

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

SIMARA SEMÍRAMIS DE ARAÚJO

**Estrutura e atividade da proteína 2-hidroxiuconato semialdeído desidrogenase (NahI)
de *Pseudomonas putida* G7, uma enzima da via de degradação do naftaleno**

Belo Horizonte

2015

UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

DEPARTAMENTO DE BIOQUÍMICA E IMUNOLOGIA

SIMARA SEMÍRAMIS DE ARAÚJO

**Estrutura e atividade da proteína 2-hidroximuconato semialdeído desidrogenase (NahI)
de *Pseudomonas putida* G7, uma enzima da via de degradação do naftaleno**

Tese apresentada ao Programa de Pós-Graduação em Bioquímica e Imunologia como requisito parcial para a obtenção do grau de Doutora em Ciências - Bioquímica.

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

Orientador: Prof. Dr. Ronaldo Alves Pinto Nagem.

Belo Horizonte

2015



ATA DA DEFESA DA TESE DE DOUTORADO DE SIMARA SEMÍRAMIS DE ARAÚJO .
 Aos trinta dias do mês de junho de 2015 às 08:30 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 ""Estrutura e atividade da proteína 2-hidroxiuconato semialdeído desidrogenase (Nahl) de *Pseudomonas Putida* G7 uma enzima da via de degradação do naftaleno"", 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 à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações: Dr. Lucas Bleicher (Universidade Federal de Minas Gerais), aprovada; Dr. Tiago Antonio da Silva Brandão (Universidade Federal de Minas Gerais), aprovada; Dr. Marcos Vicente de Albuquerque Salles Navarro (Universidade de São Paulo), aprovada; Dr. Mario de Oliveira Neto (Universidade Estadual Paulista), aprovada; Dr. Ronaldo Alves Pinto Nagem - Orientador (Universidade Federal de Minas Gerais), aprovada. Pelas indicações a candidata foi considerada APROVADA. O resultado final foi comunicado publicamente à candidata 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, 30 de junho de 2015.

Lucas Bleicher

Dr. Lucas Bleicher (UFMG)

Tiago Brandão
 Dr. Tiago Antonio da Silva Brandão (UFMG)

Albuquerque Salles Navarro
 Dr. Marcos Vicente de Albuquerque Salles Navarro (Universidade de São Paulo)

Mario de Oliveira Neto
 Dr. Mario de Oliveira Neto (Universidade Estadual Paulista)

Ronaldo Alves Pinto Nagem
 Dr. Ronaldo Alves Pinto Nagem - Orientador (UFMG)

Jader dos Santos Cruz
 Prof Jader dos Santos Cruz
 Coordenador de Curso de Pós Graduação
 em Bioquímica e Imunologia
 ICB - UFMG

*Aos meus pais, Tânia e Veríssimo, com amor,
admiração e gratidão por sua presença,
carinho, compreensão e incansável apoio ao
longo do período de elaboração deste
trabalho.*

AGRADECIMENTOS

Ao prof. Dr. Ronaldo Alves Pinto Nagem, pelo entusiasmo em ensinar, pela acessibilidade sempre presente entre aluno e orientador, pelo empenho em oferecer a estrutura adequada, segura e necessária para a realização dos experimentos, e pela vontade diária em querer aprender.

Às agências de fomento e instituições que permitiram a elaboração deste trabalho: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e VALE S.A.

Ao prof. Dr. Ricardo Aparício (Universidade Estadual de Campinas), aos Drs. Christian P. Whitman e William H. Johnson Jr. (The University of Texas), e aos colegas Cíntia Mara Leal Neves e Samuel Leite Guimarães pela contribuição efetiva neste trabalho.

Ao Laboratório Nacional de Luz Síncrotron pelo acesso às linhas de luz SAXS1, MX1 e MX2, bem como ao Laboratório Automatizado de Cristalização de Macromoléculas (CNPEM, Campinas - SP, Brasil).

Aos membros da banca examinadora, titulares e suplentes.

Ao Grupo de Biologia Estrutural do ICB / UFMG.

Aos professores e funcionários do Departamento de Bioquímica e Imunologia.

À minha família e amigos, pelo apoio.

“Se te fatigas correndo com homens que vão a pé, como poderás competir com os cavalos?”

