Há variação nas velocidades máximas entre as reações, o que implica diretamente nos valores respectivos de k_{cat} , entretanto os valores de K_M também variam na mesma proporção, mantendo a eficiência da enzima semelhante nos diversos pHs. Apesar disso, nota-se um valor bastante elevado de k_{cat} em pH 7,0 e 7,4, sendo que nesses pHs a eficiência da enzima foi ligeiramente maior quando comparado com os demais. A V_{\max} em pH 6,0 é a mais limitada, e esse valor de pH pode ser importante para definir quais resíduos do sítio ativo estão intimamente relacionados à catalise.

Em comparação com estudos feitos por Johnson em 2010 [95], nos quais foram testadas a cinética de XylJ do complexo Xyll/XylJ em pH 7,3, e cujos parâmetros cinéticos obtidos foram $K_M = 11 \pm 2 \mu\text{M}$, $k_{\text{cat}} = 360 \pm 20 \text{ s}^{-1}$, $k_{\text{cat}}^{\text{cat}}/K_M = 3,3 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$, notamos a

grande similaridade com os parâmetros de NahL na forma de complexo, e até mesmo uma eficiência catalítica ligeiramente mais elevada devido um valor mais alto de k_{cat} .

Nota-se assim que o complexo aumenta em uma ordem de grandeza a eficiência de NahK quando comparado com NahK sozinha, e que a reação da enzima NahL possui valores de k_{cat} e eficiência mais elevados que de NahK. Como consequência, a hidratação do produto de NahK ocorre mais rapidamente do que a reação de descarboxilação, impedindo que o intermediário instável se acumule.

4.15. Cristalização do complexo NahK/NahL

Os ensaios de cristalização do complexo NahK/NahL foram realizados inicialmente a partir das 96 condições dos kits Crystal Screen 1 e 2 (Hampton Research) utilizando amostra proteica purificada apenas por dois passos de IMAC. Nenhum cristal foi obtido. Posteriormente, adotou-se o passo de purificação por SEC, o que permitiu a separação do complexo de monômeros (provavelmente moléculas de NahL que se encontravam em excesso e não participavam da oligomerização). Essa amostra do complexo foi enviada ao ROBO LAB do Laboratório Nacional de Biociências – LNBio, em Campinas-SP, onde mais de 500 condições de cristalização foram testadas. Duas condições promissoras nas quais houve o crescimento de cristais de proteína ao longo de 4 meses foram selecionadas e amplificadas manualmente. Em uma das condições (MIB-PEG – malonato de sódio, imidazol e ácido bórico, pH 5,5 a 6,0 e PEG 1500 10 a 30% v/v) houve o crescimento de monocristais que difrataram no máximo em torno de 6 a 7 Å de resolução. Na segunda condição (MMT-PEG – ácido málico 20 mM, MES 40 mM, TRIS 40 mM, pH 5,5 a 7,0 e PEG 1500 20 a 27% v/v), houve o crescimento de pequenas agulhas que, após etapas de *seeding*, permitiu o crescimento de cristais com cerca de 250 µm de diâmetro que difrataram até 2,49 Å de resolução (Figura 66). Um desses cristais gerou, até o momento, um único conjunto de dados para o complexo NahK/NahL (Tabela 17).

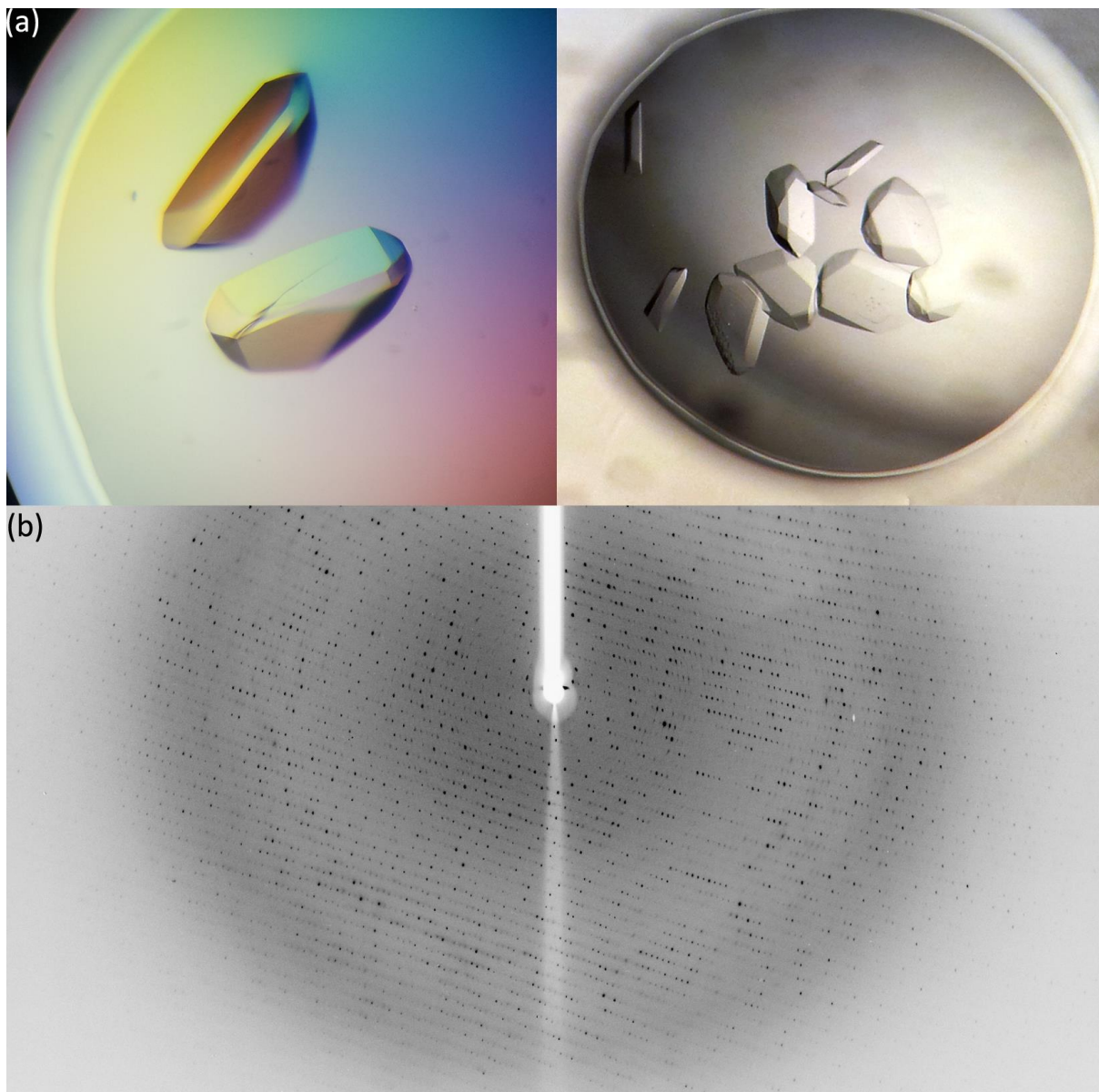


Figura 66: Cristais (a) e padrão de difração de raios-X (b) do complexo NahK/NahL.

Tabela 17: Sumário das estatísticas do conjunto de dados obtido a partir de um cristal do complexo NahK/NahL.

Os valores entre parênteses são para as zonas de maior resolução.

Conjunto de dados	7 (modelo KL)
PDB ID	-
Linha	D01B-MX2
Detector	MarMosaic 225
Comprimento de onda (Å)	1,459
Distância detector (mm)	167,0
Rotação por imagem (°)	0,1
Rotação total (°)	110
Grupo espacial	$P2_12_12_1$
	$a = 100,407$
	$b = 219,962$
	$c = 310,363$
	$\alpha = \beta = \gamma = 90$
Parâmetros de célula unitária (Å, °)	
Mosaicidade (°)	0,376
Resolução (Å)	47,66 – 2,49 (2,53 – 2,49)
No. de reflexões totais	985.382
No. de reflexões únicas	234.835
Redundância	4,2 (2,8)
Completeza (%)	97,8 (81,5)
[$I / \sigma(I)$]	13,77 (3,42)
R_{merge}^a (%)	0,074 (0,30)
Coefficiente de Matthews (Å ³ Da ⁻¹)	2,88
Conteúdo de Solvente (%)	57,39
R fator*	0,195
R_{free}^*	0,230
No. de átomos de proteína*	38.913
No. de moléculas de água*	1164
B fatores:	
átomos de proteína*	
Geral*	
rmsd:	
ângulos das ligações*	1,894
comprimento das ligações*	0,013

^a $R_{\text{merge}} = \sum_{hkl} \sum_i [|I_i(hkl) - \langle I(hkl) \rangle|] / \sum_{hkl} \sum_i I_i(hkl)$, onde $I_i(hkl)$ é a intensidade medida da reflexão hkl e $\langle I(hkl) \rangle$ é sua média.

* Parâmetros temporários já que o modelo do complexo NahK/NahL encontra-se em etapa de refinamento.

O cristal obtido a partir da condição MMT-PEG pertence ao grupo espacial $P2_12_12_1$ com parâmetros de célula unitária com valores de aproximadamente $a = 100,4$; $b = 220,0$; $c = 310,4$ Å; e ângulos $\alpha = \beta = \gamma = 90^\circ$ (Figura 67). Devido a célula unitária com essas dimensões há um grande número de famílias de planos cristalinos, e, como consequência, um grande número de pontos de difração, exigindo a rotação do cristal de apenas $0,1^\circ$ por imagem para evitar a superposição dos pontos. Ao todo foram coletadas 1100 imagens, totalizando cerca de 10 horas de coleta na linha D01B-MX2. A coleta de dados a partir desse cristal alcançou o limite operacional da linha, já que a rotação mínima permitida pelo goniômetro era de $0,1^\circ$ e

o tempo disponível de feixe é cerca de 11 horas. Com a construção do novo acelerador de elétrons brasileiro, o Sírius, espera-se que difrações de raios-X de cristais desse tipo serão coletadas em um tempo menor.

Análise do valor do Coeficiente de Matthews sugeriu a presença de cerca de 25 moléculas de proteína por unidade assimétrica. A substituição molecular foi feita com sucesso utilizando um monômero de NahK. Ao todo foram encontradas 20 posições para cadeias proteicas na unidade assimétrica, formando dois heterodecâmeros. No total são 80 cadeias na célula unitária (oito heterodecâmeros, Figura 67).

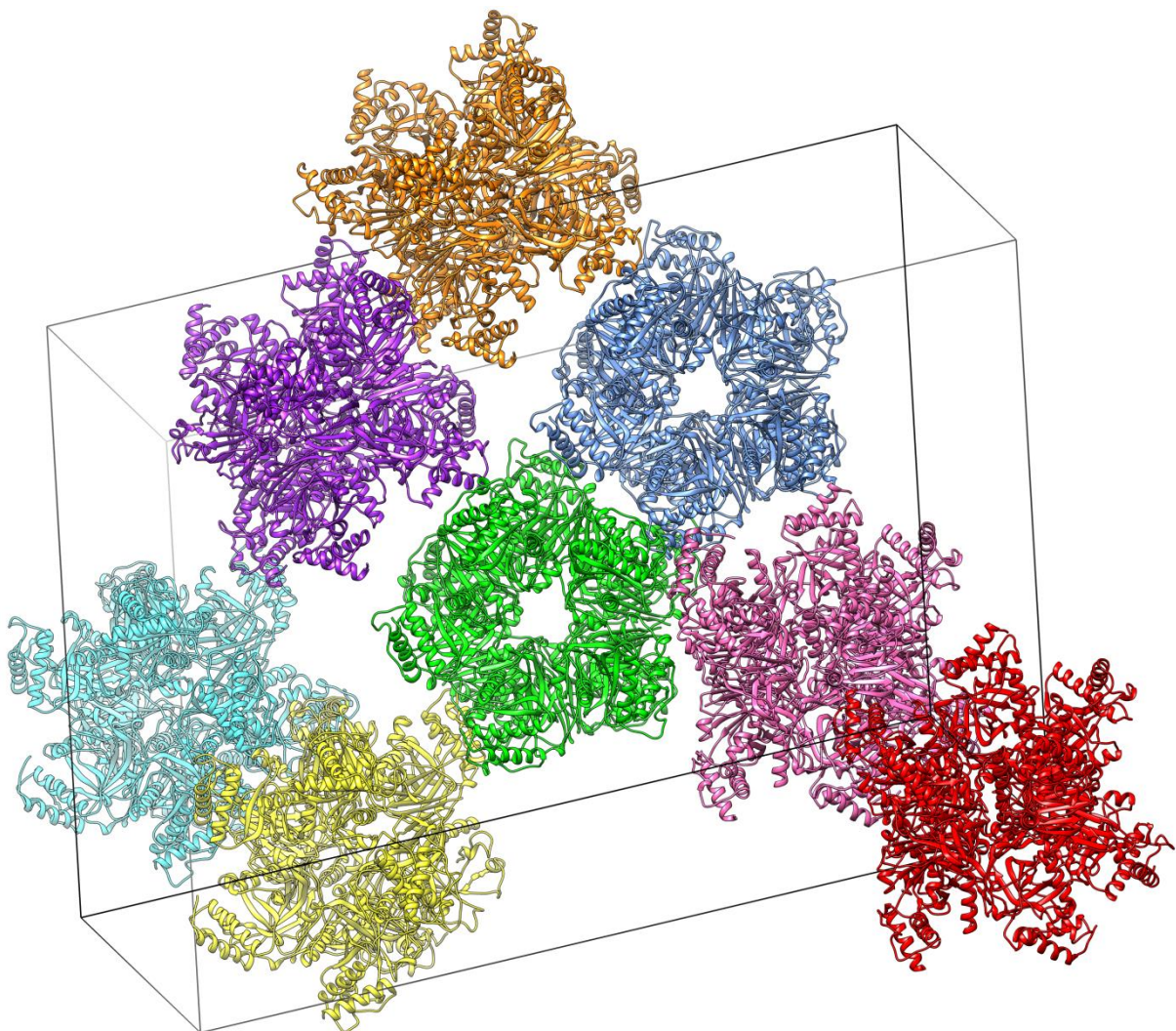


Figura 67: Representação dos oito heterodecâmeros de NahK e NahL empacotados na célula unitária.

O modelo do complexo NahK/NahL, até a data da defesa desta tese, ainda se encontra em etapa de refinamento, entretanto é clara a posição ocupada pelas 10 moléculas de proteína formadoras de cada um dos 2 complexos presentes na unidade assimétrica. Os parâmetros correspondentes à geometria do modelo (ângulos e distâncias das ligações covalentes) já se encontram em valores aceitáveis e comumente observados para os modelos depositados no PDB, e os parâmetros R fator e R_{free} , que significam o quão bem o modelo proposto explica os dados experimentais, já estão em valores médios para a resolução alcançada. No total existem aproximadamente 5250 resíduos de aminoácidos na unidade assimétrica, desses, 93,6% se encontram em valores de angulação permitidos segundo o gráfico de Ramachandran, 4,96% estão em regiões permitidas, e apenas 1,44% se encontram em regiões não favoráveis. Portanto, apesar de ainda haver melhorias a serem feitas no modelo, ele já pode ser considerado como um modelo bastante representativo do complexo formado por essas duas enzimas, e será descrito a seguir.

4.16. Análise geral da estrutura de NahL

A cadeia 'B' do modelo do complexo trata-se de uma molécula de NahL com o menor valor de B-fator dentre as outras cadeias da mesma enzima, e não possui problemas relacionados à geometria apontados durante o refinamento parcial. Portanto, essa cadeia será utilizada a seguir para o breve detalhamento das estruturas secundária e terciária de NahL.

NahL também faz parte da superfamília FAH. Além disso, como descrito para NahK e também observado para HpcG e MhpD, a enzima NahL possui um domínio N-terminal formado por hélices e pequenas voltas que também se posicionam sobre a lateral do barril- β incompleto (Figura 68), e se diferencia dos outros membros dessa superfamília por não formar dímeros (Figura 29). Em NahL o DNT é composto pelos primeiros 54 resíduos, sendo este ligeiramente mais curto em sequência do que o DNT de NahK com 58 resíduos. O servidor STRIDE identificou apenas duas hélices- α compondo o NTD, sendo que a área do barril escondida por esse domínio em NahL é de aproximadamente 1156 Å² (Tabela 13).

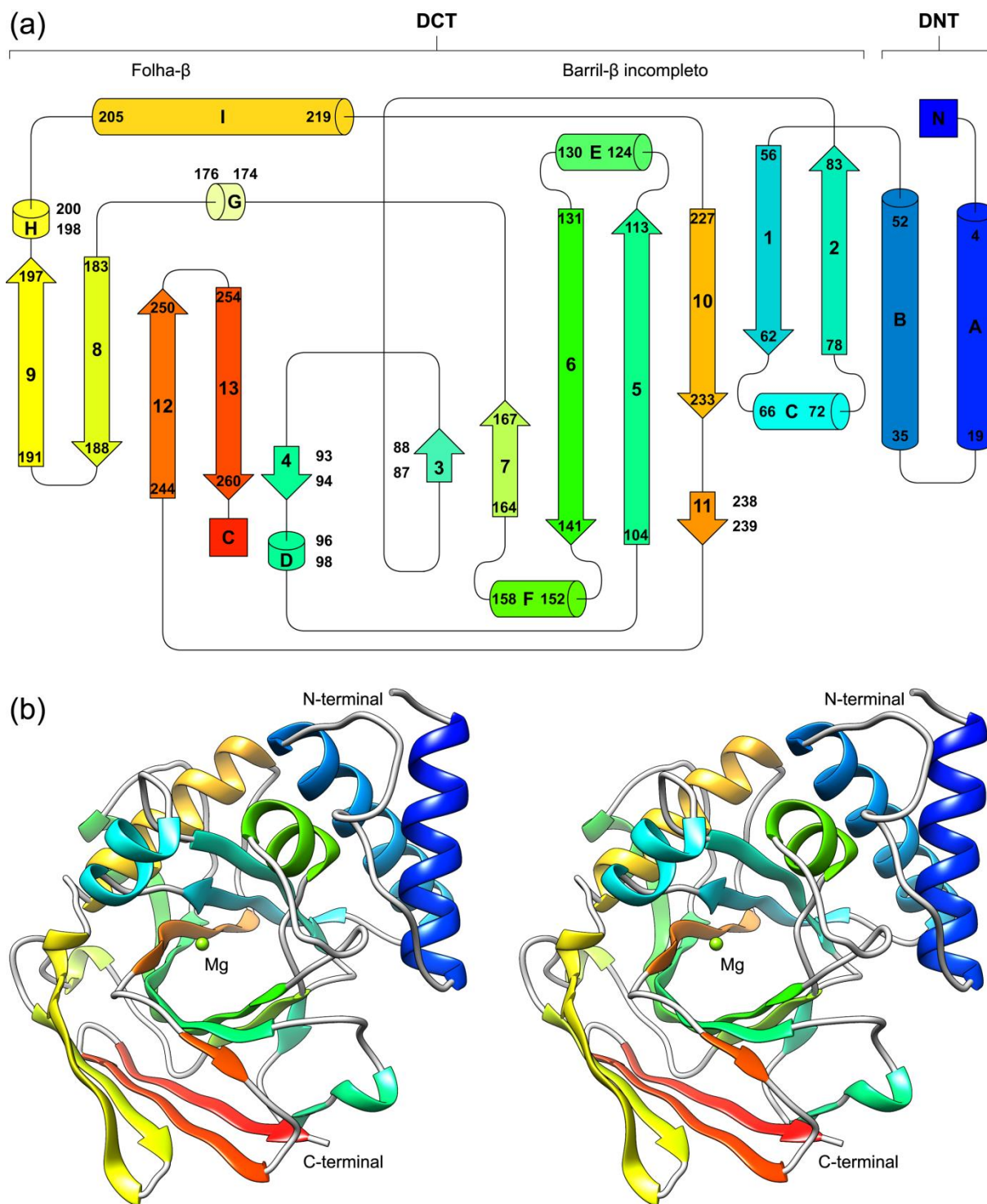


Figura 68: Representação das estruturas secundárias e terciária de NahL.

(a) Diagrama topológico de NahL: hélices- α são exibidas como tubos e nomeadas alfabeticamente; fitas- β são exibidas como setas e nomeadas numericamente. A transição de cores de azul a vermelho representa o progresso da cadeia de N a C-terminal. (b) Representação em estéreo (*cross-eyed*) da cadeia principal de NahL ligada ao magnésio, a representação de cores é a mesma de (a). O íon magnésio (esfera verde) ligado à proteína se encontra no topo da estrutura em barril- β . O diagrama topológico foi gerado com ajuda da plataforma on-line Pro-origami [96].

O domínio do barril também é composto por 8 fitas- β paralelas e antiparalelas interligadas por pequenas voltas e hélices. No geral, para NahL, o servidor STRIDE identificou menos elementos estruturais em hélices-3₁₀ do que para NahK. A ligação ao metal dá-se pelos resíduos Glu106, Glu108 e Glu139. Também se observa a folha- β C-terminal formada por 5 fitas (Figura 68). O volume da cavidade do sítio ativo em NahL, calculada pelo servidor CASTp, foi de aproximadamente 446 Å³. Essa cavidade apresenta-se menos volumosa que a de NahK, mesmo com a hélice que provavelmente forma o domínio tampa em NahL estando em posição 'aberta'. O resíduo Lys60 (correspondente ao resíduo Lys64 em NahK) também faz parte da cavidade do sítio ativo, interagindo com o Glu139 através de uma ponte salina de 2,5 Å de distância. Os demais resíduos catalíticos Lys72, Met76, Phe151, Phe153 e Ser164 detalhados em NahK são substituídos em NahL, respectivamente, pelos resíduos Val68, Met71, Trp148, Ile150 e Ser161 e também podem ser responsáveis pela interação com o substrato e/ou catálise.