Jeremias 12:5

RESUMO

ARAÚJO, S. S. **Estrutura e atividade da proteína 2-hidroxiuconato semialdeído desidrogenase (NahI) de *Pseudomonas putida* G7, uma enzima da via de degradação do naftaleno.** 2015. Tese (Doutorado) - Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais. Belo Horizonte, 2015.

Os Hidrocarbonetos Aromáticos Policíclicos (HAPs), um grupo de compostos orgânicos formados por anéis benzênicos fusionados, são contaminantes ambientais amplamente distribuídos tanto devido a fontes naturais quanto antropogênicas, sendo as atividades de produção de carvão vegetal e mineral, de refino e transporte de petróleo os principais contribuintes para o despejo localizado de HAPs. Devido à sua ocorrência ubíqua, potencial de bioacumulação e atividade carcinogênica, vários HAPs foram listados como poluentes prioritários para a remediação destacando a importância da sua remoção do meio ambiente. Diversos microrganismos têm sido extensamente estudados em virtude de sua versatilidade em degradar uma gama de compostos aromáticos. A bactéria *Pseudomonas putida* G7 possui um plasmídeo associado ao metabolismo do naftaleno, o mais simples e abundante HAP, cuja degradação ocorre através de duas vias catabólicas ditas superior e inferior. NahI, uma 2-hidroxiuconato semialdeído desidrogenase (2-HMSD) atuante na via inferior, converte 2-hidroxiuconato semialdeído (2-HMS) a 2-hidroxiuconato (2-HM) na presença de NAD^+ . Mesmo a despeito da baixa identidade de sequência, uma característica comum entre enzimas aldeído desidrogenases (ALDHs) é a manutenção de um enovelamento estrutural semelhante, salvas as inúmeras particularidades concernentes ao estado de oligomerização e às interações específicas estabelecidas entre enzima, substrato e cofator. Assim, este trabalho se propôs a caracterizar cinética e estruturalmente a enzima NahI, uma vez que, através da elucidação de sua estrutura tridimensional e comparação com as de outras ALDHs, possibilita-se inferir distinções no sítio catalítico responsáveis por mudanças sistemáticas nas propriedades cinéticas destas enzimas sobre uma variedade de substratos aldeídos. Após purificação por cromatografia de afinidade e de exclusão molecular, ensaios de espalhamento dinâmico de luz (DLS) e espalhamento de raios-X a baixo ângulo (SAXS) foram realizados para analisar o estado oligomérico da enzima em solução, identificada como um tetrâmero. No que diz respeito à estrutura cristalográfica, NahI apresentou o enovelamento α/β típico da superfamília ALDH organizado em domínio de oligomerização e domínios de ligação ao dinucleotídeo e ao substrato, este último também conhecido como catalítico. Através de operações de simetria cristalográfica, observou-se o tetrâmero de NahI formado por um dímero de dímeros. Ensaios de cinética confirmaram a preferência da enzima para o substrato alifático 2-HMS, porém não se observou qualquer atividade sobre salicilaldeído, substrato aromático natural de NahF, uma ALDH pertencente à mesma via de degradação do naftaleno. Mutações foram propostas para NahI a fim de aumentar o volume do bolso catalítico, uma vez que resíduos de cadeia lateral volumosa poderiam dificultar a atividade desta enzima sobre substratos aromáticos. Surpreendentemente, as formas mutantes não apresentaram atividade sobre salicilaldeído. Além disso, o alargamento da entrada do bolso catalítico afetou ligeiramente a afinidade da enzima para o seu substrato natural, 2-HMS. Sugere-se ainda que o resíduo L156 parece ser crucial para a oxidação do substrato. Os resultados desta investigação apresentam, pela primeira vez, características estruturais e cinéticas em relação à enzima NahI e favorecem futuros estudos sobre a identidade e função de resíduos cruciais para a catálise e a descrição de como a reação se processa.

Palavras-chave: *Pseudomonas putida* G7, degradação de naftaleno, 2-hidroxiuconato semialdeído desidrogenase, estrutura cristalográfica, atividade cinética.

ABSTRACT

ARAÚJO, S. S. **Protein structure and activity of 2-hydroxymuconate semialdehyde dehydrogenase (NahI) from *Pseudomonas putida* G7, a naphthalene-degradation pathway enzyme.** 2015. Tese (Doutorado) - Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais. Belo Horizonte, 2015.