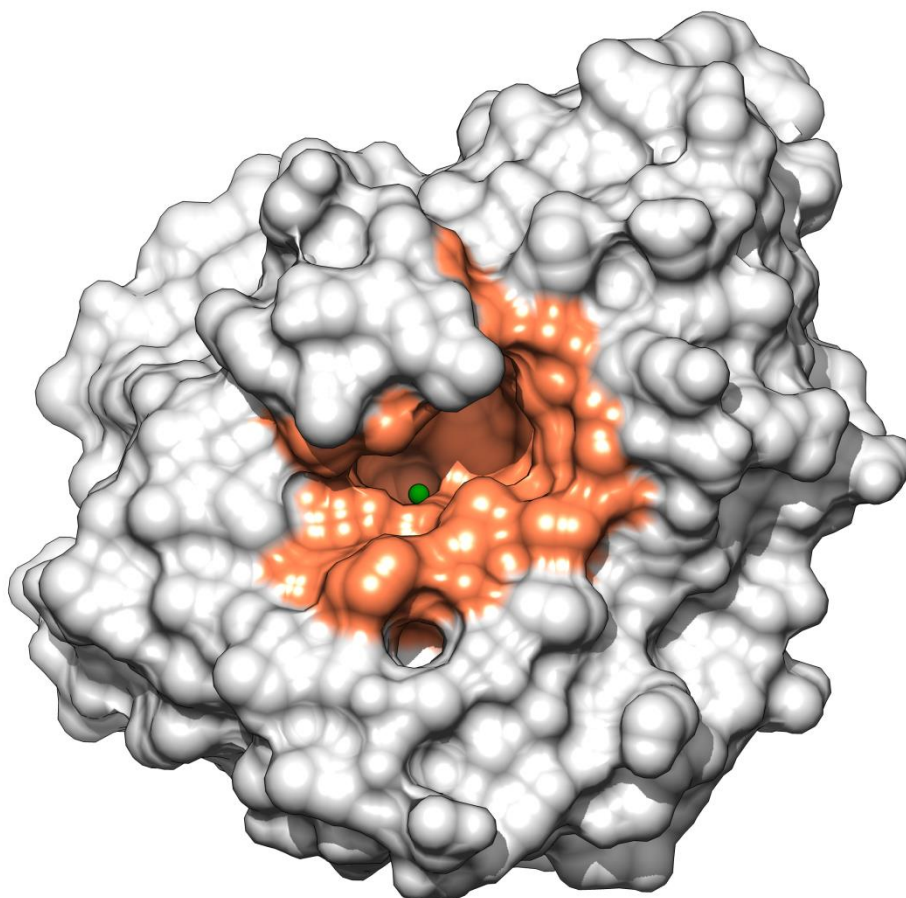


Figura 69: Modelo de superfície de NahL exibindo em destaque a cavidade do sítio ativo.

Mesmo sem a obtenção de estruturas cristalinas de NahL complexada com análogos de substrato, é provável que esta enzima também possua um domínio tampa, assim como proposto para NahK e outros membros da superfamília FAH [68,69]. É possível observar na Figura 70 que a hélice C de NahL, juntamente com a volta localizada entre a fita 6 e a hélice F possui valores de B-fator mais elevados quando comparado com o restante dos elementos do domínio do barril, sugerindo uma maior mobilidade dessas regiões.

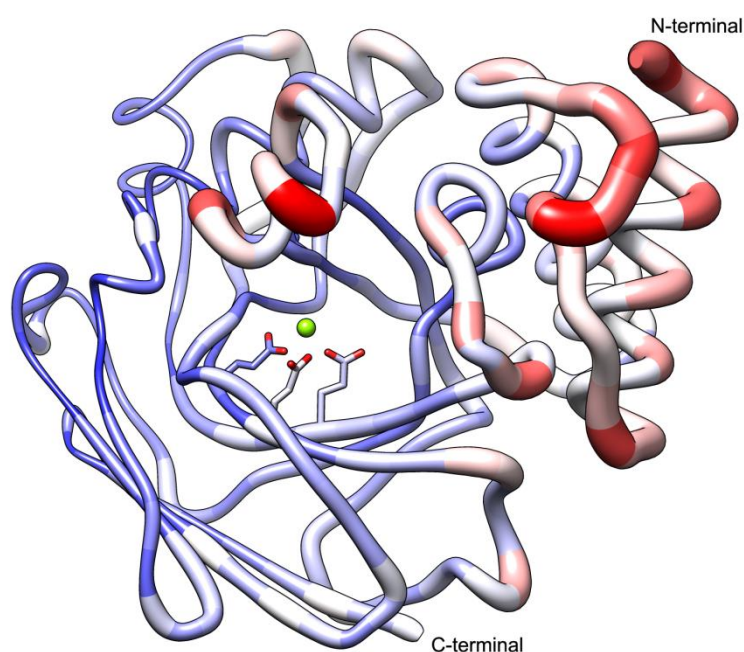


Figura 70: Representação em 'worm' do modelo de NahL.

Tanto as tonalidades de cores de azul (B-fator = 20,2) a vermelho (B-fator = 68,1), como a espessura do tubo, representam o B-fator médio dos resíduos de aminoácidos, evidenciando regiões com maior mobilidade térmica na proteína.

Mecanismos de ação para VPH e homólogas vêm sendo propostos desde 1973 [104,112,113], sendo que o produto da enzima já foi identificado como a molécula L-(S)-4-hidroxi-2-oxopentanoato (Figura 40 6c), e que a reação se inicia através de uma etapa de tautomerização do substrato de uma forma enol para a forma de cetona, seguida pelo ataque de uma molécula de água no C-4. Entretanto, até recentemente, ainda não se sabia quais resíduos poderiam ser responsáveis por essa reação de hidratação. Em 2008, Miller utilizou de análise de alinhamento de sequências de enzimas homólogas à XylJ juntamente com a análise da estrutura cristalina de MhpD (homóloga de NahL e XylJ) para propor que os resíduos Lys60, Asp78, Ser161 e Leu72 poderiam estar diretamente envolvidos na catálise

[114]. Foram então realizadas mutações sítio-dirigidas em XylJ e complexos mutantes de XylI/XylJ foram expressos, purificados e parâmetros cinéticos foram obtidos (Tabela 18).

Tabela 18: Parâmetros cinéticos de XylJ e mutantes correspondentes.

Enzima	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
XylI/XylJ nativa	13 ± 3	240 ± 23	$2,0 \times 10^7$
E109Q_XylI/K60A_XylJ	35 ± 4	$0,14 \pm 0,007$	$4,0 \times 10^3$
E109Q_XylI/D78N_XylJ	20 ± 6	$0,094 \pm 0,001$	$4,8 \times 10^3$
XylI/S161A_XylJ	7 ± 2	$11 \pm 0,9$	$1,7 \times 10^6$
XylI/L72A_XylJ	13 ± 2	33 ± 1	$2,4 \times 10^6$

Dados obtidos por Miller, 2008 [114].

Pode-se observar pelos resultados obtidos por Miller, que a mutação do resíduo Asp78 para Asn não alterou significativamente o K_M da reação, indicando que esse resíduo não teria uma função direta de ligação com o substrato. Apesar disso, essa mutação diminuiu em pelo menos 1300 vezes o valor de k_{cat} , e em quatro ordens de grandeza a eficiência da enzima, evidenciando que este é um resíduo essencial para a catálise. O resíduo Asp78 foi definido como um provável candidato para a ativação de uma molécula de água para o ataque nucleofílico no C4. Um efeito similar nos parâmetros cinéticos aconteceu com a mutação do resíduo Lys60 para Ala, entretanto, nesse caso, o valor de K_M aumentou em cerca de duas vezes juntamente com a diminuição do valor de k_{cat} , sugerindo tanto um papel de ligação ao substrato como de catálise para este resíduo. Miller sugere que a Lys60 possui papel essencial na tautomerização da forma enol do substrato para a forma cetona do intermediário. A mutação de Ser161 para Ala aumentou a afinidade da enzima pelo substrato provavelmente pelo aumento da hidrofobicidade do bolsão do sítio ativo, promovendo uma interação mais forte com a parte apolar do substrato (carbono 5). Entretanto, também houve uma diminuição no valor de k_{cat} , diminuindo assim a eficiência da enzima em uma ordem de grandeza, indicando também uma possível função catalítica desse resíduo. A mutação do resíduo Leu72 não alterou a afinidade da enzima pelo substrato, mas diminuiu o valor de k_{cat} , resultando também na diminuição da eficiência em uma ordem de grandeza. Miller não sugeriu um papel catalítico para esse resíduo, entretanto, através da comparação da estrutura cristalina de NahK e NahL, observa-se que esse resíduo faz parte da hélice C, que, por sua vez, provavelmente funciona como domínio tampa em NahL. Sendo assim, a mutação do resíduo Leu72 poderia interferir na abertura e fechamento da hélice C sobre a cavidade do sítio ativo, alterando os parâmetros cinéticos da reação.

Com base nesse estudo, Miller propôs o mecanismo de reação exibido na Figura 71. Nesse mecanismo o substrato se ligaria monodentalmente ao metal e haveria a presença de diversas moléculas de água na cavidade do sítio ativo, as quais seriam ativadas por resíduos específicos e participariam da reação. Como observamos que NahL também pode possuir um domínio tampa que poderia, por sua vez, excluir a maioria das moléculas de água da cavidade catalítica, e que o substrato poderia interagir com o metal de forma bidentada (assim como em NahK), a estrutura cristalina de NahL complexada com análogos de substrato poderia elucidar tais aspectos da reação, permitindo revisar e melhorar o modelo proposto por Miller.

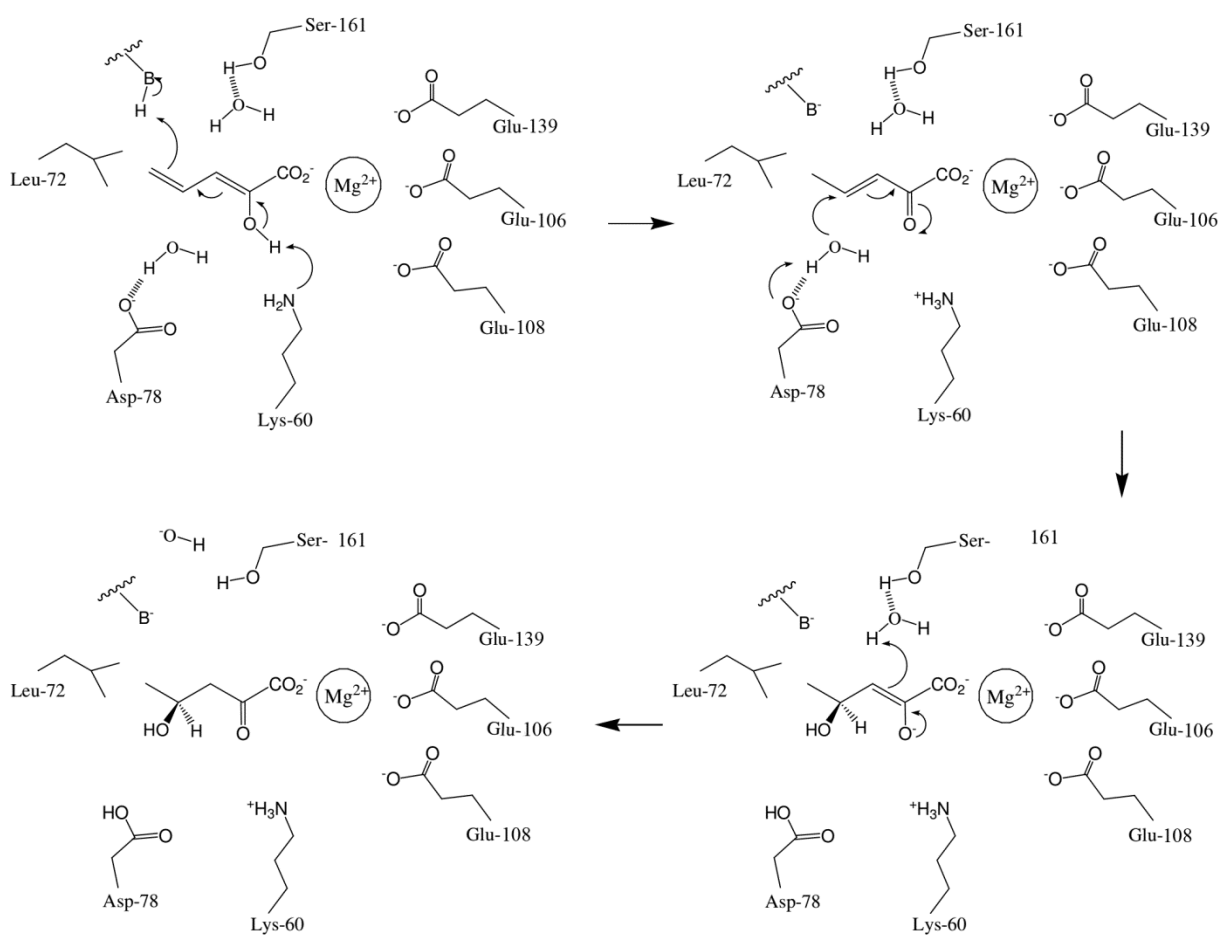


Figura 71: Mecanismo de ação de XylJ proposto por Miller a partir da estrutura de MhpD. Retirado e adaptado de Miller, 2008 [114].

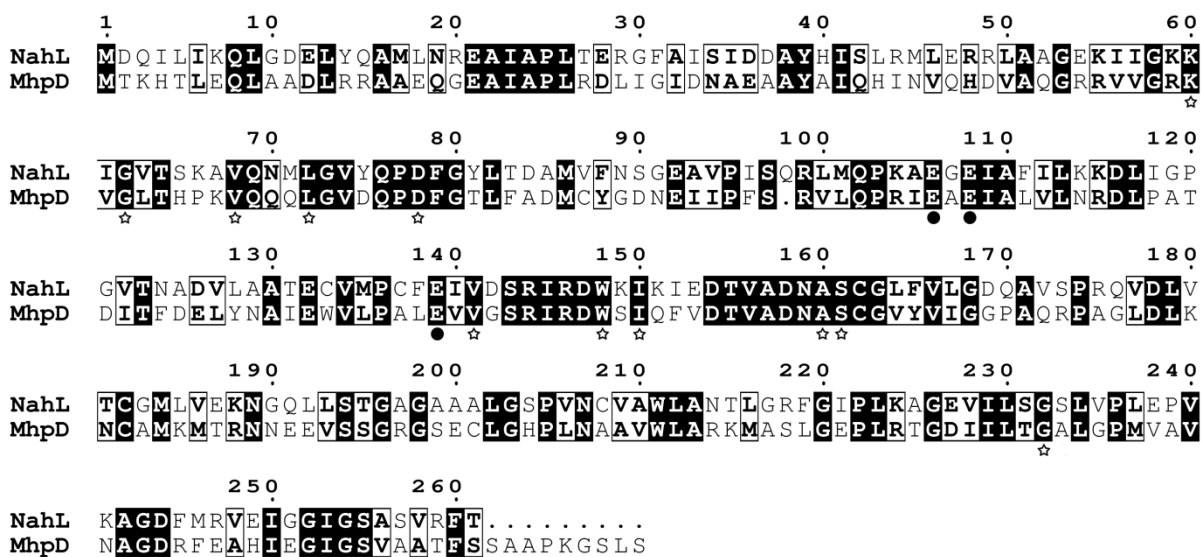


Figura 72: Alinhamento de sequência primária de NahL e MhpD.

Os resíduos que coordenam o íon metálico no sítio ativo estão destacados por um círculo, enquanto que os outros resíduos que também compõe o sítio ativo estão destacados por estrelas.

Apesar de MhpD de *E. coli* agir sobre o mesmo substrato que NahL e XylJ, ambas de *P. putida*, MhpD possui apenas 42% de identidade com NahL. Apesar disso, diversos resíduos que compõe a cavidade do sítio ativo são conversados entre essas enzimas (Figura 72), o que pode explicar a semelhança de seus parâmetros cinéticos e mecanismo de ação [114].

4.17. Estrutura geral do complexo NahK/NahL

Em cada heterodecâmero existem 4 moléculas de NahK (destacadas em azul, Figura 73) e 5 moléculas de NahL (destacadas em cinza). As estruturas de NahK na forma de complexo com NahL e na forma isolada, descrita anteriormente neste trabalho, são extremamente similares, possuindo um rmsd de 0,476 Å entre 260 pares de C α . As maiores diferenças de posição ocorrem na região formada entre os resíduos 150 e 155, que é a volta que se movimenta juntamente com o domínio tampa para o fechamento da cavidade do sítio ativo dessa enzima. Entre si, no complexo, as quatro cadeias de NahK apresentam um rmsd médio de 0,313 Å, enquanto que as cadeias de NahL possuem entre si um rmsd médio de 0,362 Å.

Não surpreendentemente, NahK e NahL possuem basicamente os mesmos elementos de estrutura secundária arranjados num enovelamento extremamente similar (Figura 27 e Figura 68). As estruturas de ambas as enzimas possuem um rmsd de 1.3 Å entre 260 pares de C α . Uma vez que ambas as proteínas são bastante semelhantes, a molécula restante que completa as 10 subunidades de cada complexo possui um mapa de densidade eletrônica que não permite definir com clareza de que se trata de uma quinta molécula de NahK ou uma sexta molécula de NahL. Como o vetor pETDuet-1 expressa independentemente cada um dos genes, e como parece ter havido um excesso de monômeros de NahL visualizados na análise por SDS-PAGE da purificação do complexo por SEC (Figura 53), é possível que parte das moléculas do complexo exista sob uma proporção exata de 5 moléculas de NahK e 5 de NahL, enquanto que parte exista numa proporção de 4 NahK e 6 NahL, gerando assim uma estrutura cristalina com uma dupla conformação de cadeias nessa posição específica (destacada em rosa na Figura 73), dificultando o refinamento dessa cadeia no modelo. Também é possível observar que as cadeias de NahK e NahL se alternam em posição para a formação dos anéis pentaméricos. Entretanto, como são apenas 5 moléculas no anel, há a repetição de uma molécula de NahK ou de NahL nessa sequência. Essa montagem em heterodecâmero foi analisada pelo servidor PISA [100] e foi avaliada como sendo estável, contando com uma superfície acessível ao solvente de 88.050 Å² e uma área enterrada de 22.060 Å².

A cadeia 'A' do modelo parcialmente refinado do complexo trata-se de uma molécula de NahK pertencente ao anel pentamérico 'superior', enquanto que a cadeia 'B' é uma molécula de NahL pertencente ao anel pentamérico 'inferior'. Ambas as cadeias se posicionam levemente deslocadas verticalmente uma com relação à outra. Como foi mencionado, as posições entre NahK e NahL se invertem ao longo do complexo. A superfície de contato entre essas duas cadeias é de apenas 588 Å². As interações nessa superfície são feitas tanto através de resíduos hidrofóbicos, quanto por ligações de hidrogênio e pontes salinas entre resíduos polares presentes nas fitas 8, 9 e 12 de NahK e NahL.

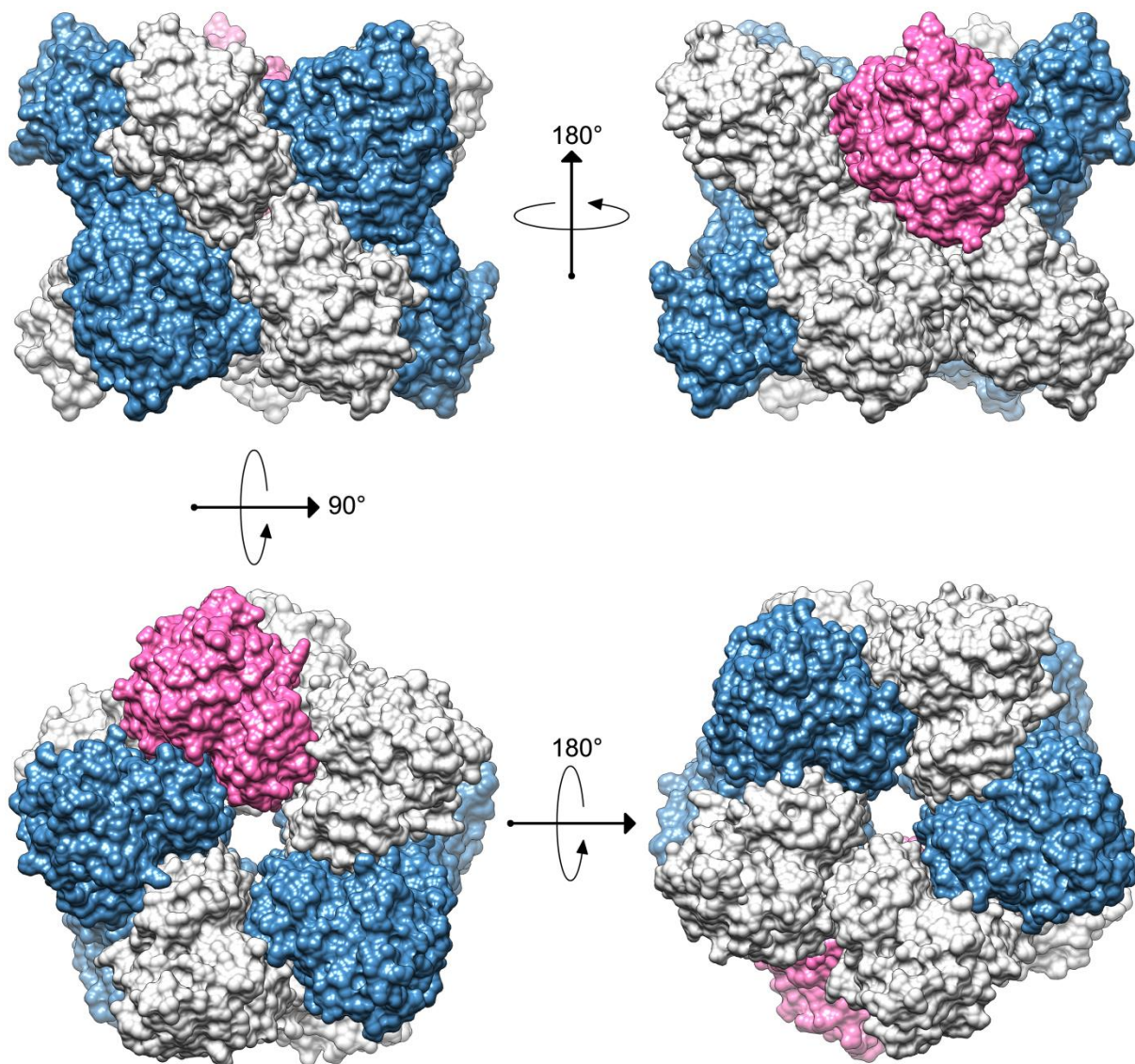


Figura 73: Modelo de superfície das subunidades formadoras dos heterodecâmeros de NahK e NahL. Os monômeros de NahK encontram-se destacados em azul, enquanto que os monômeros de NahL encontram-se em cinza. O monômero rosa representa a posição em que se encaixaria tanto um 5º monômero de NahK ou um 6º monômero de NahL.

Talvez o aspecto mais importante da formação do complexo seja o fato de que NahK e NahL exibem suas respectivas cavidades do sítio ativo uma para a outra (Figura 74). Dessa maneira, o produto instável de NahK pode rapidamente sair de seu sítio ativo e entrar na cavidade catalítica de NahL para dar continuidade ao processo de degradação.