The Polycyclic Aromatic Hydrocarbons (PAHs), a group of organic compounds consisting of fused benzene rings, are contaminants widely distributed in the environment due to natural as well as anthropogenic sources. Activities of production of coal, refining and transportation of oil are the major contributors to the contamination with PAHs. Due to their ubiquitous occurrence, potential to bio-accumulate and carcinogenic activity, a number of PAHs were listed as priority pollutants for remediation highlighting the importance of their removal from the environment. Several microorganisms have been extensively studied because of its versatility to degrade a wide range of aromatic compounds. The bacterium *Pseudomonas putida* G7 has a plasmid associated with the metabolism of naphthalene, the simplest and most abundant PAH. Its degradation by *P. putida* occurs through the upper and the lower catabolic pathways. NahI, a 2-hydroxymuconate semialdehyde dehydrogenase (2-HMSD) belonging to the lower pathway, converts 2-hydroxymuconate semialdehyde (2-HMS) to 2-hydroxymuconate (2-HM) in the presence of NAD^+ . A common feature among aldehyde dehydrogenases (ALDHs) is the similar scaffold, even despite their overall low sequence identity, modes of oligomerization and substrate specificity. Thus, this study was proposed to characterize the structure and kinetics of NahI. By elucidation of its three-dimensional structure and comparison with other ALDHs, it becomes possible to infer distinctions in the catalytic site that are responsible for systematic changes in the kinetic properties of these enzymes within a variety of aldehyde substrates. After purification by affinity and size-exclusion chromatography, dynamic light scattering (DLS) and small-angle X-ray scattering (SAXS) experiments were conducted to analyze the oligomeric state of the enzyme in solution, which was identified as being a tetramer. With regard to the crystal structure, NahI displayed a typical α/β aldehyde dehydrogenase superfamily fold with three well defined domains: the oligomerization domain, the nucleotide binding and the catalytic domains. Through crystallographic symmetry operations, the NahI tetramer formed by a dimer of dimers was observed. Kinetic assays confirmed the preference of the enzyme for the aliphatic substrate 2-HMS. On the other hand, NahI showed no activity with salicylaldehyde, the natural substrate for NahF enzyme, another ALDH belonging to the same naphthalene-degradation pathway. Mutations have been proposed for NahI in order to increase the volume of the catalytic pocket, since bulky side chain residues could hinder the activity of this enzyme with aromatic substrates. Surprisingly, the mutant forms showed no activity with salicylaldehyde. Moreover, the enlargement of the catalytic pocket entrance slightly affected the affinity of the enzyme for its natural substrate, 2-HMS. It also suggests that the L156 residue appears to be crucial for the oxidation of the substrate. For the first time, these findings show the structural and kinetic properties relative to the NahI enzyme and highlights further studies concerning the identity and function of critical residues for catalysis and description of how the reaction proceeds.

Keywords: *Pseudomonas putida* G7, naphthalene degradation, 2-hydroxymuconate semialdehyde dehydrogenase, crystal structure, kinetics.