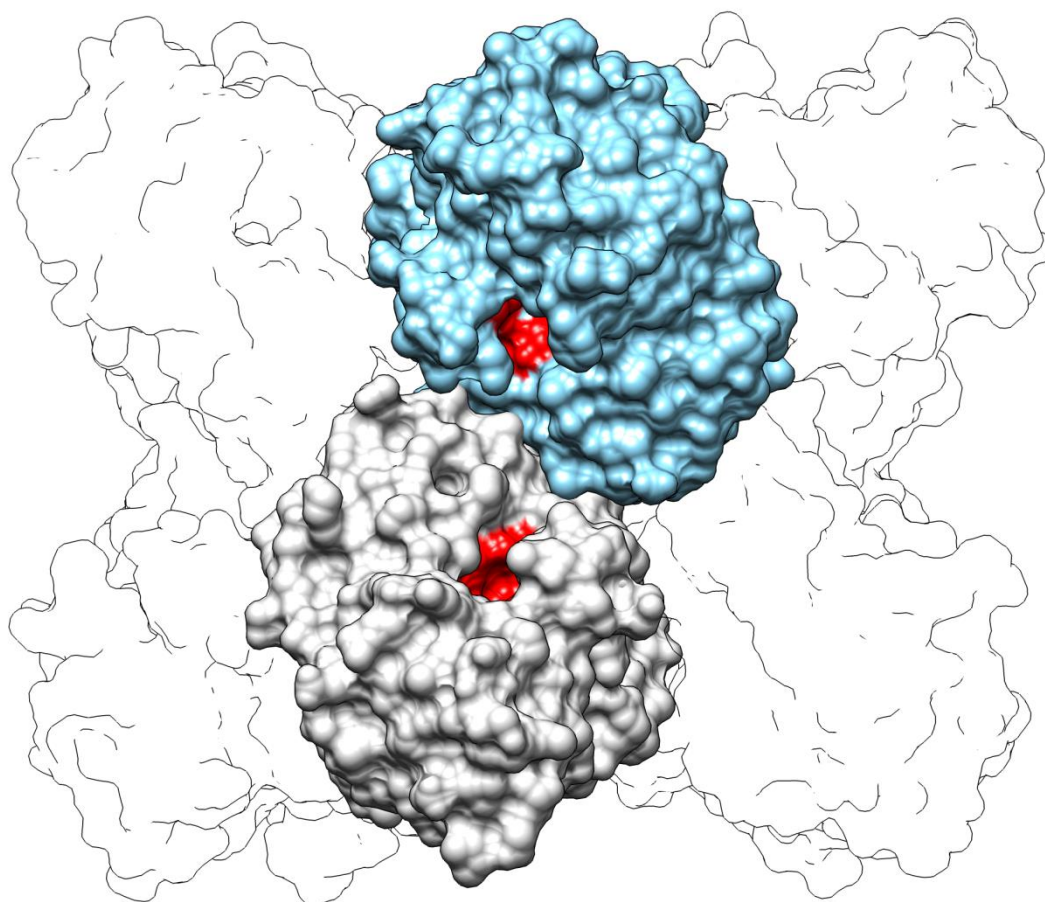


Figura 74: Modelo de superfície do complexo evidenciando um monômero de NahK (azul), um monômero de NahL (cinza) e suas respectivas cavidades catalíticas (vermelho).

A interface entre dois monômeros de NahK de anéis pentaméricos diferentes (Figura 75 a) possui uma área de 470 \AA^2 formada por resíduos polares (arginina, glutamato, asparagina, treonina e serina) e que compartilham ligações de hidrogênio e pontes salinas. Não há participação de resíduos hidrofóbicos nessa interface. Já a interface formada entre dois monômeros de NahL de anéis diferentes (Figura 75 b) possui uma área menor, de 317 \AA^2 , também possui resíduos polares mas aparentemente sem a formação de ligações de hidrogênio (arginina, glutamina, asparagina, glutamato e serina). Nessa interface há a

participação de apenas um resíduo hidrofóbico de cada cadeia, uma fenilalanina, que se posiciona sobre uma glicina da cadeia adjacente.

Ambas as interfaces de monômeros entre os anéis são formadas por resíduos provenientes da folha- β C-terminal das cadeias, evidenciando a função desse elemento estrutural na junção dos dois anéis pentaméricos para a formação do complexo de 10 subunidades.

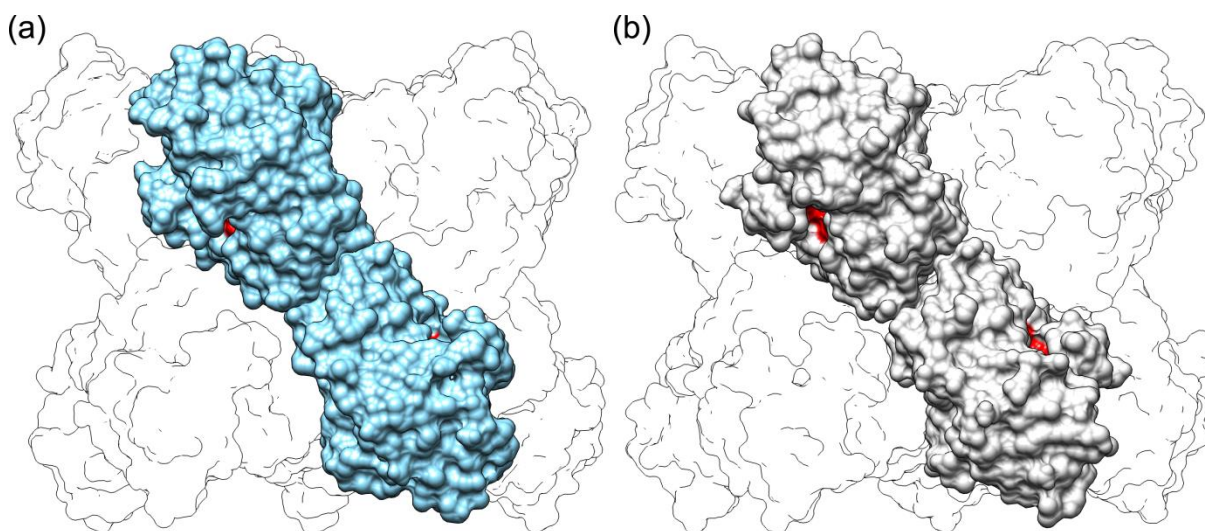


Figura 75: Modelo de superfície do complexo representando a posição de dois monômeros de NahK (em azul) e de dois monômeros de NahL (em cinza).

A área de interface entre monômeros adjacentes para a formação dos anéis pentaméricos é de aproximadamente 634 \AA^2 entre cada monômero, e se dá pela interação de diversos resíduos que formam a base do domínio barril (lado contrário ao topo do barril em que se encontra o íon metálico) com resíduos provenientes das hélices- α L e G (para NahK) ou hélices- α I e E (para NahL) (Figura 76). Essas interações também são observadas para os anéis pentaméricos de MhpD e HpcG.

A análise mais aprofundada da estrutura de NahL e do complexo será feita após a finalização do refinamento e trará um maior entendimento do mecanismo de reação dessas enzimas. Esses resultados também serão submetidos para publicação sob a forma de um artigo científico.

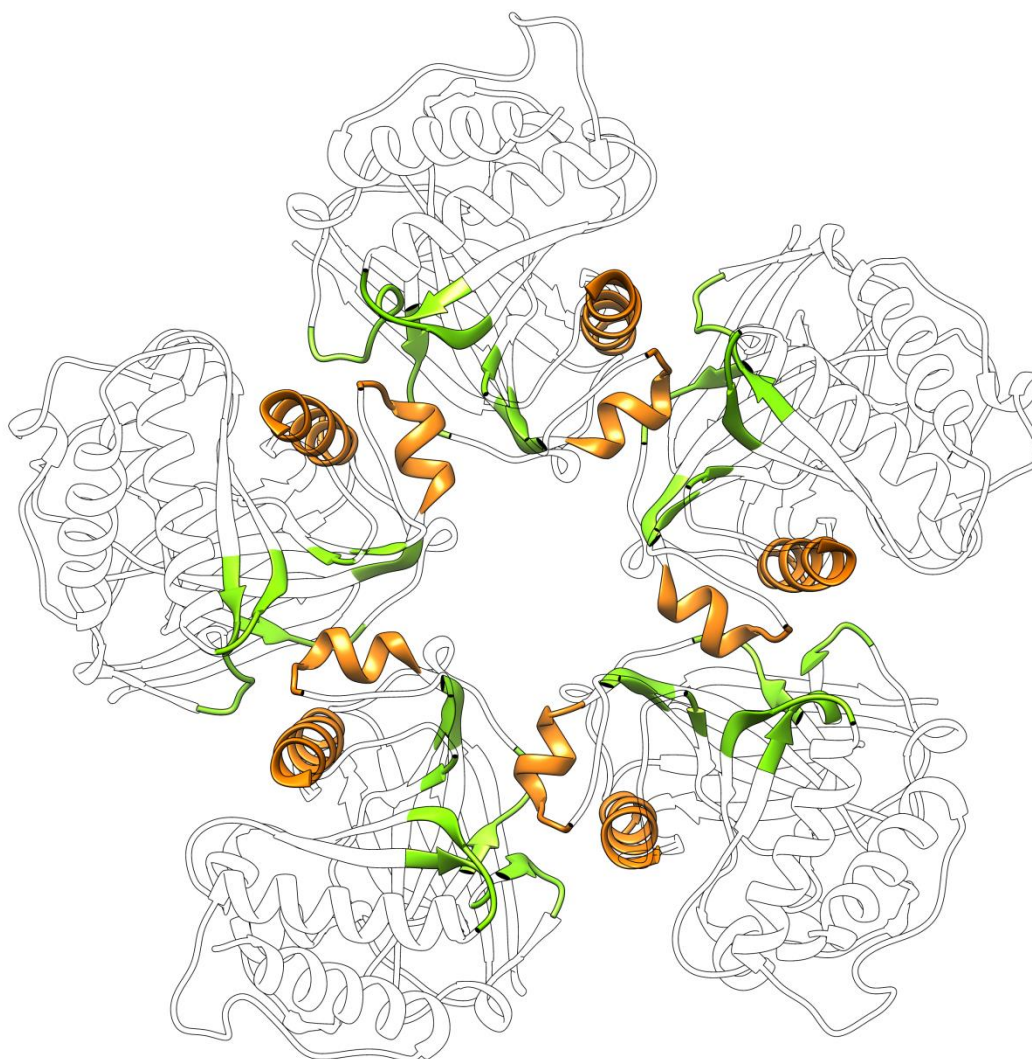


Figura 76: Elementos estruturais responsáveis pela formação do anel pentamérico. Hélices L e G (NahK) e hélices I e E (NahL) estão destacadas em laranja, enquanto que a base do barril- β incompleto de cada monômero está destacada em verde.

4.18. Ajuste dos dados de SAXS utilizando o modelo cristalográfico do complexo NahK/NahL

É possível verificar se o modelo cristalográfico sob a forma de heterodecâmero, proposto para o complexo NahK/NahL, explica a curva experimental obtida por SAXS a partir do complexo em solução. Para isso, utilizamos o servidor WAXSiS [115], o qual calcula uma curva de SAXS teórica utilizando as coordenadas atômicas do modelo cristalográfico em questão, e a compara com os dados experimentais obtidos a partir da amostra em solução. Para o cálculo dessa curva, o servidor utiliza dinâmica molecular para criar uma camada de solvatação em torno dos átomos do modelo, e leva em conta o espalhamento causado por essas moléculas de água, produzindo assim uma curva teórica mais próxima da realidade.

A Figura 77 exhibe o ajuste da curva experimental sobre a curva teórica do modelo cristalográfico obtida pelo servidor WAXSiS. O ajuste gerou um valor de $\chi = 0,054 \pm 0,008$, evidenciando que ambas as curvas são muito semelhantes. Esse resultado sugere que o modelo cristalográfico sob a forma de um heterodecâmero, proposto para o complexo NahK/NahL, corresponde a unidade biológica em solução.

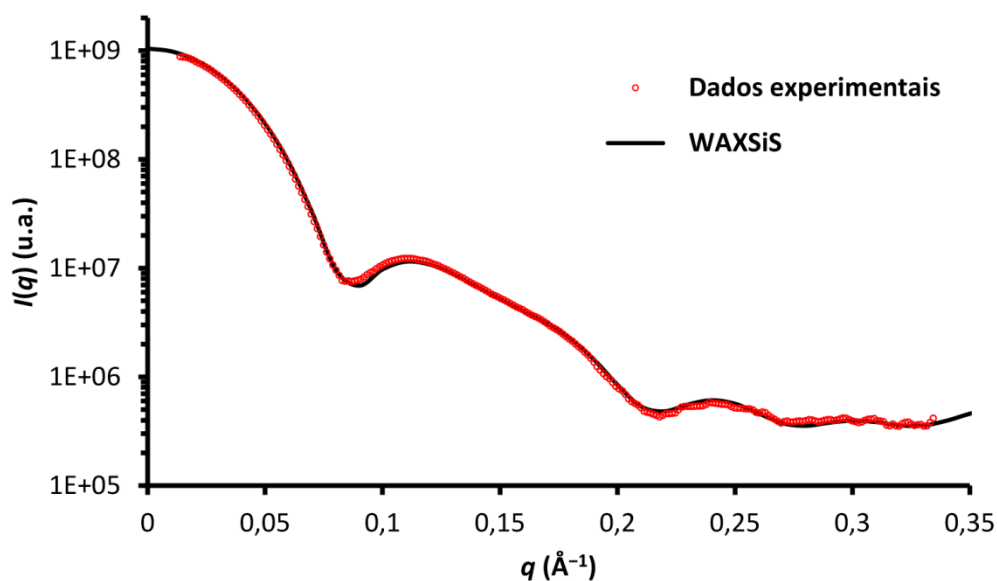


Figura 77: Ajuste da curva experimental do complexo NahK/NahL em solução obtida por SAXS com a curva de espalhamento teórico do heterodecâmero cristalográfico.

5. CONCLUSÃO E PERSPECTIVAS

O petróleo tem sido largamente utilizado como fonte de energia e matéria prima para fabricação de inúmeros produtos de grande importância, e provavelmente continuará sendo bastante utilizado pelas próximas décadas. Dessa maneira, contaminações localizadas do ambiente com petróleo e derivados são virtualmente inevitáveis. Considerando a toxicidade desses poluentes (principalmente devido à presença dos HAP), evidencia-se a necessidade de estudos para potencializar o tratamento de ambientes contaminados.

A bioremediação tornou-se uma interessante estratégia para eliminação de poluentes. Diversos organismos já foram identificados e caracterizados de acordo com sua capacidade em utilizar diversos compostos tóxicos como fonte de carbono e energia, sendo que a utilização de enzimas se mostra relevante a fim de melhorar as atuais técnicas de bioremediação. Nesse contexto, destaca-se a bactéria *P. putida* G7, capaz de degradar completamente o naftaleno. Estudos sobre as enzimas provindas dos genes *nah* são de grande interesse tanto em aspectos de conhecimentos fundamentais quanto aplicados.

As enzimas 4-OD e VPH são estudadas há mais de 44 anos em diversas espécies de bactérias. Vários autores reportaram que o complexo 4-OD/VPH, presente, entre outras, na via de degradação do naftaleno, catalisa uma sequência interessante de reações que, por conta de suas peculiaridades, levanta diversos questionamentos a respeito dos mecanismos de reação e estrutura tridimensional dessas enzimas. Whitman e colegas resolveram a incerteza a respeito dos produtos intermediários envolvidos nessa catálise. Entretanto, até o início do desenvolvimento dos experimentos desta tese, estudos estruturais dessas enzimas eram praticamente inexistentes e dificultados devido a quantidades insuficientes de proteína estável, e ao fato de que 4-OD e VPH existem na forma de um complexo proteico.

Assim, neste trabalho, apresentamos a expressão, purificação e cristalização bem sucedida de NahK e de seu complexo com NahL. Dados de CD sugeriram que NahK possui uma estrutura secundária semelhante à de HpcG e MhpD. Apesar de essas enzimas formarem grandes estruturas quaternárias, NahK recombinante se comporta como um monômero em solução, sugerindo que sua interação com NahL é essencial para oligomerização.

Ensaio cinéticos de NahK confirmaram que sua atividade é independente de sua montagem biológica com NahL, sendo exibidos valores de eficiência catalítica na ordem de grandeza de $10^6 \text{ M}^{-1} \text{ s}^{-1}$, valor semelhante ao de XylI sob a forma de complexo XylI/E106Q_XylI. Esse resultado sugere que a formação do complexo NahK/NahL não aumenta substancialmente a eficiência de NahK, uma vez que sua eficiência catalítica sob a forma de complexo é de $10^7 \text{ M}^{-1} \text{ s}^{-1}$, valor também alcançado por XylI sob a forma do complexo nativo XylI/XylI.

Ensaio cinéticos de NahL na forma de complexo também exibiram uma eficiência catalítica na ordem de $10^7 \text{ M}^{-1} \text{ s}^{-1}$. Apesar de NahK possuir atividade na ausência de NahL, a montagem quaternária dessas enzimas é essencial para o correto processamento do produto instável provindo da reação de descarboxilação.

NahK e NahL pertencem a superfamília FAH, descoberta há 16 anos e que continua crescendo, uma vez que somente 20 membros tiveram suas estruturas tridimensionais elucidadas e depositadas no PDB. Além disso, apenas parte dessas estruturas está atualmente descrita na literatura, sendo que a maioria foi apenas resolvida e depositada.

Cristais da forma apo de NahK, juntamente com sua forma complexada com diferentes ligantes foram difratados e seis conjuntos de dados foram coletados, resultando em seis modelos tridimensionais de NahK. Uma comparação de NahK com todos os modelos tridimensionais dos membros da superfamília FAH revelou que o domínio N-terminal de NahK estabiliza o barril- β incompleto, substituindo uma segunda subunidade presente em todos os outros membros dessa superfamília, permitindo também um diferente arranjo quaternário. Esse domínio N-terminal característico também está presente nas estruturas de NahL, MhpD e HpcG.

Além disso, nossos resultados sugerem que NahK pode adotar 3 conformações particulares mediante interação com o substrato, na qual o domínio tampa se encontra aberto, parcialmente fechado e fechado. A reação de descarboxilação provavelmente requer que o substrato adote uma conformação bidentada com o metal. Uma vez que o domínio tampa é fechado, o substrato adota uma geometria apropriada para a descarboxilação, a qual também é auxiliada por um ambiente hidrofóbico. O domínio tampa parece ser conservado dentre os demais membros da superfamília FAH.

Um cristal do complexo NahK/NahL foi difratado, resultando em um modelo tridimensional parcialmente refinado, mas que já exibe detalhes suficientes para sua análise. Consequentemente, com a resolução da estrutura do complexo, também se obteve a estrutura tridimensional de NahL, que se apresenta extremamente semelhante a NahK, tanto em estruturas secundárias quanto terciária. A estrutura de NahL, comparada com as estruturas de MhpD e HpcG, pode levar a um melhor entendimento do processo de hidratação realizado por essas enzimas.

Ao contrário da maioria dos membros da superfamília FAH, que adota uma forma dimérica ou a forma de um dímero de dímeros, NahK se comporta como monômero em solução na ausência de NahL. Sua oligomerização com NahL forma um heterodecâmero (5 x NahK + 5 x NahL) semelhante ao arranjo adotado pelas dez subunidades em HpcG. Modelos de SAXS corroboram essa montagem quaternária, que também foi observada na estrutura cristalina do complexo. NahK e NahL exibem suas cavidades do sítio ativo uma para a outra na tentativa de lidar rapidamente com o produto instável gerado pela descarboxilação do 4-oxalocrotonato.

A estrutura de NahK e NahL apresentada neste trabalho é inédita, constituindo as primeiras estruturas descritas de uma 4-OD e VPH, respectivamente. A estratégia e o mecanismo de reação proposto neste trabalho para NahK representa um novo mecanismo de ação para a superfamília FAH, e é o primeiro mecanismo proposto para a descarboxilação dependente de metal de um β -cetoácido vinílico. Devido a grande semelhança dentre o substrato de NahK e HpcE, é provável que essa enzima também utilize o mesmo mecanismo de descarboxilação.

Com base nos nossos resultados, mutações sítio dirigidas em resíduos de aminoácidos no sítio ativo de NahK podem auxiliar no melhor entendimento de sua reação e do modelo proposto aqui. A estrutura de NahL complexada com análogos de substrato podem auxiliar no entendimento de sua estrutura-função e também verificar a presença de um provável domínio tampa. Esses resultados abrem caminho para futuras investigações, possibilitando o uso de engenharia de proteínas para aumentar a estabilidade e atividade a fim de que essas enzimas possam ser utilizadas futuramente em processos biotecnológicos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] SOLOMONS, T. W. G. *et al.* *Química orgânica*; 7ª ed.; ed. Rio de Janeiro, 2001.
- [2] VAN HAMME, J. D. *et al.* (2003). Recent advances in petroleum microbiology. *Microbiol Mol Biol Rev*, 67, 503.
- [3] ANTOS, G. J. *et al.* *Catalytic Naphtha Reforming*; 2th ed.; CRC Press: New York, , 2004.
- [4] PERRONS, R. K. (2013). Assessing the damage caused by Deepwater Horizon: Not just another Exxon Valdez. *Mar Pollut Bull*.
- [5] PEREZ, S. *et al.* (2001). Occurrence of polycyclic aromatic hydrocarbons in sewage sludge and their contribution to its toxicity in the toxalert 100 bioassay. *Chemosphere*, 45, 705.
- [6] GAN, S. *et al.* (2009). Remediation of soils contaminated with polycyclic aromatic hydrocarbons (PAHs). *J Hazard Mater*, 172, 532.
- [7] YEUNG, P. Y. *et al.* (1997). Biodegradation of petroleum hydrocarbons in soil as affected by heating and forced aeration. *Journal of Environmental Quality*, 26, 1511.
- [8] HOLLIGER, C. *et al.* (1997). Contaminated environments in the subsurface and bioremediation: organic contaminants. *FEMS Microbiol Rev*, 20, 517.
- [9] NETTO, A. D. P. *et al.* (2000). Avaliação da contaminação humana por hidrocarbonetos policíclicos aromáticos (HPAs) e seus derivados nitrados (nHPAs): uma revisão metodológica. *Química Nova*, 23, 765.
- [10] PURCARO, G. *et al.* (2013). Overview on polycyclic aromatic hydrocarbons: occurrence, legislation and innovative determination in foods. *Talanta*, 105, 292.
- [11] JACQUES, R. J. S. *et al.* (2007). Biodegradação de hidrocarbonetos aromáticos policíclicos. *Ciência e Natura*, 29, 7.
- [12] JOHNSEN, A. R. *et al.* (2005). Principles of microbial PAH-degradation in soil. *Environ Pollut*, 133, 71.
- [13] BAMFORTH, S. M. *et al.* (2005). Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions. *Journal of Chemical Technology and Biotechnology*, 80, 723.
- [14] TIBURTIUS, E. R. L. *et al.* (2004). Contaminação de águas por BTXs e processos utilizados na remediação de sítios contaminados. *Química Nova*, 27, 441.
- [15] MASTRANGELO, G. *et al.* (1996). Polycyclic aromatic hydrocarbons and cancer in man. *Environ Health Perspect*, 104, 1166.