Lista de Figuras

Figura 1: Hidrocarbonetos aromáticos policíclicos.....	2
Figura 2: Mecanismos de formação de adutos entre HAPs e DNA.	3
Figura 3: Representação esquemática dos operons <i>nah</i> e <i>sal</i> de NAH7.	7
Figura 4: Oxidação inicial do naftaleno a cis-1,2-dihidroxi-1,2-dihidronaftaleno pelo complexo da naftaleno dioxigenase (NahAaAbAcAd).	8
Figura 5: Vias superior e inferior de degradação do naftaleno.	9
Figura 6: Organização dos genes catabólicos <i>xyl</i> , <i>nah</i> e <i>dmp</i> nos operons de via inferior / meta-clivagem codificados por pWW0, NAH7 e pVI150 para a degradação de tolueno/xileno, naftaleno e dimetilfenol, respectivamente.	11
Figura 7: Estrutura tridimensional de uma ALDH classe 3 resolvida a 2,6 Å de resolução...	14
Figura 8: Mecanismo de ação proposto para ALDHs.	15
Figura 9: Comparação das estruturas primária e terciária de ALDH2 (1AG8) e ALDH3 (1AD3).	16
Figura 10: Estrutura quaternária de ALDH2 (1AG8) e ALDH3 (1AD3).	16
Figura 11: A ALDH hexamérica P5CDH de <i>Thermus thermophilus</i>	17
Figura 12: A enzima NahF, uma salicilaldeído desidrogenase da via de degradação do naftaleno de <i>Pseudomonas putida</i> G7.	18
Figura 13: 2-hidroxiuconato semialdeído (1) e 2-hidroxiuconato (2).	18
Figura 14: Aldeído desidrogenases, XylC e XylG, envolvidas na degradação de tolueno e xilenos por <i>Pseudomonas putida</i> mt-2.	20
Figura 15: Desenho esquemático do vetor de expressão pET28a-TEV (5357 pb).	24
Figura 16: Mapas das condições de cristalização testadas para os mutantes de NahI.	31
Figura 17: Expressão heteróloga e purificação de 6xHis-NahI.	36
Figura 18: Análises de estrutura secundária e termoestabilidade de NahI na presença e ausência de NAD ⁺	38
Figura 19: Estado oligomérico de NahI avaliado por cromatografia de exclusão molecular.	40

Figura 20: Dados de SAXS e parâmetros globais.	44
Figura 21: Cristais de 6xHis-NahI, uma 2-hidroxiuconato semialdeído desidrogenase de <i>Pseudomonas putida</i> G7, crescidos em variadas condições de cristalização.	46
Figura 22: Estrutura cristalográfica da enzima NahI recombinante, uma 2-hidroxiuconato semialdeído desidrogenase de <i>Pseudomonas putida</i> G7, resolvida por difração de raios-X a 1,85 e 2,15 Å de resolução.	53
Figura 23: Modelo cristalográfico do tetrâmero da enzima NahI resolvido por difração de raios-X a 1,85 Å de resolução.	54
Figura 24: Localização do cofator NAD ⁺ na estrutura terciária de NahI resolvida por difração de raios-X a 2,15 Å de resolução.	55
Figura 25: Interações entre o NAD e resíduos do sítio de ligação ao cofator em NahI.	56
Figura 26: Ligações de hidrogênio e interações hidrofóbicas observadas no complexo NahI-NAD ⁺	58
Figura 27: Sítio de ligação ao substrato.	59
Figura 28: Atividade cinética de NahI.	60
Figura 29: Substratos de XylG.	61
Figura 30: Comparação estrutural de possíveis resíduos interagentes com o substrato em NahI e Sp2771.	62
Figura 31: Alinhamento de sequencias de aminoácidos de NahI e três outras 2-HMSDs.	63
Figura 32: Localização da porção compreendida por Glu88 e Ala108 na estrutura cristalográfica de NahI.	64
Figura 33: Alinhamento das sequencias nucleotídicas de NahI (NAH7) e XylG (pWW0).	65
Figura 34: Estruturas diméricas das enzimas NahI e NahF de <i>P. putida</i> G7.	66
Figura 35: Alinhamento entre as sequencias de NahI e NahF.	70
Figura 36: Análise comparativa dos sítios de ligação ao substrato de NahI e NahF.	68
Figura 37: Expansão dos sítios de ligação ao substrato em formas mutantes de NahI.	69
Figura 38: Cromatogramas obtidos durante as etapas de purificação das formas mutantes de NahI.	71

Figura 39: Perfil eletroforético das formas mutantes de NahI após a etapa cromatográfica de exclusão molecular.	71
Figura 40: Cristais de formas mutantes de NahI, crescidos em condições variadas de formato de sódio.....	72
Figura 41: Célula unitária do cristal de NahI-L156G-F448S.....	74
Figura 42: Entrada do sítio catalítico de NahI e de dois mutantes.....	75