- [16] CERNIGLIA, C. E. (1984). Microbial metabolism of polycyclic aromatic hydrocarbons. *Adv Appl Microbiol*, 30, 31.
- [17] BRÄNDLI, R. *et al.* (2004). Occurrence and relevance of organic pollutants in compost, digestate and organic residue. *EPF Lausanne - ENAC - ISTE - CECOTOX / Agroscope FAL Reckenholz*.
- [18] PÉREZ, S. *et al.* (2001). Occurrence of polycyclic aromatic hydrocarbons in sewage sludge and their contribution to its toxicity in the toxalert 100 bioassay. *Chemosphere*, 45, 705.
- [19] DUA, M. *et al.* (2002). Biotechnology and bioremediation: successes and limitations. *Appl Microbiol Biotechnol*, 59, 143.
- [20] HARITASH, A. K. *et al.* (2009). Biodegradation aspects of polycyclic aromatic hydrocarbons (PAHs): a review. *J Hazard Mater*, 169, 1.
- [21] CERNIGLIA, C. E. (1997). Fungal metabolism of polycyclic aromatic hydrocarbons: past, present and future applications in bioremediation. *J Ind Microbiol Biotechnol*, 19, 324.
- [22] SAMANTA, S. K. *et al.* (2002). Polycyclic aromatic hydrocarbons: environmental pollution and bioremediation. *Trends Biotechnol*, 20, 243.
- [23] BAMFORTH, S. M. *et al.* (2005). Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions. *J Chem Technol Biot*, 80, 723.
- [24] GALLEGO, J. L. *et al.* (2001). Bioremediation of diesel-contaminated soils: evaluation of potential in situ techniques by study of bacterial degradation. *Biodegradation*, 12, 325.
- [25] SHEN, T. *et al.* (2015). Biodegradation of different petroleum hydrocarbons by free and immobilized microbial consortia. *Environ Sci Process Impacts*.
- [26] EL FANTROUSSI, S. *et al.* (2005). Is bioaugmentation a feasible strategy for pollutant removal and site remediation? *Curr Opin Microbiol*, 8, 268.
- [27] ATLAS, R. M. *et al.* (2011). Oil biodegradation and bioremediation: a tale of the two worst spills in U.S. history. *Environ Sci Technol*, 45, 6709.
- [28] PETERSON, C. H. *et al.* (2003). Long-term ecosystem response to the Exxon Valdez oil spill. *Science*, 302, 2082.
- [29] BOUCHEZ, T. *et al.* (2000). Ecological study of a bioaugmentation failure. *Environ Microbiol*, 2, 179.
- [30] PAUL, D. *et al.* (2005). Accessing microbial diversity for bioremediation and environmental restoration. *Trends Biotechnol*, 23, 135.
- [31] AHUJA, S. K. *et al.* (2004). Utilization of enzymes for environmental applications. *Crit Rev Biotechnol*, 24, 125.

- [32] ALCALDE, M. *et al.* (2006). Environmental biocatalysis: from remediation with enzymes to novel green processes. *Trends in Biotechnology*, 24, 281.
- [33] SUTHERLAND, T. D. *et al.* (2004). Enzymatic bioremediation: from enzyme discovery to applications. *Clin Exp Pharmacol Physiol*, 31, 817.
- [34] VALDMAN, E. *et al.* (2004). On-line detection of low naphthalene concentrations with a bioluminescent sensor. *Process Biochemistry*, 39, 1217.
- [35] SEO, J. S. *et al.* (2009). Bacterial degradation of aromatic compounds. *Int J Environ Res Public Health*, 6, 278.
- [36] JIA, C. *et al.* (2010). A critical review of naphthalene sources and exposures relevant to indoor and outdoor air. *Int J Environ Res Public Health*, 7, 2903.
- [37] NORTH, D. W. *et al.* (2008). A review of whole animal bioassays of the carcinogenic potential of naphthalene. *Regul Toxicol Pharmacol*, 51, S6.
- [38] STOHS, S. J. *et al.* (2002). Naphthalene toxicity and antioxidant nutrients. *Toxicology*, 180, 97.
- [39] EPA, U. S. *Toxicological review of naphthalene*; U.S. Environmental Protection Agency: Washington, 2004.
- [40] SOTA, M. *et al.* (2006). Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and its transposon Tn4655 suggests catabolic gene spread by a tyrosine recombinase. *J Bacteriol*, 188, 4057.
- [41] DUNN, N. W. *et al.* (1973). Transmissible plasmid coding early enzymes of naphthalene oxidation in *Pseudomonas putida*. *J Bacteriol*, 114, 974.
- [42] HABE, H. *et al.* (2003). Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. *Biosci Biotechnol Biochem*, 67, 225.
- [43] GRIMM, A. C. *et al.* (1999). NahY, a catabolic plasmid-encoded receptor required for chemotaxis of *Pseudomonas putida* to the aromatic hydrocarbon naphthalene. *J Bacteriol*, 181, 3310.
- [44] SALA-TREPAT, J. M. *et al.* (1972). The metabolic divergence in the meta cleavage of catechols by *Pseudomonas putida* NCIB 10015. Physiological significance and evolutionary implications. *Eur J Biochem*, 28, 347.
- [45] HARAYAMA, S. *et al.* (1987). Evolutionary relationships between catabolic pathways for aromatics: conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. *Mol Gen Genet*, 210, 241.
- [46] HARAYAMA, S. *et al.* (1984). Transposon mutagenesis analysis of meta-cleavage pathway operon genes of the TOL plasmid of *Pseudomonas putida* mt-2. *J Bacteriol*, 160, 251.

- [47] HARAYAMA, S. *et al.* (1993). Comparison of the nucleotide sequences of the meta-cleavage pathway genes of TOL plasmid pWW0 from *Pseudomonas putida* with other meta-cleavage genes suggests that both single and multiple nucleotide substitutions contribute to enzyme evolution. *Mol Gen Genet*, 239, 81.
- [48] ROPER, D. I. *et al.* (1993). Purification, nucleotide sequence and some properties of a bifunctional isomerase/decarboxylase from the homoprotocatechuate degradative pathway of *Escherichia coli* C. *Eur J Biochem*, 217, 575.
- [49] ROPER, D. I. *et al.* (1993). The *Escherichia coli* C homoprotocatechuate degradative operon: hpc gene order, direction of transcription and control of expression. *Mol Gen Genet*, 237, 241.
- [50] ROPER, D. I. *et al.* (1995). Sequence of the hpcC and hpcG genes of the meta-fission homoprotocatechuate acid pathway of *Escherichia coli* C: nearly 40% amino-acid identity with the analogous enzymes of the catechol pathway. *Gene*, 156, 47.
- [51] RHODES, G. *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*; 2^a ed.; Academic Press, 2000.
- [52] BERMAN, H. M. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res*, 28, 235.
- [53] BROWN, E. N. *et al.* (2008). Determining Rieske cluster reduction potentials. *J Biol Inorg Chem*, 13, 1301.
- [54] KITA, A. *et al.* (1999). An archetypical extradiol-cleaving catecholic dioxygenase: the crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from *Pseudomonas putida* mt-2. *Structure*, 7, 25.
- [55] CARREDANO, E. *et al.* (2000). Substrate binding site of naphthalene 1,2-dioxygenase: functional implications of indole binding. *J Mol Biol*, 296, 701.
- [56] ARAÚJO, S. S. D. *et al.* (2015). Structural and kinetic characterization of recombinant 2-hydroxymuconate semialdehyde dehydrogenase from *Pseudomonas putida* G7. *Archives of Biochemistry and Biophysics*, 579, 8.
- [57] SUBRAMANYA, H. S. *et al.* (1996). Enzymatic ketonization of 2-hydroxymuconate: specificity and mechanism investigated by the crystal structures of two isomerases. *Biochemistry*, 35, 792.
- [58] THOMPSON, L. C. *et al.* (2007). 2-Hydroxychromene-2-carboxylic acid isomerase: a kappa class glutathione transferase from *Pseudomonas putida*. *Biochemistry*, 46, 6710.
- [59] COITINHO, J. B. *et al.* (2012). Expression, purification and preliminary crystallographic studies of NahF, a salicylaldehyde dehydrogenase from *Pseudomonas putida* G7 involved in naphthalene degradation. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 68, 93.
- [60] TAME, J. R. *et al.* (2002). The crystal structure of HpcE, a bifunctional decarboxylase/isomerase with a multifunctional fold. *Biochemistry*, 41, 2982.

- [61] ALTSCHUL, S. F. *et al.* (1990). Basic local alignment search tool. *J Mol Biol*, 215, 403.
- [62] MONTGOMERY, M. G. *et al.* (2010). Assembly of a 20-nm protein cage by *Escherichia coli* 2-hydroxypentadienoic acid hydratase. *J Mol Biol*, 396, 1379.
- [63] IZUMI, A. *et al.* (2007). Structure and mechanism of HpcG, a hydratase in the homoprotocatechuate degradation pathway of *Escherichia coli*. *J Mol Biol*, 370, 899.
- [64] HARAYAMA, S. *et al.* (1989). Physically associated enzymes produce and metabolize 2-hydroxy-2,4-dienoate, a chemically unstable intermediate formed in catechol metabolism via meta cleavage in *Pseudomonas putida*. *J Bacteriol*, 171, 6251.
- [65] STANLEY, T. M. *et al.* (2000). Expression and stereochemical and isotope effect studies of active 4-oxalocrotonate decarboxylase. *Biochemistry*, 39, 3514.
- [66] TIMM, D. E. *et al.* (1999). Crystal structure and mechanism of a carbon-carbon bond hydrolase. *Structure*, 7, 1023.
- [67] MANJASETTY, B. A. *et al.* (2004). X-ray structure of fumarylacetoacetate hydrolase family member *Homo sapiens* FLJ36880. *Biol Chem*, 385, 935.
- [68] BROUNS, S. J. *et al.* (2008). Structural insight into substrate binding and catalysis of a novel 2-keto-3-deoxy-D-arabinonate dehydratase illustrates common mechanistic features of the FAH superfamily. *J Mol Biol*, 379, 357.
- [69] RAN, T. *et al.* (2013). Crystal structures of Cg1458 reveal a catalytic lid domain and a common catalytic mechanism for the FAH family. *Biochem J*, 449, 51.
- [70] GUIMARÃES, S. L. Dissertação de Mestrado, Universidade Federal de Minas Gerais, 2011.
- [71] INVITROGEN 17 de junho de 2007; Vol. 2011.
- [72] TAYLOR, R. G. *et al.* (1993). *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res*, 21, 1677.
- [73] LOENEN, W. A. (2003). Tracking EcoKI and DNA fifty years on: a golden story full of surprises. *Nucleic Acids Res*, 31, 7059.
- [74] VIEIRA, J. *et al.* (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*, 19, 259.
- [75] STUDIER, F. W. *et al.* (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, 189, 113.
- [76] SANGER, F. *et al.* (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74, 5463.
- [77] ARTIMO, P. *et al.* (2012). ExpASy: SIB bioinformatics resource portal. *Nucleic Acids Res*, 40, W597.

- [78] BOHM, G. *et al.* (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng*, 5, 191.
- [79] PETOUKHOV, M. V. *et al.* (2012). New developments in the program package for small-angle scattering data analysis. *J Appl Crystallogr*, 45, 342.
- [80] PETERSEN, E. F. *et al.* (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, 25, 1605.
- [81] FISCHER, H. *et al.* (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography*, 43, 101.
- [82] BATTYE, T. G. *et al.* (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*, 67, 271.
- [83] OTWINOWSKI, Z. *et al.* (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology*, 276, 307.
- [84] WINN, M. D. *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*, 67, 235.
- [85] MCCOY, A. J. *et al.* (2007). Phaser crystallographic software. *J Appl Crystallogr*, 40, 658.
- [86] COWTAN, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr*, 62, 1002.
- [87] ADAMS, P. D. *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66, 213.
- [88] EMSLEY, P. *et al.* (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66, 486.
- [89] HASPEL, J. *et al.* (2001). System for cleavable Fc fusion proteins using tobacco etch virus (TEV) protease. *Biotechniques*, 30, 60.
- [90] BANACHOWICZ, E. (2006). Light scattering studies of proteins under compression. *Biochim Biophys Acta*, 1764, 405.
- [91] DEV, S. *et al.* (2006). Dynamic light scattering study of peanut agglutinin: size, shape and urea denaturation. *J Biosci*, 31, 551.
- [92] FERRE-D'AMARE, A. R. *et al.* (1994). Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. *Structure*, 2, 357.
- [93] SREERAMA, N. *et al.* (2004). Computation and analysis of protein circular dichroism spectra. *Methods Enzymol*, 383, 318.

- [94] FRISHMAN, D. *et al.* (1995). Knowledge-based protein secondary structure assignment. *Proteins*, 23, 566.
- [95] JOHNSON, W. H., JR. *et al.* (2004). 4-Oxalocrotonate tautomerase, its homologue YwhB, and active vinylpyruvate hydratase: synthesis and evaluation of 2-fluoro substrate analogues. *Biochemistry*, 43, 10490.
- [96] STIVALA, A. *et al.* (2011). Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics*, 27, 3315.
- [97] HOLM, L. *et al.* (1995). Dali: a network tool for protein structure comparison. *Trends Biochem Sci*, 20, 478.
- [98] BAUGH, L. *et al.* (2015). Increasing the structural coverage of tuberculosis drug targets. *Tuberculosis (Edinb)*, 95, 142.
- [99] KUBE, M. *et al.* (2013). Genome sequence and functional genomic analysis of the oil-degrading bacterium *Oleispira antarctica*. *Nat Commun*, 4, 2156.
- [100] KRISSINEL, E. *et al.* (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 372, 774.
- [101] DUNDAS, J. *et al.* (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res*, 34, W116.
- [102] PERMYAKOV, E. A. *et al.* *Calcium binding proteins*; Wiley: Hoboken, N.J., 2010.
- [103] BLACK, C. B. *et al.* (1994). Biological coordination chemistry of magnesium, sodium, and potassium ions. Protein and nucleotide binding sites. *Coordination Chemistry Reviews*, 135–136, 165.
- [104] COLLINSWORTH, W. L. *et al.* (1973). Stereospecific enzymes in the degradation of aromatic compounds by *Pseudomonas putida*. *J Bacteriol*, 113, 922.
- [105] BAYLY, R. C. *et al.* (1969). Oxoenic acids as metabolites in the bacterial degradation of catechols. *Biochemical Journal*, 111, 303.
- [106] CATTERALL, F. A. *et al.* (1971). The coexistence of two pathways for the metabolism of 2-hydroxymuconic semialdehyde in a naphthalene-grown *Pseudomonas*. *Biochem Biophys Res Commun*, 43, 463.
- [107] O'LEARY, M. H. In *The Enzymes: Mechanisms of catalysis*; Elsevier Science: Burlington, 1992; Vol. 20, p 235.
- [108] LI, T. *et al.* (2012). Decarboxylation mechanisms in biological system. *Bioorg Chem*, 43, 2.

- [109] LIAN, H. *et al.* (1994). Stereochemical and Isotopic Labeling Studies of 4-Oxalocrotonate Decarboxylase and Vinylpyruvate Hydratase: Analysis and Mechanistic Implications. *Journal of the American Chemical Society*, 116, 10403.
- [110] GRISSOM, C. B. *et al.* (1986). Carbon isotope effects on the metal ion catalyzed decarboxylation of oxalacetate. *Journal of the American Chemical Society*, 108, 5582.
- [111] KRISHNAMURTHY, S. (1982). The principle of vinylogy. *Journal of Chemical Education*, 59, 543.
- [112] POLLARD, J. R. *et al.* (1998). Purification, characterisation and reaction mechanism of monofunctional 2-hydroxypentadienoic acid hydratase from *Escherichia coli*. *Eur J Biochem*, 251, 98.
- [113] BURKS, E. A. *et al.* (1998). Stereochemical and Isotopic Labeling Studies of 2-Oxohept-4-ene-1,7-dioate Hydratase: Evidence for an Enzyme-Catalyzed Ketonization Step in the Hydration Reaction. *Journal of the American Chemical Society*, 120, 7665.
- [114] MILLER, S. G. Tese de Doutorado, The University of Texas at Austin, 2008.
- [115] KNIGHT, C. J. *et al.* (2015). WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res*, 43, W225.

7. ANEXO I

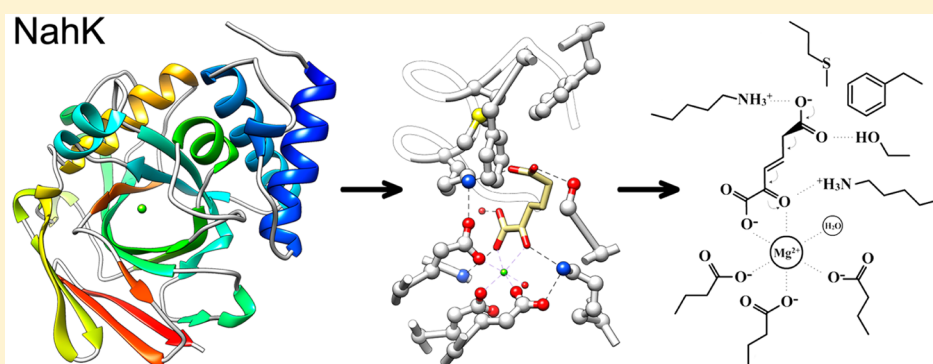
Crystal Structures of Apo and Liganded 4-Oxalocrotonate Decarboxylase Uncover a Structural Basis for the Metal-Assisted Decarboxylation of a Vinylogous β -Keto Acid

Samuel L. Guimarães,[‡] Juliana B. Coitinho,[‡] Débora M. A. Costa,[‡] Simara S. Araújo,[‡] Christian P. Whitman,[§] and Ronaldo A. P. Nagem^{*,‡}

[‡]Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 31270-901, Brazil

[§]Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas, Austin, Texas 78712-1071, United States

S Supporting Information



ABSTRACT: The enzymes in the catechol *meta*-fission pathway have been studied for more than 50 years in several species of bacteria capable of degrading a number of aromatic compounds. In a related pathway, naphthalene, a toxic polycyclic aromatic hydrocarbon, is fully degraded to intermediates of the tricarboxylic acid cycle by the soil bacteria *Pseudomonas putida* G7. In this organism, the 83 kb NAH7 plasmid carries several genes involved in this biotransformation process. One enzyme in this route, NahK, a 4-oxalocrotonate decarboxylase (4-OD), converts 2-oxo-3-hexenedioate to 2-hydroxy-2,4-pentadienoate using Mg^{2+} as a cofactor. Efforts to study how 4-OD catalyzes this decarboxylation have been hampered because 4-OD is present in a complex with vinylpyruvate hydratase (VPH), which is the next enzyme in the same pathway. For the first time, a monomeric, stable, and active 4-OD has been expressed and purified in the absence of VPH. Crystal structures for NahK in the apo form and bonded with five substrate analogues were obtained using two distinct crystallization conditions. Analysis of the crystal structures implicates a lid domain in substrate binding and suggests roles for specific residues in a proposed reaction mechanism. In addition, we assign a possible function for the NahK N-terminal domain, which differs from most of the other members of the fumarylacetoacetate hydrolase superfamily. Although the structural basis for metal-dependent β -keto acid decarboxylases has been reported, this is the first structural report for that of a vinylogous β -keto acid decarboxylase and the first crystal structure of a 4-OD.

Naphthalene degradation by *Pseudomonas putida* G7 has been studied extensively since its 83-kilobase plasmid was first isolated.^{1,2} The NAH7 plasmid carries the catabolic genes (*nah*) organized in two operons. The gene products encoded by the upper *nah* operon convert naphthalene (1, Figure 1) to salicylate (2, Figure 1) and pyruvate, whereas the enzymes encoded by the genes of the lower *nah* operon process 2 to a second pyruvate molecule and acetyl-CoA. The overall structure and gene organization of the lower pathway of the naphthalene degradation are very similar to those observed in the catechol *meta*-fission pathways in different *Pseudomonas* strains, which promote the aerobic degradation of monocyclic aromatic compounds such as benzene, toluene, and xylenes

(e.g., the *xyl* genes of the TOL plasmid pWW0 from *P. putida* mt-2).³

In *P. putida* G7, the fifth reaction in the lower pathway is carried out by NahK, a 4-oxalocrotonate decarboxylase (4-OD, EC 4.1.1.77). The enzyme converts a vinylogous β -keto acid, 2-oxo-3-hexenedioate (3, Figure 1), to 2-hydroxy-2,4-pentadienoate (4, Figure 1) and CO_2 requiring only a magnesium ion as cofactor. Study of this 4-OD is complicated by the observation

Received: January 20, 2016

Revised: April 11, 2016

Published: April 15, 2016

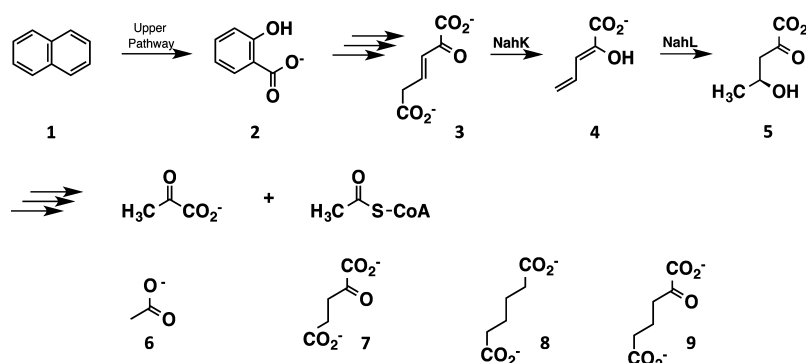


Figure 1. Naphthalene-degradation pathway in *P. putida* G7. The upper pathway degrades naphthalene (1) to pyruvate and salicylate (2) using the enzymes NahAaAbAcAdBCDEF. The lower pathway is composed of the enzymes NahGHIJKLMNO and converts 2 to pyruvate and acetyl-CoA. The intermediates, 2-oxo-3-hexenedioate (3), 2-hydroxy-2,4-pentadienoate (4), and 4-hydroxy-2-oxopentanoate (5) are also depicted. Molecules with structures similar to NahK's substrate are shown: acetate (6), α -ketoglutarate (7), adipate (8), and 2-oxoadipate (9).

that it forms a complex with NahL, a vinylpyruvate hydratase (VPH, EC 4.2.1.80), the next enzyme in the same degradation pathway. A similar observation has been made for XylI and XylJ from *P. putida* mt-2,⁴ which are the NahK and NahL homologues, respectively. VPH then converts 4 to 4-hydroxy-2-oxopentanoate (5, Figure 1).

The reactions carried out by the 4-OD/VPH complex are interesting and raise mechanistic and structural questions.^{5–8} Among these questions is the structural basis for both reactions and how 4-OD catalyzes the decarboxylation reaction. A second question is whether the two enzymes have separate active sites where the unstable 4-OD product would be channeled to the VPH active site or a single active site. There are also evolutionary questions. Both enzymes belong to the fumarylacetoacetate hydrolase (FAH) superfamily, which includes MhpD from *Escherichia coli* K-12,⁹ a similar hydratase in the phenylpropionate degradation pathway, and HpcE¹⁰ and HpcG¹¹ from *E. coli* C, a decarboxylase and hydratase, respectively, in the homoprotocatechuate degradation pathway. NahK and NahL share 39% sequence identity and likely have the same fold. These observations raise the question of whether a duplication event gave rise to one of the enzymes. Thus, for more than 15 years, different research groups have tried to solve the 3-D structure of the 4-OD/VPH complex, but so far without success.¹²

Attempts to address these questions in the 4-OD/VPH complex from *P. putida* mt-2 have been limited by the fact that XylI is unstable in the absence of XylJ.^{6,12,13} By using NahK from *P. putida* G7, we have now produced a stable 4-OD in large quantities that has kinetic parameters similar to those of 4-OD in the XylI/XylJ native complex. Crystallographic analysis of the apo form and substrate analogue complexes has uncovered the structural basis for the metal-assisted decarboxylation. In this mechanism, Glu109, Glu111, and Glu142 function as metal binding residues while Lys64, Lys72, and Ser164 play binding and catalytic roles. The side chains of Lys72 and Ser164 position the departing carboxylate group almost perpendicular to the plane of the substrate and place it in front of the hydrophobic side chains of Met76, Phe151, and Phe153. The side chain of Lys64 provides the proton to the dienolate intermediate to form the final dienol product.

This work describes, for the first time, a structural basis for the metal-assisted decarboxylation of a vinylous β -keto acid and the first crystal structure of a 4-OD, and sets the stage for future structural studies of the NahK/NahL complex. The

results also have implications for the structural diversity of the FAH superfamily as well as the common features.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification. The coding sequence of *nahK* was amplified from the NAH7 plasmid² using the forward primer 5'-GCT-AGC-ATG-AAC-CGT-ACC-CTG-ACC-CGC-3', and the reverse primer 5'-AAG-CTT-TCA-GAT-AAA-GCG-CGC-GGA-TAC-CG-3'. The sites for the restriction enzymes *Nhe*I and *Hind*III, respectively, are underlined. The amplified PCR fragment was purified, cloned into pGEM-T vector (Promega), and subcloned into pET28a-TEV.¹⁴ The expression vector was inserted into *E. coli* BL21(DE3) (Novagen). The correct position and reading frame of the inset were verified by DNA sequencing using T7 primers in a MegaBACE instrument (GE Healthcare Life Sciences).

For protein production, 20 mL of LB medium supplemented with 50 μ g/mL kanamycin was inoculated with a *nahK*-transformed clone of *E. coli* BL21(DE3) and grown at 37 °C for 16 h. The culture was diluted 1:100 in 2 L of LB medium supplemented with kanamycin, and grown at 37 °C until an OD₆₀₀ of 0.5–0.7 was reached. The expression was induced with isopropyl β -D-1-thiogalactopyranoside at 0.5 mM, and the cells were grown for an additional 5 h at 30 °C. After expression, cells were harvested by centrifugation at 5000g, and the pellet was stored at –20 °C. Prior to cell lysis, the pellet was resuspended in 40 mL of lysis buffer (20 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl, 45 mM imidazole, and 100 μ g/mL lysozyme). Cells were lysed with three freeze–thaw cycles and sonication was carried out in an ice–water bath. Soluble and insoluble fractions were separated by centrifugation at 12000g for 30 min at 4 °C.

The soluble fraction was immediately loaded onto a 5 mL Ni²⁺ affinity column HisTrap HP (GE Healthcare) previously equilibrated with binding buffer A (20 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl, and 45 mM imidazole). The column was washed with binding buffer A, and the protein was recovered using an increasing imidazole gradient obtained with elution buffer A (20 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl, and 500 mM imidazole). The eluted fractions were analyzed by Coomassie blue stained 15% SDS-PAGE. Fractions containing the recombinant protein were pooled, desalted with a HiPrep 26/10 desalting column (GE Healthcare) previously

equilibrated and eluted with the *final buffer* (50 mM Tris-HCl buffer, pH 7.4, and 50 mM NaCl).

The 6xHis-tag was removed by adding TEV protease 5% w/w and 1 mM dithiothreitol (final concentration) directly to the protein solution, and the reaction was carried out for 16 h at 4 °C. The resulting sample was loaded onto a 5 mL Ni²⁺ affinity column HisTrap HP for a second purification step by the procedure described above using *binding buffer B* (50 mM Tris-HCl buffer, pH 7.4, 50 mM NaCl, and 30 mM imidazole) and *elution buffer B* (50 mM Tris-HCl buffer, pH 7.4, 50 mM NaCl, and 500 mM imidazole).

The resulting NahK sample was concentrated, whenever required, by ultrafiltration using a Vivaspin sample concentrator (10 kDa MWC, GE Healthcare), and then loaded onto a Superdex 200 10/300 GL column (GE Healthcare) for a final gel-filtration purification step using the *final buffer*. The peak fractions were pooled and the sample purity was verified by Coomassie blue stained 12–15% SDS-PAGE. The protein concentration at each purification step was determined from the absorbance at 280 nm using the appropriate extinction coefficient. The 6xHis-NahK and tag-free NahK molar extinction coefficients (21 430 and 19 940 M⁻¹ cm⁻¹, respectively) were estimated from the amino acid composition by the ProtParam tool¹⁵ using the ExPASy Server.¹⁶

Dynamic Light Scattering and Circular Dichroism Analysis. The hydrodynamic radius of NahK molecules and the polydispersity of the purified sample were estimated using the dynamic light scattering (DLS) assay performed on a Zetasizer Nano ZS system (Malvern Instruments). NahK samples were centrifuged at 15000g for 10 min at 4 °C prior to any measurement. The final protein concentration for DLS measurements was about 1 mg/mL in *final buffer*. Each sample was measured in a 10 mm path length quartz cuvette at 25 °C. Data were collected and analyzed using the Malvern Zetasizer software v. 6.01, where the “Protein Utility” routine was used to estimate particle mass based on the hydrodynamic radius.

The secondary structure content of NahK and its thermal stability were analyzed in a circular dichroism (CD) spectroscopy assay with a Jasco J-815 spectropolarimeter (Jasco Corporation). The final protein concentration for CD measurements was 0.2 mg/mL in buffer containing 5 mM Tris-HCl pH 7.4 and 5 mM NaCl. Far UV CD spectra (190–260 nm) were collected using quartz cuvettes (1 or 0.5 mm path length) in the temperature range 20–80 °C (at 10 °C intervals). Deconvolution of CD spectra was performed using the software CDNN v.2.1.¹⁷

Kinetic Assays. The NahK decarboxylation activity at 25 °C was monitored by following the rate of appearance of the product 2-hydroxy-2,4-pentadienoate at 265 nm ($\epsilon = 7850 \text{ M}^{-1} \text{ cm}^{-1}$) in an Agilent 8453 UV–vis spectrophotometer (Agilent Technologies). The reaction mixtures consisted of 20 mM sodium phosphate buffer in the pH range 6–8, 5 mM MgCl₂, and 0.5 $\mu\text{g/mL}$ of NahK. The reaction was initiated by addition of the substrate (1.6–48.0 μM) in a final reaction volume of 1.0 mL. The stock solution of substrate (2-oxo-3-hexenedioate) was generated enzymatically by the action of 4-oxalocrotonate tautomerase (XylH⁶) on 2 mM 2-hydroxymuconate. The 2-hydroxymuconate was previously dissolved in ethanol to 20 mM then made up in 20 mM sodium phosphate buffer pH 7.4. The V_{max} and K_{M} values, and their standard errors, were estimated by nonlinear regression of the initial rate of the reactions versus the concentration of the substrate to the

Michaelis–Menten equation using the software GraFit (Eratich Software Ltd., Staines, U.K.).

Crystallization, Data Collection, and Structure Determination. Pure and monodisperse protein samples were concentrated to approximately 10–12 mg/mL using a Vivaspin sample concentrator (10 kDa MWC, GE Healthcare) and then used in crystallization trials with sparse-matrix screens¹⁸ at 18 °C. The hanging-drop vapor-diffusion method was used, and 98 precipitant solutions from Crystal Screen 1 and 2 crystallization kits (Hampton Research) were tested. Crystallization drops consisting of 1 or 2 μL of protein solution mixed with 1 μL of each screening solution were equilibrated against 1 mL reservoir of the same screening solution. Conditions where protein crystals formed were manually optimized by grid screening of the precipitant concentration and pH of the crystallization conditions.

NahK crystals obtained from crystallization conditions were harvested from the crystallization drop with a nylon loop and immersed for a few seconds into a cryocooling solution (reservoir solution with 10% v/v ethylene glycol final concentration). Protein–ligand complexes were obtained by soaking different ligands (6–9, Figure 1) into apo crystals. The Mg²⁺/NahK and Mg²⁺/ α -ketoglutarate/NahK complexes were obtained when apo crystals were immersed for almost 1 min into a cryocooling solution containing 0.4 M MgCl₂ and/or 0.2 M α -ketoglutarate. Alternatively, a solution containing 100 mM MgCl₂ and 15 mM of adipate or 2-oxoadipate was added to the crystallization drop (after crystal growth) for the Mg²⁺/adipate/NahK and Mg²⁺/2-oxoadipate/NahK complexes, respectively. The Ca²⁺/acetate/NahK complex was obtained using only the screening solution in the crystallization trials. Once prepared, all crystals were flash-cooled to 100 K in a cold nitrogen stream prior to data collection.

In total, six X-ray diffraction data sets were collected at the Laboratório Nacional de Luz Síncrotron (LNLS, Campinas-SP, Brazil). Four of them were collected on a MarCCD 165 mm detector at the D03B-MX1 beamline¹⁹ while two were collected on a MarMosaic 225 mm detector at the W01B-MX2 beamline.²⁰ All data sets were processed using the HKL-2000 software.²¹ The crystal structures of NahK and the NahK complexes were solved by the molecular replacement (MR) method using the software Phaser from the CCP4 package.^{22,23} The template structure used for MR was a single monomer of the *E. coli* hydratase HpcG (PDB entry 2EB4), an enzyme which shares 41% amino acid sequence identity with NahK. Preliminary model building was performed using the software Buccaneer.²⁴ Cycles of structure refinement and manual model adjustments were conducted using the graphical interface of *phenix.refine* within the PHENIX package²⁵ and Coot,²⁶ respectively.

RESULTS AND DISCUSSION

Recombinant NahK Is Soluble and Stable in the Absence of NahL. In 1989, it was reported that XylI is physically associated with XylJ (a 4-OD and a VPH from TOL plasmid pWW0 present in *P. putida* mt-2), which share 94 and 90% sequence identity with NahK and NahL, respectively.⁴ In the same work, the authors expressed XylI in *E. coli* BL21(DE3) in the absence of XylJ, and partially purified the enzyme by anion-exchange chromatography and ammonium sulfate precipitation. This separately expressed and partially purified XylI showed low 4-OD activity, and the activity could not be recovered after passage through a hydrophobic interaction

chromatography column. The partially purified XylI rapidly lost activity when concentrated or stored at 4 °C. Attempts to stabilize the enzyme using purification buffers containing Mg^{2+} and glycerol were also not successful.^{6,12}

In the present work, we used a different strategy for expression and purification than previously described. The *nahK* gene from NAH7 plasmid present in *P. putida* G7 was subcloned into a pET28a-TEV expression vector, and transformed *E. coli* BL21(DE3) cells were used for protein expression. The expression of the protein was followed by SDS-PAGE analysis (Figure S1 A1, lanes 0h and 5h). The recombinant 6xHis-NahK enzyme was expressed as a partially soluble molecule (Figure S1 A1, lanes S and I). The soluble fraction was purified by Ni^{2+} -affinity chromatography, and the 6xHis-tag was cleaved by recombinant TEV protease. The resulting NahK sample was purified by another Ni^{2+} -affinity chromatography step, and the protein was recovered in the flow-through. Subsequently, NahK was purified to homogeneity by size-exclusion chromatography (Figure S1 A2, lane GF), where the elution volume of the main peak indicated a protein with molecular weight of approximately 25–30 kDa (Figure S1B), suggesting that NahK behaves as a monomer in solution. Using this methodology a stable form of NahK can be expressed in the absence of NahL. The enzyme can be concentrated and stored at 4 °C, and it retains decarboxylation activity during and after the purification steps.

NahK and XylI share 94% amino acid sequence identity where 17 out of 264 amino acid residues are different. Although most of the amino acid substitutions are equivalent in size and/or charge, 15 of them are located on the protein surface with their side chains interacting with the solvent. Thus, despite the high identity, this slight difference between XylI and NahK could be responsible for a higher solubility and stability of NahK compared to XylI. Another possible explanation for this behavior might be the use of a milder purification protocol that does not involve precipitation with ammonium sulfate or hydrophobic interaction chromatography.

Oligomeric State and Secondary Structure Content.

The oligomeric state of NahK was estimated by DLS. The Figure S1C presents particle-size distribution by mass, indicating that the sample was mainly composed of a monodisperse particle with 9.7% of polydispersity, and with a hydrodynamic diameter of 4.7 nm. Mass estimation based on hydrodynamic radius suggests a protein with molecular mass of approximately 26 kDa, indicating that recombinant NahK behaves as a monomer in solution. This result is in good agreement with the size-exclusion chromatography profile, corroborating the monomeric state of NahK in solution when expressed as a single component without NahL.

A search in the RCSB Protein Data Bank performed by the DALI server^{27,28} using the NahK model obtained in this work identified HpcG (2-oxo-hept-3-ene-1,7-dioate hydratase, PDB entry 2EB4) and MhpD (2-hydroxypentadienoic acid hydratase, PDB entry 2WQT) as the two most similar proteins. They share 41 and 34% amino acid sequence identity with NahK, respectively. It has been reported that MhpD forms pentamers in solution, while the biological unit of HpcG is a homodecamer made up of two pentamers similar to MhpD. Since recombinant NahK is a monomer in solution, the presence of NahL would be essential to form any quaternary structure, as observed in XylI and XylJ.

The secondary structure content of NahK was characterized by CD and confirmed later by the crystal structure.

Deconvolution of the data provided an estimated content of 37% α -helices, 31% β -strands, and 31% random coils (Figure S1D). According to the STRIDE server,²⁹ this result is in good agreement with the secondary structure of HpcG which shows 34% helical and 32% β -strand content, and with MhpD, which shows 32% α -helices and 31% β -strand structures, reflecting the similar folds among these proteins. CD spectra collected at different temperatures suggest NahK is a relatively stable molecule with a melting temperature between 60 and 70 °C (Figure S1D inset).

Kinetic Assays of NahK in the Absence of NahL. Five reaction conditions were examined to determine the effect of pH on the decarboxylase activity of recombinant NahK. The steady-state kinetic values obtained for each condition are shown in Table 1. Two trends can be seen. First, K_M decreases

Table 1. Variation of the Steady-State Kinetics of NahK with pH

pH	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1} s^{-1}$)
6.0	51.7 ± 14.2	22.5 ± 4.1	4.4×10^5
6.5	64.7 ± 8.5	108.4 ± 9.8	1.7×10^6
7.0	22.2 ± 2.4	104.2 ± 5.5	4.7×10^6
7.4	19.9 ± 1.1	98.8 ± 2.7	5.0×10^6
8.0	4.9 ± 0.3	21.8 ± 0.3	4.5×10^6

with increasing pH. This likely reflects the ionization of substrate and its tighter binding (assuming K_M reflects a dissociation constant). It could also result from the titration of the Mg^{2+} -binding residues and the tighter binding of Mg^{2+} . Second, k_{cat} and k_{cat}/K_M are highest in the range of 7.0–8.0. The steady-state kinetic values determined for XylI in the XylI/XylJ complex were $K_M = 6.6 \pm 0.6 \mu M$, $k_{cat} = 68 \pm 2 s^{-1}$, and $k_{cat}/K_M = 1.0 \times 10^7 M^{-1} s^{-1}$; and for the XylI/E106QXylJ mutant, in which XylJ is devoid of hydratase activity, $K_M = 13 \pm 1 \mu M$, $k_{cat} = 67 \pm 2 s^{-1}$, and $k_{cat}/K_M = 5.1 \times 10^6 M^{-1} s^{-1}$ (at pH 7.3 and 5 mM $MgCl_2$).¹² Hence, the steady-state kinetics parameters for NahK are similar to those obtained for XylI in the XylI/E106QXylJ complex. These observations suggest that the decarboxylase activity of NahK occurs separately from the NahL-catalyzed reaction. The Michaelis–Menten fit of the data and the ultraviolet–visible spectra for the general reaction catalyzed by NahK are presented in Figure S2.

Structure Determination of NahK and Its Complexes with Substrate Analogues. NahK samples were concentrated and crystallized using the hanging-drop method. After a few days, some nucleation was observed in drops consisting of NahK without the 6xHis-tag in condition #22 from Crystal Screen 1 crystallization kit (CS1–22:0.2 M sodium acetate trihydrate, 0.1 M Tris-HCl buffer, pH 8.5, and 30% w/v PEG 4000). Needle-shaped crystals suitable for diffraction (Figure S3A) were grown in a few weeks under optimized conditions derived from CS1–22 with a pH range from 8.5 to 10.5, and precipitant concentration of 26–30%. In addition, single rectangular-shaped crystals (Figure S3B) were observed after 4 months in drops consisting of 6xHis-NahK in condition #16 from Crystal Screen 1 crystallization kit (CS1–16:0.1 M Na HEPES buffer, pH 7.5, and 1.5 M lithium sulfate monohydrate). Crystals from CS1–16 and optimized CS1–22 conditions were used for ligand soaking and in X-ray diffraction experiments.

Three diffraction data sets of crystals obtained in CS1–16 condition were collected (Table 2, data sets 1–3). The first

Table 2. Summary of Statistics for All Diffraction Data Sets of the NahK Crystals^a

	1 - Apo model	2 - Mg model	3 - AKG model	4 - ACT model	5 - AA model	6 - 20A model
ligands		Mg ²⁺	Mg ²⁺ , α-ketoglutarate	Ca ²⁺ , acetate	Mg ²⁺ , adipate	Mg ²⁺ , 2-oxoadipate
PDB ID	5D2F	5D2G	5D2H	5D2I	5D2J	5D2K
Beamline	D03B-MX1	D03B-MX1	D03B-MX1	D03B-MX1	W01B-MX2	W01B-MX2
detector	MarCCD 165	MarCCD 165	MarCCD 165	MarCCD 165	MarMosaic 225	MarMosaic 225
wavelength (Å)	1.437	1.437	1.437	1.430	1.459	1.461
detector distance (mm)	60.0	80.0	80.0	70.0	80.0	80.0
rotation per image (deg)	1.0	1.0	1.0	1.0	0.7	1.0
total rotation range (deg)	180	99	90	300	200	280
space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁
unit cell param. (Å, °)	<i>a</i> = 43.8 <i>b</i> = 44.9 <i>c</i> = 125.1	<i>a</i> = 42.8 <i>b</i> = 45.0 <i>c</i> = 124.7	<i>a</i> = 43.7 <i>b</i> = 44.9 <i>c</i> = 125.3	<i>a</i> = 77.5 <i>b</i> = 43.8 <i>c</i> = 79.5 <i>β</i> = 118.2	<i>a</i> = 78.8 <i>b</i> = 44.5 <i>c</i> = 81.3 <i>β</i> = 118.8	<i>a</i> = 79.3 <i>b</i> = 44.4 <i>c</i> = 81.6 <i>β</i> = 118.7
mosaicity (deg)	0.5	0.7	0.9	0.4	1.0	0.4
resolution range (Å)	31.35–1.74 (1.78–1.74)	40.45–1.90 (1.96–1.90)	42.27–1.94 (2.00–1.94)	26.68–1.78 (1.82–1.78)	40.59–1.72 (1.76–1.72)	40.78–1.57 (1.60–1.57)
no. reflections	180,312	66,421	58,992	276,928	196,894	385,525
no. of unique reflections	25,510	18,966	18,138	45,433	51,584	69,868
redundancy	7.1 (6.3)	3.5 (2.7)	3.3 (2.5)	6.1 (5.9)	3.8 (2.9)	5.5 (4.2)
completeness (%)	97.2 (94.5)	95.7 (72.6)	94.8 (70.7)	100.0 (100.0)	97.3 (80.6)	99.9 (99.8)
<1/σ(<i>I</i>)>	20.2 (2.8)	10.2 (1.8)	9.9 (1.4)	29.3 (3.0)	15.4 (3.0)	29.7 (8.0)
<i>R</i> _{merge} ^b	8.0 (51.6)	10.8 (43.5)	10.4 (51.4)	6.0 (49.1)	7.5 (35.0)	5.4 (17.7)
Matthews coeff. (Å ³ Da ⁻¹)	2.00	1.95	2.00	2.05	2.16	2.18
solvent content (%)	38.6	37.1	38.6	40.1	43.1	43.6
<i>R</i> work (%)	17.0	15.4	16.4	14.5	15.4	14.8
<i>R</i> free (%)	21.2	20.4	23.2	18.6	19.1	17.1
no. of protein atoms	2,008	1,991	1,999	3,968	3,996	4,034
no. of water molecules	398	307	271	608	525	651
no. of ligands ^c	0	1	2	4	4	4
average <i>B</i> factor (Å ²)						
protein atoms	15.9	15.4	22.0	17.8	18.8	15.8
ligand atoms ^c		9.15	29.9	17.8	25.0	22.3
water molecules	29.6	25.8	29.8	29.3	30.1	27.8
overall	18.5	17.0	22.6	18.9	20.2	17.3
RMSD from ideal						
bond angles (deg)	1.174	1.206	1.097	1.175	1.059	1.099
bond lengths (Å)	0.007	0.008	0.007	0.007	0.006	0.007

^aValues in parentheses are for the highest resolution shells. ^b $R_{\text{merge}} = \sum_{hkl} \sum_i [I_i(hkl)] / \sum_{hkl} [I_i(hkl)] - \langle I_i(hkl) \rangle / \sum_{hkl} [I_i(hkl)] \times 100$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl* and $\langle I_i(hkl) \rangle$ is its average. ^cRefers to the ligands (magnesium, calcium and substrate analogs) inside the active site of the enzymes in the asymmetric unit of the crystal.