Lista de Quadros e Tabelas

Quadro 1: Construções Plasmidiais	23
Quadro 2: Antibióticos.....	24
Tabela 1: As 13 famílias de ALDHs..	13
Tabela 2: Espalhamento dinâmico de luz de NahI na ausência e presença de NAD^+	41
Tabela 3: Estatísticas dos dados de difração dos cristais de NahI e NahI- NAD^+	47
Tabela 4: Número de moléculas de proteína presentes na unidade assimétrica, coeficiente de Matthews e conteúdo de solvente calculados para cristais de NahI e NahI- NAD^+	48
Tabela 5: Estatísticas de refinamento e de validação dos modelos cristalográficos de NahI e NahI- NAD^+	50
Tabela 6: Concentrações (em mg/mL) das formas mutantes de NahI utilizadas nos ensaios de cristalização.	70
Tabela 7: Estatísticas dos dados de difração e de processamento dos cristais das formas mutantes de NahI. Os valores estatísticos para as faixas de menor e maior resolução são dadas entre parênteses.....	73

Lista de Abreviaturas e Siglas

2-AMS	2-aminomuconato 6-semialdeído
2-HM	2-hidroximuconato
2-HMS	2-hidroximuconato semialdeído
2-HMSD	2- hidoximuconato semialdeído desidrogenase
ALDH	aldeído desidrogenase
a.u.	unidades arbitrárias
CD	dicroísmo circular, do inglês <i>circular dichroism</i>
D_h	diâmetro hidrodinâmico
DLS	espalhamento dinâmico de luz, do inglês <i>dynamic light scattering</i>
DO	densidade óptica
DSF	fluorimetria de varredura diferencial, do inglês <i>differential scanning fluorimetry</i>
DTT	ditiotreitól
HAP	hidrocarbonetos aromáticos policíclicos
HMSALDH	família hidroximuconato semialdeído desidrogenase
IPTG	isopropil- β -D-tiogalactopiranosídeo
k_{cat}	constante catalítica
kDa	kiloDalton
K_M	constante de <i>Michaelis-Menten</i>
kV	kilovolts
M	concentração molar (moles/L)
mM	concentração milimolar (10^{-3} moles/L)
μ M	concentração micromolar (10^{-6} moles/L)
MM	massa molecular
nm	nanômetro (10^{-9} m)
NAD	dinucleotídeo de nicotinamida-adenina
NADP	dinucleotídeo de nicotinamida-adenina fosfato
NahF	salicilaldeído desidrogenase de <i>Pseudomonas putida</i> G7
NahI	2- hidoximuconato semialdeído desidrogenase de <i>Pseudomonas putida</i> G7
pb	pares de bases
PCR	reação em cadeia da polimerase, do inglês <i>polymerase chain reaction</i>
R_g	raio de giro
R_h	raio hidrodinâmico
R.m.s.D.	desvio quadrático médio, do inglês <i>root-mean-square deviation</i>

rpm	rotações por minuto
SAXS	espalhamento de raios-X a baixo ângulo, do inglês <i>small-angle x-ray scattering</i>
SDS	dodecil sulfato de sódio
Tm	temperatura de transição
XylG	2- hidoximuconato semialdeído desidrogenase de <i>Pseudomonas putida</i> mt-2

Abreviações (em uma e três letras) 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

Sumário

1. INTRODUÇÃO E JUSTIFICATIVA	1
1.1. HIDROCARBONETOS AROMÁTICOS POLICÍCLICOS	1
1.2. BIOREMEDIAÇÃO DE HIDROCARBONETOS AROMÁTICOS POLICÍCLICOS	4
1.3. BIODEGRADAÇÃO DE NAFTALENO POR <i>Pseudomonas putida</i> G7	6
1.4. A VIA DE META-CLIVAGEM DO CATECOL	10
1.5. ALDEÍDO DESIDROGENASES	11
2. OBJETIVOS	22
2.1. GERAL	22
2.2. ESPECÍFICOS	22
2.2.1. <i>NahI</i>	22
2.2.2. <i>Mutantes de NahI</i>	22
3. MATERIAIS E MÉTODOS	23
3.1. PLASMÍDEOS	23
3.2. ANTIBIÓTICOS	24
3.3. CEPAS BACTERIANAS	24
3.4. PREPARO DE CÉLULAS ELETROCOMPETENTES	25
3.5. TRANSFORMAÇÃO DE BACTÉRIAS POR ELETROPORAÇÃO	26
3.6. EXPRESSÃO HETERÓLOGA EM <i>E. coli</i>	26
3.7. LISE CELULAR	27
3.8. PURIFICAÇÃO DAS PROTEÍNAS RECOMBINANTES	27
3.9. DETERMINAÇÃO DA CONCENTRAÇÃO PROTEICA	28
3.10. ESPALHAMENTO DINÂMICO DE LUZ	29
3.11. DICROÍSMO CIRCULAR	30
3.12. CRISTALIZAÇÃO DE PROTEÍNAS E DIFRAÇÃO DE RAIOS-X	30
3.13. RESOLUÇÃO DA ESTRUTURA CRISTALOGRÁFICA	32
3.14. ESPALHAMENTO DE RAIOS-X A BAIXO ÂNGULO	33
3.15. ENSAIO DE ATIVIDADE ENZIMÁTICA	34
4. RESULTADOS E DISCUSSÃO	36
4.1. ESTRUTURA SECUNDÁRIA DE <i>NahI</i> NA AUSÊNCIA E NA PRESENÇA DE NAD^+	37
4.2. ESTADO OLIGOMÉRICO DE <i>NahI</i> EM SOLUÇÃO	39