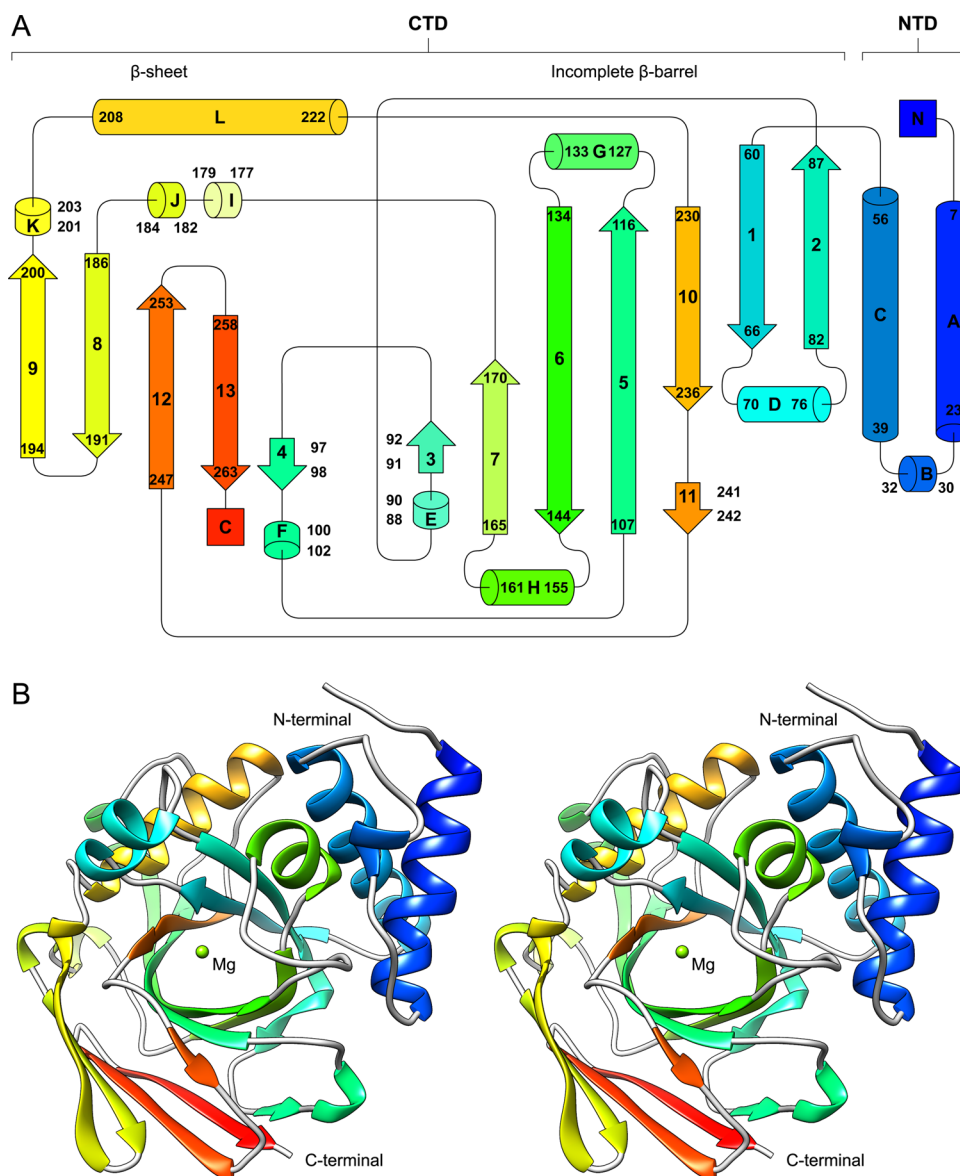


Figure 2. Different representations of the overall fold of NahK. Topology diagram (A) and a cross-eyed stereo cartoon of NahK in complex with magnesium (B). (A) Helices are shown as tubes and labeled alphabetically, and β -strands are shown as arrows and labeled numerically. Color shades from blue to red represent the progress of the chain from N to C-terminus. (B) Mg^{2+} cation (green sphere) bonded to the protein (colored as in A) lies at the center of the incomplete β -barrel structure.

crystal contains the apoenzyme (*Apo* model, 1.74 Å resolution), the second crystal was immersed into a soaking solution containing magnesium ions (*Mg* model, 1.90 Å resolution), and the third crystal was soaked in the magnesium solution and then soaked in a solution with α -ketoglutarate (*AKG* model, 1.94 Å resolution). These crystals belong to the orthorhombic $P2_12_12_1$ space group with average unit-cell parameters of $a = 43.4$, $b = 45.0$, $c = 125.0$ Å. MR for all these data sets was performed successfully using one molecule of HpcG as the search model.

Three other data sets were collected using NahK crystals grown in the optimized CS1–22 condition (Table 2, data sets 4–6). The first data set refers to a crystal in which an acetate ion was observed in the catalytic site of the enzyme (*ACT* model, 1.78 Å resolution), the second one is from a crystal soaked with magnesium and adipate (*AA* model, 1.72 Å resolution), while the last data set was collected with a crystal soaked with magnesium and 2-oxoadipate (*2OA* model, 1.57 Å

resolution). These crystals belong to the monoclinic $P2_1$ space group with average unit cell parameters of $a = 78.5$, $b = 44.2$, $c = 80.8$ Å, $\beta = 118.6^\circ$. MR using a single monomer of HpcG as the search model yields two best solutions and both were used to construct the NahK/ligand models.

The final models display a good geometry with an average root-mean-square deviation (rmsd) of 0.007 Å for bond lengths and 1.14° for angles. On average, considering all polypeptide chains in the six structures, 96.6% (255/chain) and 3.0% (8/chain) of the amino acid residues are located in the favored and allowed regions of the Ramachandran plot, respectively.

Overall Structure of NahK. NahK is a member of the FAH superfamily. Initially, the crystal structure of the mouse FAH was solved (PDB entry 1QCN³⁰) and described as having a unique fold, which is conserved from bacteria to mammals and defined the overall scaffold of the family. The FAH superfamily has grown considerably since this first report and now includes several other proteins with diverse enzymatic functions such as

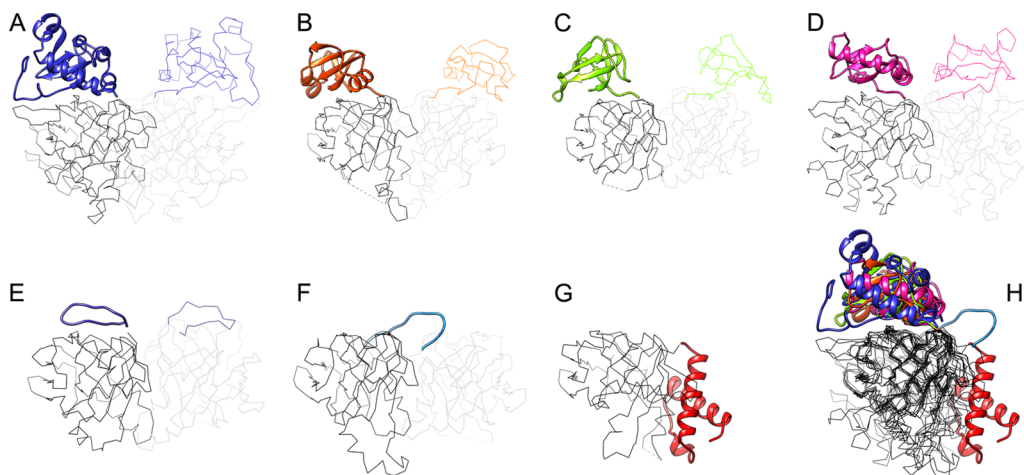


Figure 3. Comparative analysis of the N-terminal domains of FAH superfamily members. (A–G) All 21 FAH members available in the Protein Data Bank to date were grouped according to the overall structure of their N-terminal domains. Only one member of each cluster (quite often a dimeric protein) was chosen to represent the overall fold of the group. Representative tertiary and quaternary structures of enzymes from (A) *Mus musculus* (PDB entry 1QCN³⁰) and *Burkholderia cenocepacia* (PDB entry 4QKU); (B) *Sulfolobus solfataricus* (PDB entry 2Q18³²); (C) *Mycobacterium marinum* (PDB entry 3QDF³⁹), *Thermus thermophilus* (PDB entry 2DFU and PDB entry 1WZO), *M. abscessus* (PDB entry 3RR6³⁹), *Corynebacterium glutamicum* (PDB entry 4DBF³³), and *M. smegmatis* (PDB entry 4PFZ); (D) *M. abscessus* (PDB entry 3R6O³⁹) and *Sinorhizobium meliloti* (PDB entry 3LZK); (E) *Homo sapiens* (PDB entry 1SAW³¹), *E. coli* (PDB entry 1NR9), *Yersinia pestis* (PDB entry 3SS2), *Oleispira antarctica* (PDB entry 3V77⁴⁰), *B. cenocepacia* (PDB entry 4MAQ) and *Saccharomyces cerevisiae* (PDB entry 1NKQ); (F) *E. coli* HpcE (PDB entry 1GTT¹⁰) and (G) *E. coli* HpcG (PDB entry 2EB4¹¹), *E. coli* MhpD (PDB entry 2WQT⁹) and *P. putida* NahK (this work). (H) Superposition of representative monomers showing the conserved CTD (incomplete β -barrel and β -sheet) and the position of each NTD.

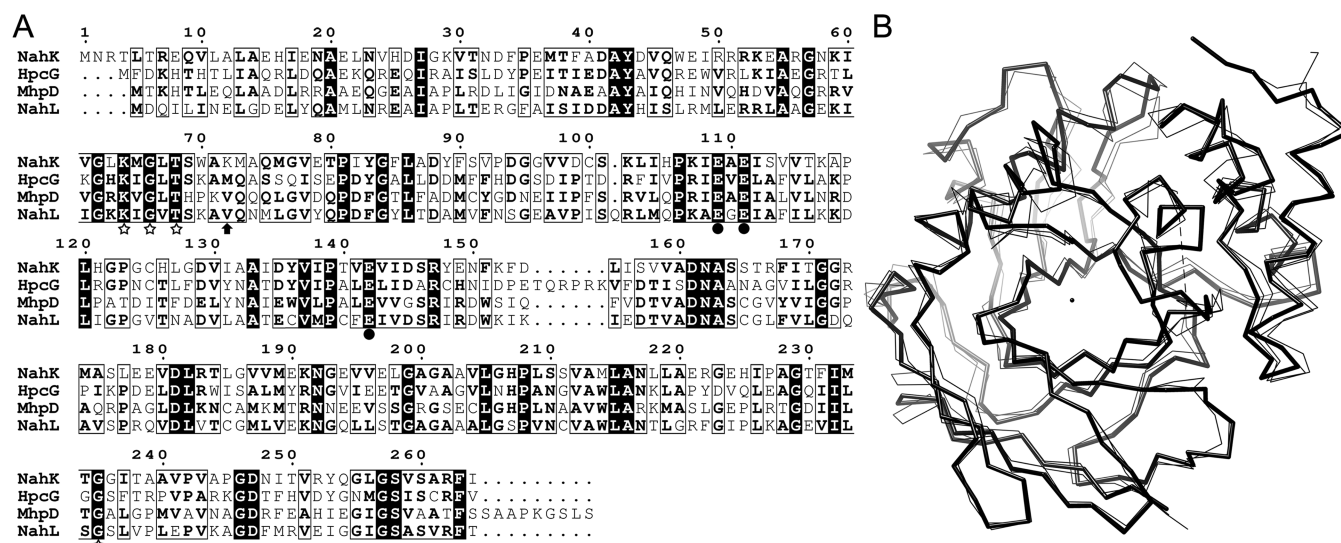


Figure 4. Multiple alignments between NahK and similar proteins. (A) Sequence alignment of NahK, HpcG, MhpD, and NahL. The glutamate residues that coordinate the metal ion are indicated with black circles, while other conserved residues at the active site are indicated with a star. Lys72, a specific residue in the catalytic site of NahK, is indicated with an arrow. The figure was prepared using ESPrpt 3.0.⁴¹ (B) Superposition of the C α backbone of NahK (thickest line) with HpcG and MhpD (thinnest line). The dot at the center of the structure shows the position of the metal ion.

hydrolysis,³¹ isomerization,¹⁰ dehydration,³² hydration,^{9,11} and decarboxylation.^{10,33}

The FAH superfamily scaffold comprises two main domains; the N- and C-terminal domains. The amino-terminal domain (NTD), consisting of the first 120 amino acid residues, is characterized by an α/β -sandwich structure. The carboxy-terminal domain (CTD), made up of the last 300 residues, is divided into a barrel-like (incomplete β -barrel) and a β -sheet structure. Despite the very low sequence identity between NahK and mouse FAH (9% identity), the C-terminal domains show similar folds, which is unique to the FAH superfamily.

The major differences are observed in the N-terminal domains. Contrary to the mouse FAH, the NTD of NahK, with only 58 residues, folds into an helical domain composed of one short and two long helices. A topology diagram and a ribbon representation of the overall fold of NahK are shown in Figure 2A,B.

HpcG, MhpD, and probably NahL also share a similar fold with NahK. The NTDs are similar to each other and differ from other FAH members (Figure 3). A sequence alignment of NahK, HpcG, MhpD, and NahL is shown in Figure 4A. Interestingly, while MhpD is a homopentamer in solution and

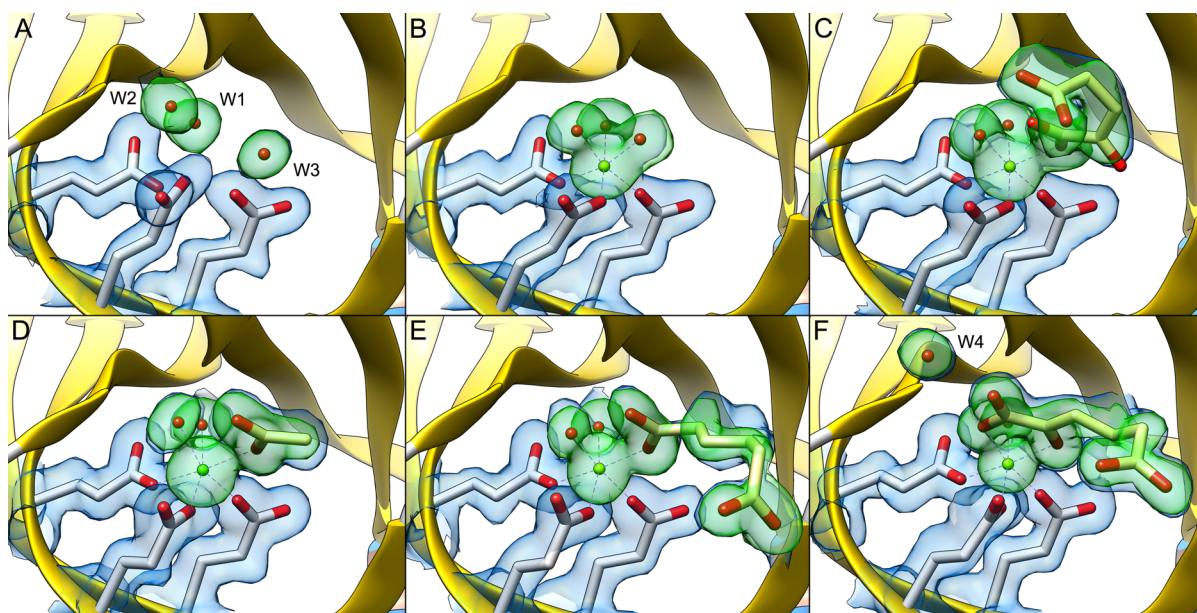


Figure 5. Electron density maps illustrating the position of metals (green spheres), water molecules (red spheres), and ligands (yellow and red stick-models) at the active site of crystallized NahK. The $2F_o - F_c$ map (blue) and omit map (green) are contoured at 1.0 and 4.0 σ , respectively. Water molecules W1, W2, W3, and W4 are named and described in the text. (A) Apo model: active site with the electron density map of three water molecules sharing hydrogen bonds with three glutamate residues. (B) Mg model: active site with a magnesium ion coordinated by three glutamate residues and three water molecules. (C) AKG model: in addition to the presence of one magnesium ion, the electron density above the metal suggests the presence of an α -ketoglutarate molecule. Two water molecules also coordinate the metal. (D) ACT model: the electron density suggests the presence of a metal ion being coordinated by the glutamate residues. An acetate ion from the crystallization buffer is also present along with two water molecules coordinating the metal. (E) AA model: the electron density shows that one magnesium ion and one adipate molecule are present in the active site. Two water molecules coordinate the metal. (F) 2OA model: the electron density from a 2-oxoadipate molecule is observed near the magnesium ion. One water molecule (W1; behind the plane of the ligand) is interacting with the metal.

HpcG is a dimer of homopentamers, NahK adopts a monomeric form. NahK might only assemble into a complex in the presence of NahL. As previously reported for HpcG and MhpD, the NTDs are not involved in oligomerization. They face the solvent without any contacts with neighboring chains. This observation suggests that the NTDs of NahK and NahL are not involved in NahK/NahL complex assembly. We propose that the NTD of NahK is involved in the stabilization of the barrel-like structure in the CTD through interactions between several residues from the β -barrel (located in strand 2, helices E and H, and loop 6H) and the NTD. Hydrogen bonds and hydrophobic interactions are responsible for a buried surface area of 1325 \AA^2 between these domains. Thus, in contrast to what has been proposed for HpcG,¹¹ deletion of the NahK NTD might destabilize the incomplete β -barrel and compromise the folding and activity of NahK. As we have observed, residues adjacent to helix H (in the β -barrel) compose the loop which forms the entry of the active site and adopts different conformations upon ligand binding—suggesting that destabilization of this region would decrease (or eliminate) the enzyme activity. Even with small differences in some surface loops around their active sites, it is possible to superpose HpcG, MhpD, and NahK monomers (246 C^α atom pairs superposed) with a rmsd close to 1.3 \AA . The superposed C^α traces are shown in Figure 4B.

A complete survey of all available crystal structures of FAH superfamily members, excluding HpcG and MhpD, revealed that they primarily adopt a dimeric form as the quaternary structure in which the incomplete β -barrel edges of each monomer interact with each other generating a mean buried surface area of 1,665 \AA^2 (varying from 1103 to 2778 \AA^2). As

shown in Figure 3, the equivalent position of the second incomplete β -barrel in these homodimeric structures is occupied by the NTD of NahK, HpcG, and MhpD. This observation is consistent with our proposal that these specific NTDs stabilize the barrel and allow a different quaternary arrangement in the superfamily.

In the MhpD homopentamer, helices G and L (using the NahK nomenclature) in the CTD provide the amino acid residues forming the interface between monomers. In the HpcG homododecamer, these same interactions are observed within each pentamer. In addition, the dimer of pentamers is constructed primarily by interactions between the β -sheet structures (CTD) present in each ring.

The similar fold of NahK and HpcG, and the sequence identity between NahK and NahL raise the possibility that NahK and NahL might interact like the subunits of HpcG (homododecamer) to form a heterododecamer structure. If so, the β -sheets at the CTD of NahK and NahL would be involved in oligomerization. To the best of our knowledge, this 280 kDa heterododecamer ($5 \times \text{NahK} + 5 \times \text{NahL}$) has never been proposed, although results first obtained in 1973 suggest such a possibility.⁵ Collinsworth and colleagues obtained a 287 kDa native complex from *P. putida* (determined by analytical ultracentrifugation) that was only tested for VPH activity. Even though they observed two protein bands in a SDS-polyacrylamide gel (27.6 and 28.0 kDa each), they believed both bands were derived from the same VPH enzyme, suggesting its native form was composed of 10 subunits. In 1989, an equivalent complex from *P. putida* mt-2 was reported that showed 4-OD and VPH activities and two bands in the SDS-PAGE were also observed (27 and 28 kDa). However, gel

filtration chromatography results were not conclusive about the size of the complex because a value of 130 kDa was obtained⁴ and the previous 287 kDa value obtained by Collinsworth was never explored.

A superposition of all NahK models obtained in this work results in a rmsd of 0.55 Å between C^α atom pairs, with residues from helix D and loop 6H (between strand 6 and helix H; residues 149–154) displaying the highest values up to 9.5 Å. These residues are located at the sides of the cavity that lead to the active site and could be subjected to conformational changes upon ligand binding. In addition, the same residues have the highest B-factors among the other surface residues of the protein, confirming their predisposition to a greater mobility. Not surprisingly, the same regions in HpcG and MhpD also have the highest B-factors or are not visible in the models.

The Active Site of NahK. 4-OD is a metal-dependent enzyme requiring magnesium ion for activity. Johnson and colleagues identified conserved and potential catalytic residues in XylI and homologues, and constructed XylI mutants that were expressed and purified as a complex with XylJ (VPH).¹³ Two mutants (E109Q and E111Q) lacked the decarboxylase activity of 4-OD, but retained the hydratase activity of VPH. It has been proposed that Glu109 and Glu111 (XylI or NahK numbering) might be involved in metal binding. The crystallographic model of NahK provides the structural basis for these observations and assigns roles to these and other residues.