4.2.1. <i>Por cromatografia de exclusão molecular e espalhamento dinâmico de luz</i>	39
4.2.2. <i>Por espalhamento de raios-X a baixo ângulo</i>	43
4.3. RESOLUÇÃO DA ESTRUTURA CRISTALOGRÁFICA DE NahI	46
4.4. MODELO CRISTALOGRÁFICO PROPOSTO PARA NahI	52
4.5. SÍTIO DE LIGAÇÃO AO DINUCLEOTÍDEO E SUAS INTERAÇÕES COM O NAD ⁺	55
4.6. SÍTIO DE LIGAÇÃO AO SUBSTRATO E ATIVIDADE CINÉTICA DE NahI	59
4.7. ANÁLISE COMPARATIVA DOS SÍTIOS DE LIGAÇÃO AO SUBSTRATO DE NahI E NahF	66
4.8. PURIFICAÇÃO DAS VARIANTES DE NahI	70
4.9. CRISTALIZAÇÃO, DIFRAÇÃO DE RAIOS-X E ESTRUTURA DAS VARIANTES DE NahI	72
4.10. ATIVIDADE CINÉTICA DAS VARIANTES DE NahI	76
5. CONCLUSÕES E PERSPECTIVAS	79
6. REFERÊNCIAS BIBLIOGRÁFICAS	82
7. ANEXO	90

1. INTRODUÇÃO E JUSTIFICATIVA

1.1. HIDROCARBONETOS AROMÁTICOS POLICÍCLICOS

Os Hidrocarbonetos Aromáticos Policíclicos (HAPs) constituem um grupo de compostos orgânicos com dois ou mais anéis benzênicos fusionados, classificados como “leves” (2 a 3 anéis) ou “pesados” (4 ou mais anéis) (Purcaro et al 2013).

Devido à possibilidade de fusão de um número variável de anéis e de diferentes posições em que estes anéis podem se ligar entre si, há atualmente mais de 100 HAPs reconhecidos pela União Internacional de Química Pura e Aplicada (*IUPAC*) (Jacques, Bento et al. 2007), dentre os quais, 16 compostos são considerados poluentes prioritários pela Agência de Proteção Ambiental dos Estados Unidos (*US EPA*). São eles: acenafteno, acenaftileno, antraceno, benzo(*a*)antraceno, benzo(*a*)pireno, benzo(*b*)fluoranteno, benzo(*k*)fluoranteno, benzo(*g,h,i*)perileno, criseno, dibenzo(*a,h*)antraceno, fenantreno, fluoranteno, fluoreno, indeno(1,2,3-*c,d*)pireno, naftaleno e pireno (Figura 1) (Habe & Omori 2003; Purcaro et al 2013).

Os HAPs ocorrem naturalmente em combustíveis fósseis, como carvão e petróleo, e também são formados durante a combustão incompleta de matéria orgânica. Fontes naturais incluem incêndios de florestas e pastagens, além de erupções vulcânicas. Entretanto, a principal causa de contaminação ambiental por HAPs é a geração antropogênica de tais compostos, associada principalmente às atividades de produção de carvão vegetal, de extração e gaseificação do carvão mineral e à cadeia de extração, transporte, refino, transformação e utilização do petróleo e seus derivados (Bamforth & Singleton 2005; Das & Chandran 2011; Haritash & Kaushik 2009; Kanaly & Harayama 2000).

Em virtude da estabilidade da estrutura química dos HAPs, dada pela ressonância dos anéis aromáticos, e da baixa solubilidade em água, tais compostos apresentam forte tendência de adsorção às partículas orgânicas e minerais do solo e dos sedimentos (Jacques et al 2007), de forma que a permanência dos HAPs no ambiente dependerá, portanto, de vários fatores, como estrutura química, concentração, dispersão e biodisponibilidade do contaminante. Somado a isto, o tipo e a estrutura do solo, pH e temperatura, bem como níveis adequados de oxigênio, nutrientes e água influenciam na capacidade da biota microbiana local em degradar

essas moléculas. De modo geral, quanto mais elevado é o peso molecular do HAP, maiores são a hidrofobicidade, a toxicidade e a persistência no ambiente (Bamforth & Singleton 2005; Kanaly & Harayama 2000).

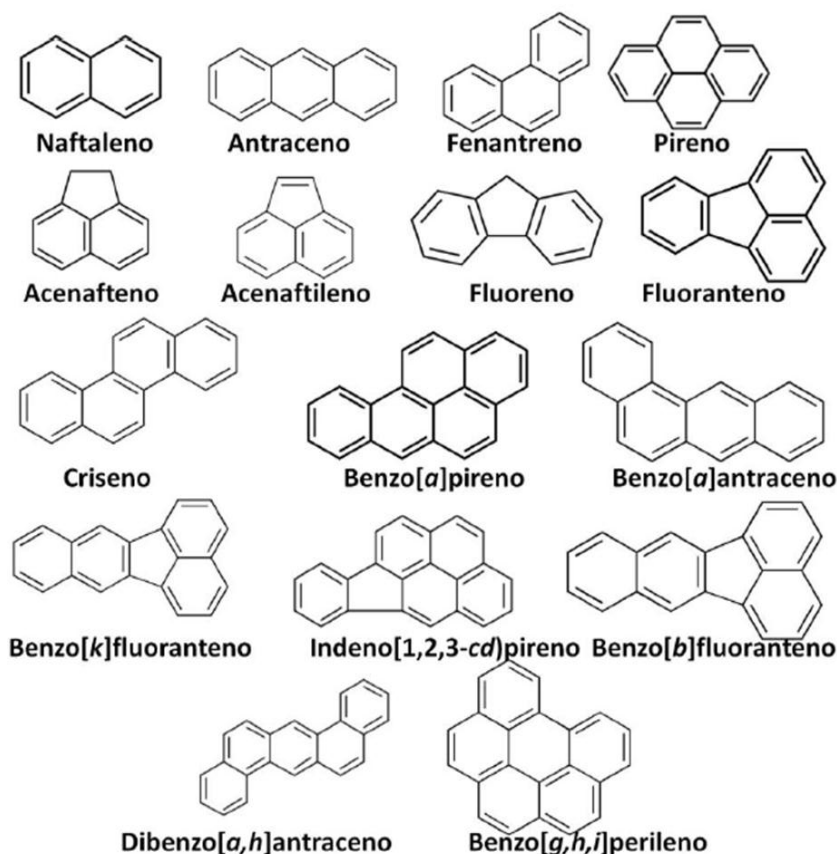


Figura 1: Hidrocarbonetos aromáticos policíclicos. Estruturas químicas dos 16 HAPs listados como poluentes prioritários pela Agência de Proteção Ambiental dos Estados Unidos (US EPA). Adaptado de (Habe & Omori 2003).

Quando hidrocarbonetos aromáticos são incorporados pela flora e fauna, devido ao seu caráter lipofílico tornam-se parte da reserva lipídica, podendo ser transferidos ao longo da cadeia alimentar sem alteração de suas estruturas (Tiburtius et al 2004). Ainda assim, mesmo sendo prontamente absorvidos no trato gastrointestinal quando ingeridos pelo homem, a principal rota de entrada dos HAPs no organismo é via absorção dérmica. Por isso, em geral, o número de anéis benzênicos se relaciona diretamente à toxicidade do hidrocarboneto aromático (Bamforth & Singleton 2005; Cerniglia 1984).