In the *Apo* and *Mg* models the active site is identified as a large cavity approximately 15 Å deep and 19 Å wide. The bottom of this cavity is located in the top of the barrel-like structure (in the CTD). According to the CASTp server,³⁴ this cavity has an average volume of approximately 985 Å³. Analysis of the electron density map at the bottom of this cavity in the *Apo* model shows the presence of three water molecules sharing hydrogen bonds with the side chains of three glutamate residues, with no bound metal (Figure 5A). However, the *Mg* model clearly shows the incorporation of one Mg²⁺ at the active site (Figure 5B). Three water molecules are still present in this model and together with three oxygen^e atoms from residues Glu109, Glu111, and Glu142 bind to the metal in an octahedral geometry. One of these water molecules (W1, Figure 5A) is found in the active sites of all NahK models and is buried among the side chains of the protein. This water molecule is stabilized by hydrogen bonds with the backbone carbonyl oxygen of Met65 (2.6 Å), and the oxygen^e atoms from glutamates in the active site (2.7–2.9 Å). The two other coordinating water molecules (W2 and W3, Figure 5A) are displaced to allow the binding of a ligand.

In addition to the three acidic residues, the NahK active site is also composed of basic amino acids (Lys64 and Lys72). The nitrogen^c atom of Lys64 forms a salt bridge with one oxygen^e atom of Glu142 (2.6 Å) in all models. This same group makes an additional hydrogen bond (2.8 Å) with either a coordinating water molecule (W3) or an oxygen atom from the ligand (Figure 5C–F). This lysine is conserved in the same position in the active sites and/or amino acid sequences of HpcG, MhpD, and NahL (Lys61, Lys60, and Lys60, respectively; Figure 4A), suggesting it plays a role in catalysis of these metal-dependent enzymes.

Contrary to the conserved Lys64, Lys72 appears to be specific to NahK. While Lys64 interacts with the α -keto acid group of the ligand at the metal binding site, Lys72 is located in

the opposite side of the active site cavity, and provides a positively charged group to interact with the vinylous β -keto carboxylate group of the ligand. This charged amino acid is replaced by hydrophobic residues in HpcG, MhpD, and NahL (Met69, Val68, Val68; respectively) reflecting the different biological substrates.

The AKG, ACT, AA, and 2OA models show the ligands α -ketoglutarate, acetate, adipate, and 2-oxoadipate, respectively, in the enzyme active site (Figure 5C–F). All these ligands displace at least one coordinating water molecule (W3) to interact with the metal as a monodentate ligand. This interaction occurs between the metal and one oxygen atom of their carboxylate group. In the 2OA model, at least in one of the polypeptide chains, the ligand adopts a bidentate conformation in which the α -keto acid group displaces two water molecules (W3 and W2) to interact with the metal (Figure 5F). While one of the oxygen atoms from the first carboxylate group (C-1) interacts with Mg²⁺, the second one makes hydrogen bonds with the main chain amide of Leu67 (2.9 Å) and another water molecule (W4, 2.6 Å) trapped inside the active site. Because of the inability of NahK to decarboxylate the substrate analogue 2-oxoadipate (9, Figure 1), the conformation adopted by the Mg²⁺/2-oxoadipate/NahK complex could reflect what might happen to the substrate and enzyme just before decarboxylation. A detailed list of polar interactions between ligands and protein residues, metal ion and water molecules in the active site is presented in Table 3.

NahK Has a Lid Domain. Most protein chains in the NahK models show a large active site cavity with an average volume of 845 Å³, which is considered as an *open* conformation. This conformation is obtained either in the presence or absence of a

Table 3. Main Polar Interactions Observed between the Ligands and Protein Residues, Metal Ion and Water Molecules in the Active Site of NahK Complexes

chain:residue [atom]	distance (Å)	chain:residue [atom]
α -Ketoglutarate		
D:AKG [O1]	2.35	C:MG 1 [Mg]
D:AKG [O1]	2.76	A:LYS 64 [N ζ]
D:AKG [O3]	2.98	A:LEU 67 [N]
D:AKG [O5]	2.81	A:ALA 163 [O]
D:AKG [O2]	2.50	W:HOH 2 [O]
Acetate		
G:ACT 1 [O]	2.36	C:CA 1 [Ca]
G:ACT 1 [O]	2.86	A:LYS 64 [N ζ]
G:ACT 1 [OXT]	2.54	W:HOH 2 [O]
Adipate		
D:0L1 [O1]	2.19	C:MG 1 [Mg]
D:0L1 [O1]	2.95	A:LYS 64 [N ζ]
D:0L1 [O3]	2.67	A:SER 164 [O γ]
D:0L1 [O4]	3.64	A:LYS 72 [N ζ]
D:0L1 [O2]	2.57	W:HOH 2 [O]
2-Oxoadipate		
D:OOG 1 [O1]	2.09	C:MG 1 [Mg]
D:OOG 1 [O3]	2.20	C:MG 1 [Mg]
D:OOG 1 [O1]	2.74	A:GLY 236 [N]
D:OOG 1 [O2]	2.97	A:LEU 67 [N]
D:OOG 1 [O3]	2.75	A:LYS 64 [N ζ]
D:OOG 1 [O4]	2.79	A:SER 164 [O γ]
D:OOG 1 [O5]	2.71	A:LYS 72 [N ζ]
D:OOG 1 [O2]	2.67	W:HOH 2 [O]

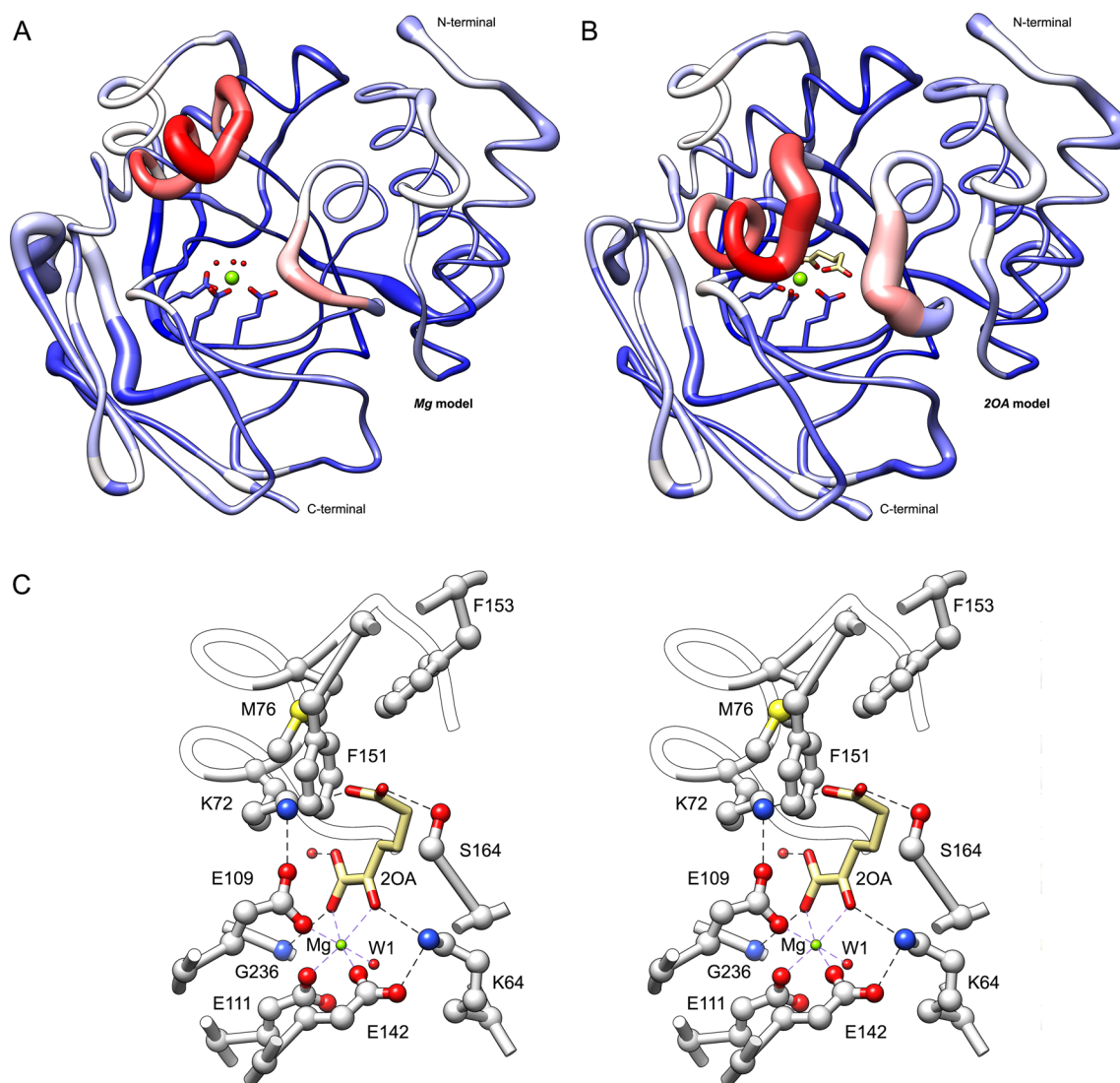


Figure 6. Conformational changes in NahK upon ligand binding. The *open* conformation of the Mg model (A) and the *closed* conformation of the 2OA model (B) are depicted as a “worm” model. The thickness of the chains is proportional to the B-factor values per residue in each model, indicating the parts of the protein with more thermal motion. The color shades from blue (rmsd value of 0.482 Å), to white (rmsd value of 1.0 Å), and red (rmsd value of 9.5 Å) represent the rmsd between C α pairs from the two models. Red color shades indicate the parts of the protein with the highest difference between the positions of the C α pairs. The three glutamate residues coordinating the magnesium ion are shown at the bottom of the active site in both models, and in the 2OA model (B) the 2-oxoadipate molecule is represented as a stick model. (C) A cross-eyed stereo representation of the catalytic site of the Mg $^{2+}$ /2-oxoadipate/NahK complex in the *closed* conformation is shown. Main hydrogen bonds, salt bridges, and metal coordination are represented by dashed lines. Helix D composing the lid domain is depicted as a helical tube. The α -keto acid moiety of 2-oxoadipate binds to Mg $^{2+}$ in a bidentate mode, while the second carboxylate group is positioned 70° outside the plane of the ligand, making it almost parallel to the π -orbitals of the vinyllogous ketone. The departing carboxylate group could be positioned in this conformation by Lys72 and Ser164. The hydrophobic residues Met76, Phe151 and Phe153 which might facilitate decarboxylation of the substrate are also shown. The figures were prepared using Chimera software.⁴²

magnesium ion. However, as observed in a few chains, NahK adopts a *partially closed* and a *closed* conformation upon ligand binding. In the *partially closed* conformation, observed in chain A from the AA model, the enzyme has a reduced active site volume of approximately 690 Å 3 caused by the slight movement of helix D (approximately 1.6 Å) toward the ligand. This movement allows the side chain of residue Lys72 to form a salt bridge with one oxygen atom from the second carboxylate group (C-6) of the adipate molecule. In addition, the second oxygen of this carboxylate group shares a hydrogen bond with the oxygen $^{\prime}$ from Ser164. Interestingly, the movement of helix D causes the rearrangement of the residues Phe151 and Phe153 in loop 6H.

One of the biggest changes in the structure is the movement of helix D in chain A of the 2OA model. In this conformation (*closed* conformation), helix D is moved significantly toward the ligand (maximum distance of 9.5 Å) and together with loop 6H fills the rest of the active site cavity, reducing its volume to approximately 232 Å 3 and shielding the ligand from the aqueous environment. Figure 6A shows the difference in the positions of helix D and loop 6H in the *open* and *closed* conformations. The 2-oxoadipate ligand lacks a double bond between C-3 and C-4 atoms when compared to 2-oxo-3-hexenedioate (9 and 3, Figure 1). Thus, being more flexible, this substrate analogue may represent the conformation of the substrate in the active site just before decarboxylation. Among

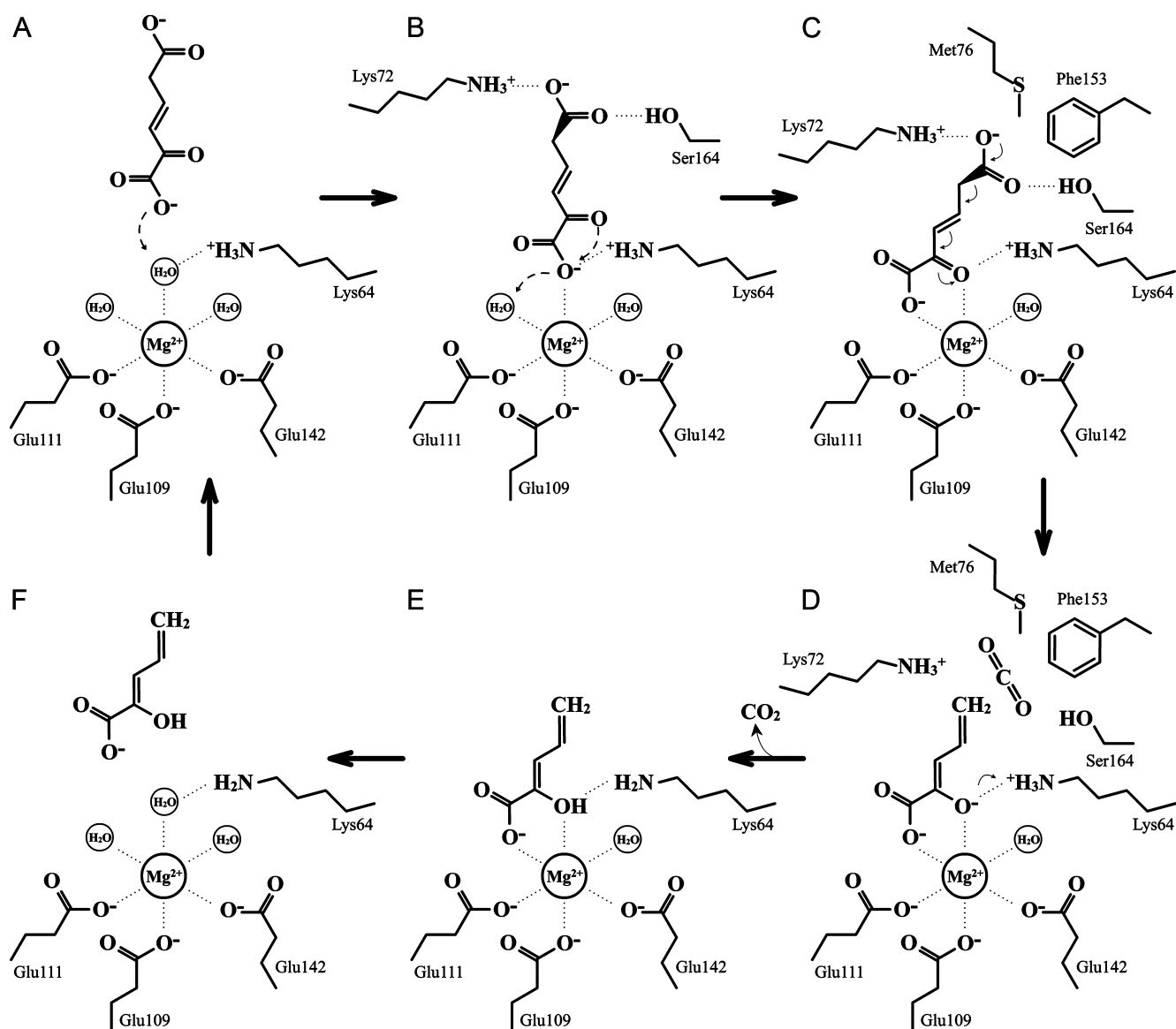


Figure 7. Proposed decarboxylation mechanism for NahK. The (A) *open*, (B) *partially closed*, and (C, D) *closed* conformations of the substrate/enzyme complex are presented with major catalytic residues. (D) Decarboxylation of the dienolate form of the intermediate takes place in the catalytic site with a subsequent (E) formation of the dienol form of the product. (F) The product is released from the enzyme catalytic site allowing a new decarboxylation cycle to occur.

the ligands examined, this was the only one that displaced the second coordinating water molecule (W2) and interacted with the metal in a bidentate mode through two oxygen atoms from the α -keto acid group (Figure SF).

In the *closed* conformation, helix D brings the side chain of Lys72 even closer to the ligand, making it possible to form a hydrogen bond with oxygen^e of Glu109 (2.5 Å), and a salt bridge with one oxygen atom from the second carboxylate group (C-6) of the ligand (2.8 Å) (Figure 6C). The second oxygen of the C-6 carboxylate group forms a hydrogen bond with the oxygen^f from Ser164. In addition to these conformational changes, the side chain of Met76, which also moves toward the ligand, is placed together with Phe151 and Phe153 in the vicinity of the C-6 carboxylate group of the ligand, forming a hydrophobic environment favorable for decarboxylation. These three hydrophobic residues are, in average, 3.3 Å apart from the C-6 carboxylate group of 2-oxoadipate.

The corresponding region of helix D is missing in most crystal structures of FAH superfamily members.^{10,11,31} However, in models showing the enzyme bound to inhibitors or substrate analogs,³⁰ this region is visible and closes the top of the active site cavity. Therefore, our results reinforce the proposal that this region functions as a lid in FAH superfamily.^{32,33} The availability of three different conformational states (*open*, *partially closed*, and *closed*) for NahK allowed us to simulate what might be the movement of helix D and loop 6H upon substrate binding, highlighting the coordinate rearrangement of catalytic residues to trap and orientate the substrate in the active site just prior decarboxylation (Video S1 and S2).

The Decarboxylation Mechanism of NahK. To date, there are three characterized decarboxylases that belong to the FAH superfamily: 4-oxalocrotonate decarboxylase (NahK, XylI and other homologues), 5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase (HpcE¹⁰), and oxaloacetate decarboxylase

(Cg1458³³). Ran and colleagues proposed that Cg1458 uses a glutamic acid-histidine-water catalytic mechanism to facilitate decarboxylation. Although NahK and Cg1458 belong to the same superfamily and utilize Mg^{2+} as a cofactor, NahK lacks the equivalent histidine residue in the catalytic site. Moreover, our structural data suggest that the only water molecules available in the active site of NahK after lid closing are the water molecules W1 and W4, and both are too far from C-6 of the substrate to interact with it. HpcE and NahK have very similar substrates, with the only difference being a carboxymethyl group attached to C-5 in HpcE's substrate. However, a catalytic mechanism for HpcE is still unclear. The metal is proposed to only hold the residues of the active site in place and not to play a role in decarboxylation.¹⁰

Various decarboxylation mechanisms have been described. As NahK does not use any cofactor other than a divalent ion, decarboxylation mechanisms dependent on pyridoxal-5'-phosphate, biotin, flavin, $NAD^+/NADP^+$, and other cofactors can be excluded.^{35,36} A Schiff base mechanism for decarboxylation has also been eliminated.⁶ In 2000, Stanley and colleagues suggested the 4-OD mechanism could proceed in a similar manner to the metal ion catalyzed decarboxylation of oxaloacetate,³⁷ in which the metal ion forms a complex through the α -keto acid group of the substrate and acts as an electron sink to stabilize the enolate intermediate and facilitates the decarboxylation. Taken together with our crystallographic data, we propose the enzymatic reaction mechanism described below and shown in Figure 7.

In the *open* conformation of NahK (Figure 7A), the substrate is allowed to enter into the large active site cavity, displacing one coordinating water molecule (W3) in order to form a monodentate interaction with the metal through one oxygen atom from the C-1 carboxylate group. In this monodentate conformation, the complex adopts a *partially closed* conformation in which helix D moves slightly toward the ligand allowing the side chain of Lys72 to form a salt bridge with one oxygen atom from the C-6 carboxylate group of the substrate. The second oxygen atom from this C-6 carboxylate group is stabilized by the side chain of Ser164 (Figure 7B). Displacement of a second coordinating water molecule (W2) allows the substrate to form a bidentate interaction with Mg^{2+} through its α -keto acid group. In this bidentate interaction, the substrate moves deeper into the active site allowing the movement of helix D to the *closed* conformation. This conformation places Met76, Phe151, and Phe153 in the vicinity of the departing C-6 carboxylate group, which is positioned by residues Lys72 and Ser164 almost perpendicular (approximately 70°) to the main plane of the substrate. The hydrophobic environment created by Met76, Phe151, and Phe153 and the geometry adopted by the departing carboxylate group favor decarboxylation. At the same time, the C-2 carbonyl oxygen interacts directly with the metal and the nitrogen⁶ atom of Lys-64, favoring the enolate form of the substrate (Figure 7C). The formation of the enolate moves the position of the double bond in the conjugated ketone from C-3/C-4 to C-2/C-3 with the consequent destabilization of the carbon-carbon bond between C-5 and C-6 atoms, transforming it into a dienolate and resulting in the loss of CO_2 (Figure 7D).³⁸ Finally, Lys64 donates a proton to the enolate oxygen to form the enol tautomer, which corresponds to the product of NahK (Figure 7E). The helix D comes back to its *open* conformation, releasing the products and allowing the enzyme to perform a new catalytic cycle (Figure 7F).

CONCLUSION

In this work we present the successful expression, purification, and crystallization of an active 4-OD separately from VPH (NahK and NahL from the NAH7 plasmid, respectively). The secondary structure content of NahK is similar to HpcG and MhpD, and although these enzymes form large quaternary structures, recombinant NahK behaves as a monomer in solution, suggesting the interaction with NahL could be essential to oligomerization. The kinetic parameters confirmed that the activity of NahK does not depend on its assembly with NahL. Moreover, the similarity of the k_{cat}/K_M value to that of XylI in the native XylI/XylJ complex ($\sim 10^7 M^{-1} s^{-1}$) suggests that a NahK/NahL complex formation may not substantially improve the activity of the NahK enzyme, but might be required for correct processing of NahK unstable product by NahL. The apo form of the enzyme and different complexes with substrate analogues were crystallized and X-ray diffraction data sets were collected up to 1.57 Å resolution. This allowed the determination of six three-dimensional models of NahK.

A comparison among all available crystal structures of FAH superfamily members revealed that the N-terminal domain of NahK stabilizes the incomplete β -barrel and allows a different quaternary arrangement for this molecule. Contrary to most FAH superfamily members which adopt a dimeric form, NahK is a monomer in solution and might form a heterodecamer complex ($5 \times NahK + 5 \times NahL$) with NahL in a similar arrangement observed for HpcG.

In addition, our results suggest that the enzyme adopts at least three particular conformations (*open*, *partially closed*, and *closed*) depending on the type of the ligand and the way it is positioned in the catalytic site. Presumably, the decarboxylation reaction performed by NahK requires that the substrate (with six carbon chain) adopt a bidentate interaction with the metal to move the substrate deeper into the active site and allow a more pronounced movement of helix D and loop 6H. The enzyme adopts a *closed* conformation that seals off the active site from the outside environment and generates a proper geometry and hydrophobic environment for decarboxylation. The presence of this lid domain seems to be a conserved feature of the FAH superfamily enzymes.

The enzymatic catalytic mechanism proposed in this work represents a new decarboxylation strategy for the FAH superfamily and is the first one described for a vinylogous β -keto acid metal dependent decarboxylase. Because of the similarities between NahK and HpcE and their substrates, it is likely that HpcE uses the same decarboxylation mechanism.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.6b00050.

SDS-PAGE analysis of NahK expression and purification (Figure S1A); Size-exclusion chromatography profile of NahK (Figure S1B); Dynamic light scattering measurements of NahK (Figure S1C); Secondary structure analysis and thermal stability of NahK by Circular dichroism spectroscopy (Figure S1D); Fit of the data to the Michaelis-Menten equation of the measured rates of the decarboxylation at 25 °C as a function of pH (Figure S2A); Ultraviolet-visible spectra for the decarboxylation of 2-oxo-3-hexenedioate by NahK in the absence of

NahL (Figure S2B); NahK crystals obtained from CS1–16 and CS1–22 conditions (Figure S3A/B); X-ray diffraction pattern for the Mg²⁺/2-oxoadipate/NahK complex crystal (Figure S3C) (PDF)

Video S1 (AVI)

Video S2 (AVI)

Accession Codes

The atomic coordinates and structure factors obtained in this work have been deposited in PDB (PDB entry 5D2F for the apo form of NahK and entry 5D2G for the Mg²⁺/NahK, entry 5D2H for the Mg²⁺/ α -ketoglutarate/NahK, entry 5D2I for the Ca²⁺/acetate/NahK, entry 5D2J for the Mg²⁺/adipate/NahK, and entry 5D2K for the Mg²⁺/2-oxoadipate/NahK complexes).

AUTHOR INFORMATION

Corresponding Author

*E-mail: nagem@icb.ufmg.br. Phone: +55 (31) 3409-2626. Fax: +55 (31) 3409-2614.

Funding

This work was supported by FAPEMIG (EDT-0185-0.00-07 and Rede-170/08), CAPES, CNPq (482173/2010-6), VALE S.A (RDP-00174-10), the National Institutes of Health Grant GM-41239 (USA) and the Robert A. Welch Foundation (F-1334).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

In addition to the funding sources, we would like to thank Dr. Masataka Tsuda and Dr. Masahiro Sota for kindly providing us the *P. putida* G7 strain, Dr. William H. Johnson Jr. for providing 2-hydroxyruconate for kinetic assays, and the Laboratório Nacional de Luz Síncrotron (CNPEM, Campinas, Brazil) for provision of synchrotron-radiation facilities.

ABBREVIATIONS

4-OD, 4-oxalocrotonate decarboxylase; VPH, vinylpyruvate hydratase; FAH, fumarylacetoacetate hydrolase; DLS, dynamic light scattering; CD, circular dichroism; MR, molecular replacement; NTD, amino-terminal domain; CTD, carboxy-terminal domain

REFERENCES

- (1) Dunn, N. W., and Gunsalus, I. C. (1973) Transmissible plasmid coding early enzymes of naphthalene oxidation in *Pseudomonas putida*. *J. Bacteriol.* 114, 974–979.
- (2) Tsuda, M., and Iino, T. (1990) Naphthalene degrading genes on plasmid NAH7 are on a defective transposon. *Mol. Gen. Genet.* 223, 33–39.
- (3) Harayama, S., Reik, M., Wasserfallen, A., and Bairoch, A. (1987) Evolutionary relationships between catabolic pathways for aromatics: conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. *Mol. Gen. Genet.* 210, 241–247.
- (4) Harayama, S., Reik, M., Ngai, K. L., and Ornston, L. N. (1989) Physically associated enzymes produce and metabolize 2-hydroxy-2,4-dienoate, a chemically unstable intermediate formed in catechol metabolism via meta cleavage in *Pseudomonas putida*. *J. Bacteriol.* 171, 6251–6258.
- (5) Collinsworth, W. L., Chapman, P. J., and Dagle, S. (1973) Stereospecific enzymes in the degradation of aromatic compounds by *Pseudomonas putida*. *J. Bacteriol.* 113, 922–931.

- (6) Lian, H., and Whitman, C. P. (1994) Stereochemical and isotopic labeling studies of 4-oxalocrotonate decarboxylase and vinylpyruvate hydratase: analysis and mechanistic implications. *J. Am. Chem. Soc.* 116, 10403–10411.

- (7) Kim, S., Kweon, O. K., Kim, Y., Kim, C. K., Lee, K. S., and Kim, Y. C. (1997) Localization and sequence analysis of the phnH gene encoding 2-hydroxy-2,4-dienoate hydratase in *Pseudomonas sp.* strain DJ77. *Biochem. Biophys. Res. Commun.* 238, 56–60.

- (8) Burks, E. A., Johnson, W. H., and Whitman, C. P. (1998) Stereochemical and isotopic labeling studies of 2-oxo-hept-4-ene-1,7-dioate hydratase: evidence for an enzyme-catalyzed ketonization step in the hydration reaction. *J. Am. Chem. Soc.* 120, 7665–7675.

- (9) Montgomery, M. G., Coker, A. R., Taylor, I. A., and Wood, S. P. (2010) Assembly of a 20-nm protein cage by *Escherichia coli* 2-hydroxypentadienoic acid hydratase. *J. Mol. Biol.* 396, 1379–1391.

- (10) Tame, J. R., Namba, K., Dodson, E. J., and Roper, D. I. (2002) The crystal structure of HpcE, a bifunctional decarboxylase/isomerase with a multifunctional fold. *Biochemistry* 41, 2982–2989.

- (11) Izumi, A., Rea, D., Adachi, T., Unzai, S., Park, S. Y., Roper, D. I., and Tame, J. R. (2007) Structure and mechanism of HpcG, a hydratase in the homoprotocatechuate degradation pathway of *Escherichia coli*. *J. Mol. Biol.* 370, 899–911.

- (12) Stanley, T. M., Johnson, W. H., Jr., Burks, E. A., Whitman, C. P., Hwang, C. C., and Cook, P. F. (2000) Expression and stereochemical and isotope effect studies of active 4-oxalocrotonate decarboxylase. *Biochemistry* 39, 3514.

- (13) Johnson, W. H., Jr., Wang, S. C., Stanley, T. M., Czerwinski, R. M., Almrud, J. J., Poelarends, G. J., Murzin, A. G., and Whitman, C. P. (2004) 4-Oxalocrotonate tautomerase, its homologue YwhB, and active vinylpyruvate hydratase: synthesis and evaluation of 2-fluoro substrate analogues. *Biochemistry* 43, 10490–10501.

- (14) Carneiro, F. R., Silva, T. C., Alves, A. C., Haline-Vaz, T., Gozzo, F. C., and Zanchin, N. I. (2006) Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1. *Biochem. Biophys. Res. Commun.* 343, 260–268.

- (15) Gasteiger, E., Wilkins, M. R., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 112, 531–552.

- (16) Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–603.

- (17) Bohm, G., Muhr, R., and Jaenicke, R. (1992) Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng., Des. Sel.* 5, 191–195.

- (18) Jancarik, J., and Kim, S.-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* 24, 409–411.

- (19) Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C., and Craievich, A. (1998) Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. *J. Synchrotron Radiat.* 5, 72–76.

- (20) Guimaraes, B. G., Sanfelici, L., Neuenschwander, R. T., Rodrigues, F., Grizzoli, W. C., Raulik, M. A., Piton, J. R., Meyer, B. C., Nascimento, A. S., and Polikarpov, I. (2009) The MX2 macromolecular crystallography beamline: a wiggler X-ray source at the LNLS. *J. Synchrotron Radiat.* 16, 69–75.

- (21) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* 276, 307–326.

- (22) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.

- (23) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 67, 235–242.
- (24) Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 62, 1002–1011.
- (25) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 213–221.
- (26) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 486–501.
- (27) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
- (28) Holm, L., and Sander, C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* 20, 478–480.
- (29) Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins: Struct., Funct., Genet.* 23, 566–579.
- (30) Timm, D. E., Mueller, H. A., Bhanumoorthy, P., Harp, J. M., and Bunick, G. J. (1999) Crystal structure and mechanism of a carbon-carbon bond hydrolase. *Structure* 7, 1023–1033.
- (31) Manjasetty, B. A., Niesen, F. H., Delbruck, H., Gotz, F., Sievert, V., Bussow, K., Behlke, J., and Heinemann, U. (2004) X-ray structure of fumarylacetoacetate hydrolase family member Homo sapiens FLJ36880. *Biol. Chem.* 385, 935–942.
- (32) Brouns, S. J., Barends, T. R., Worm, P., Akerboom, J., Turnbull, A. P., Salmon, L., and van der Oost, J. (2008) Structural insight into substrate binding and catalysis of a novel 2-keto-3-deoxy-D-arabinonate dehydratase illustrates common mechanistic features of the FAH superfamily. *J. Mol. Biol.* 379, 357–371.
- (33) Ran, T., Gao, Y., Marsh, M., Zhu, W., Wang, M., Mao, X., Xu, L., Xu, D., and Wang, W. (2013) Crystal structures of Cg1458 reveal a catalytic lid domain and a common catalytic mechanism for the FAH family. *Biochem. J.* 449, 51–60.
- (34) Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* 34, W116–118.
- (35) Li, T., Huo, L., Pulley, C., and Liu, A. (2012) Decarboxylation mechanisms in biological system. *Bioorg. Chem.* 43, 2–14.
- (36) O'Leary, M. H. (1992) Catalytic Strategies in Enzymic Carboxylation and Decarboxylation, In *The Enzymes: Mechanisms of Catalysis*, pp 235–269, Elsevier Science, Burlington, MA.
- (37) Grissom, C. B., and Cleland, W. W. (1986) Carbon isotope effects on the metal ion catalyzed decarboxylation of oxalacetate. *J. Am. Chem. Soc.* 108, 5582–5583.
- (38) Casiraghi, G., Zanardi, F., Appendino, G., and Rassa, G. (2000) The vinylogous aldol reaction: a valuable, yet understated carbon-carbon bond-forming maneuver. *Chem. Rev.* 100, 1929–1972.
- (39) Baugh, L., Phan, I., Begley, D. W., Clifton, M. C., Armour, B., Dranow, D. M., Taylor, B. M., Muruthi, M. M., Abendroth, J., Fairman, J. W., Fox, D., 3rd, Dieterich, S. H., Staker, B. L., Gardberg, A. S., Choi, R., Hewitt, S. N., Napuli, A. J., Myers, J., Barrett, L. K., Zhang, Y., Ferrell, M., Mundt, E., Thompkins, K., Tran, N., Lyons-Abbott, S., Abramov, A., Sekar, A., Serbzhinskiy, D., Lorimer, D., Buchko, G. W., Stacy, R., Stewart, L. J., Edwards, T. E., Van Voorhis, W. C., and Myler, P. J. (2015) Increasing the structural coverage of tuberculosis drug targets. *Tuberculosis (Oxford, U. K.)* 95, 142–148.
- (40) Kube, M., Chernikova, T. N., Al-Ramahi, Y., Beloqui, A., Lopez-Cortez, N., Guazzaroni, M. E., Heipieper, H. J., Klages, S., Kotsyurbenko, O. R., Langer, I., Nechitaylo, T. Y., Lunsdorf, H., Fernandez, M., Juarez, S., Ciordia, S., Singer, A., Kagan, O., Egorova, O., Petit, P. A., Stogios, P., Kim, Y., Tchigvintsev, A., Flick, R., Denaro, R., Genovese, M., Albar, J. P., Reva, O. N., Martinez-Gomariz, M., Tran, H., Ferrer, M., Savchenko, A., Yakunin, A. F., Yakimov, M. M., Golyshina, O. V., Reinhardt, R., and Golyshin, P. N. (2013) Genome sequence and functional genomic analysis of the oil-degrading bacterium *Oleispira antarctica*. *Nat. Commun.* 4, 2156.
- (41) Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–324.
- (42) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.

Crystal structures of apo and liganded 4-oxalocrotonate decarboxylase uncover a structural basis for the metal-assisted decarboxylation of a vinylogous β -keto acid

Samuel L. Guimarães[‡], Juliana B. Coitinho[‡], Débora M. A. Costa[‡], Simara S. Araújo[‡], Christian P. Whitman[§], and Ronaldo A. P. Nagem^{‡}*

[‡] Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 31270-901, Brazil, and [§] Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas, Austin, Texas 78712-1071, United States

* To whom correspondence should be addressed.

KEYWORDS

biodegradation, crystal structure, decarboxylase, enzyme mechanism, naphthalene, *Pseudomonas putida*

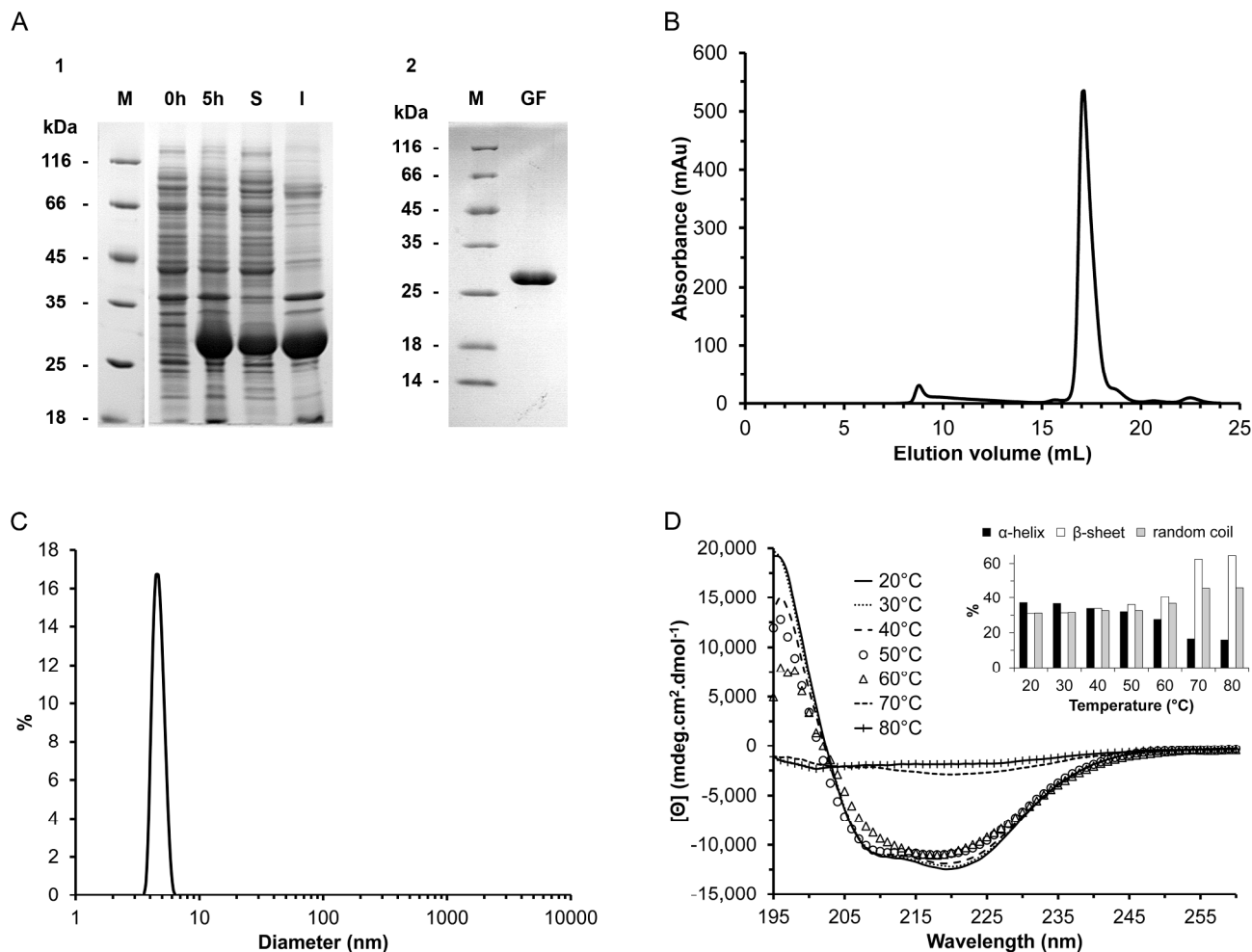


Figure S1. (A1) SDS-PAGE (12%) analysis of NahK expression. Lane M: Protein molecular weight marker (Fermentas); Lane 0h: crude bacterial extract before induction with IPTG; Lane 5h: crude bacterial extract past induction (5 h); Lane S and I: soluble and insoluble fractions, respectively, after cell lysis and centrifugation. (A2) SDS-PAGE (15%) analysis of final purification step. Lane M: Protein molecular weight marker (Fermentas); Lane GF: the recombinant NahK after two affinity chromatography steps and gel filtration. (B) Size-exclusion elution profile of NahK using Superdex 200 10/300 GL column (GE Healthcare). Protein elution was monitored by measuring absorbance at 280 nm. (C) Dynamic light scattering measurements of purified NahK. The peak corresponds to a particle with a hydrodynamic diameter of 4.7 nm, equivalent to a globular protein with molecular mass of 26 kDa. The polydispersity is 9.7%. The protein concentration was approximately 1 mg/mL. (D) Secondary structure analysis and thermal stability of NahK by CD spectroscopy. Figure inset: the percentage of each type of secondary structure calculated by the

CDNN program. The spectra show the thermal denaturation of NahK at 0.2 mg/mL from 20 to 80 °C in 5 mM Tris-HCl buffer, pH 7.4, and 5 mM NaCl. As temperature is increased, a gradual loss of the α -helical structures is observed, which is accompanied by an increase of parallel and antiparallel β -sheet, and random coil content. These changes are consistent with a major transition of unfolding with a temperature of melting between 60 and 70 °C.

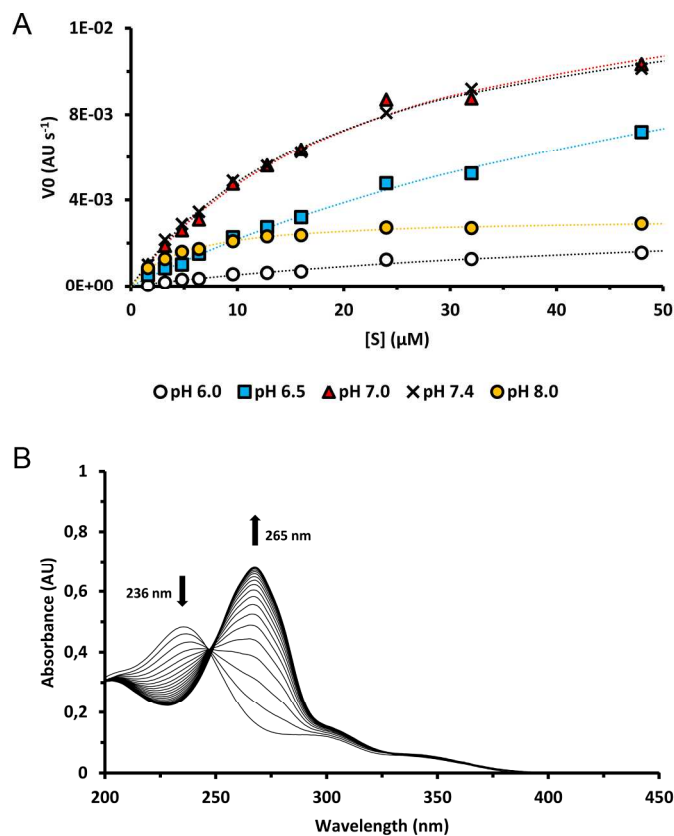


Figure S2. (A) Michaelis-Menten kinetics of NahK as a function of pH (at 25 °C). The points are measured rates and the line is the least-squared minimized fit of the data to the Michaelis-Menten equation. (B) Ultraviolet-visible spectra for the decarboxylation of 2-oxo-3-hexenedioate by NahK in the absence of NahL.

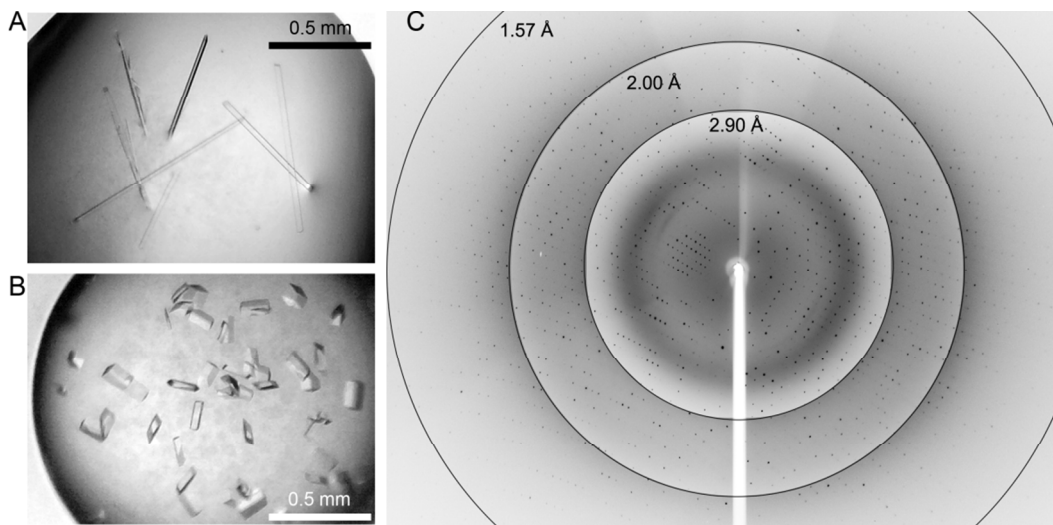


Figure S3. (A) Needle-shaped crystals of recombinant NahK grown under CS1-22 optimized conditions. (B) Rectangular-shaped crystals of recombinant NahK grown under the CS1-16 condition. (C) Typical X-ray diffraction pattern for the Mg^{2+} /2-oxoadipate/NahK complex crystal. The crystal-to-detector distance was set to give a resolution limit of 1.57 Å at the outer edge of the image. The diffraction pattern background around the delimited circles of resolution ranges was altered to allow better visualization of the reflections.

Video S1. Simulation of hypothesized conformational changes that occur in NahK upon substrate binding, highlighting the rearrangement of catalytic residues that trap and orientate the substrate in the active site prior decarboxylation. A cartoon representation of the protein molecule was used in this simulation. The models used in this simulation was: *Mg* model (open), *AA* model (partially closed), and *2OA* model (closed). The video was prepared using the Morph Conformations tool into the Chimera software.

Video S2. The same simulation of Video S1 where a surface representation of the protein molecule was used. Red surface highlights the active site cavity. Lys72 and Ser164 surfaces are shown in blue. Met76, Phe151 and Phe153 surfaces are shown in green.