O metabolismo dos HAPs em mamíferos visa à conversão de tais compostos lipofílicos em metabólitos hidrossolúveis, de forma a serem mais facilmente excretados pelo organismo, o que, todavia, elicit as propriedades mutagênicas, genotóxicas e carcinogênicas destas moléculas (Cerniglia 1984).

Alguns mecanismos têm sido propostos para explicar a ativação metabólica dos hidrocarbonetos aromáticos (Baird et al 2005; Rubin 2001; Saeed et al 2009; Stansbury et al 1994). No mecanismo por formação de diol-epóxido, monoxigenases dependentes de citocromo P₄₅₀ são responsáveis pela oxidação enzimática dos HAPs, formando compostos epóxidos que podem espontaneamente produzir fenóis ou, por ação das epóxido hidrolases, di-hidrodióis. Estes podem sofrer nova epoxidação levando à formação de di-hidrodiolépóxidos, capazes de reagir com as bases nucleotídicas do DNA e, eventualmente, iniciar um processo mutagênico (Figura 2 A). No mecanismo via formação de quinona, a desidrogenação enzimática dos di-hidrodióis produz quinonas capazes de reagirem diretamente com o DNA ou com oxigênio molecular, assim gerando espécies reativas de oxigênio, como radicais hidroxila ou ânions superóxido, que reagem com o DNA (Figura 2 B). Estes mecanismos não são excludentes e podem ocorrer simultaneamente (Netto et al 2000).

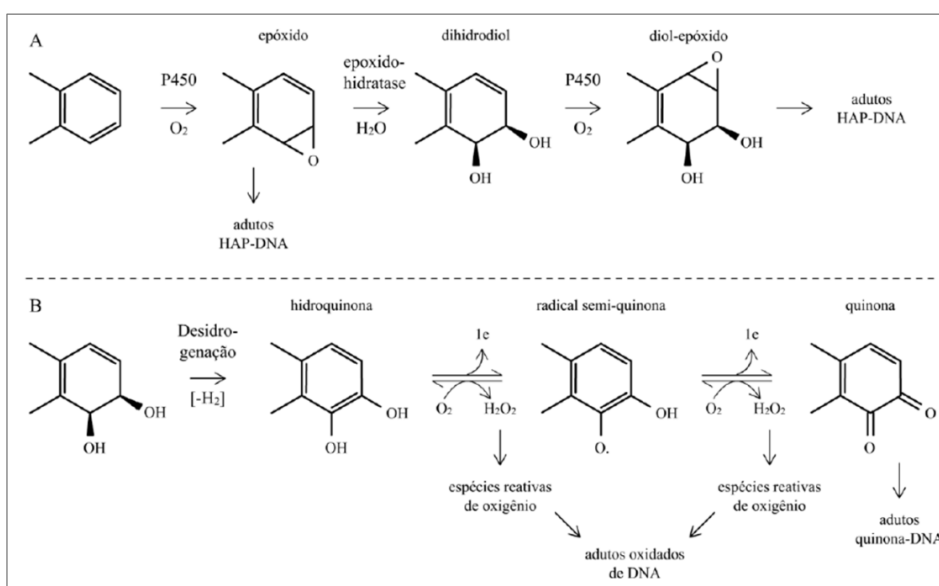


Figura 2: Mecanismos de formação de adutos entre HAPs e DNA. A ativação metabólica dos HAPs envolve uma série de enzimas. A. Oxidação enzimática seguida de hidrólise com a formação de diolépóxidos. B. Desidrogenação enzimática dos metabólitos di-hidrodióis produz quinonas capazes de reagirem diretamente com o DNA ou com O₂, gerando espécies reativas de oxigênio que reagem com o DNA. Adaptado de (Netto et al 2000).

1.2. BIOREMEDIAÇÃO DE HIDROCARBONETOS AROMÁTICOS POLICÍCLICOS

A bioremediação envolve a utilização de microrganismos, seja de ocorrência natural (nativos) ou cultivada, para degradar ou imobilizar contaminantes orgânicos em águas subterrâneas e em solos. O objetivo principal é a mineralização completa dos contaminantes, ou seja, a sua biotransformação em produtos com pouca ou nenhuma toxicidade, como CO₂ e água. Para tanto, os microrganismos geralmente utilizados são bactérias, fungos filamentosos e leveduras (Bamforth & Singleton 2005).

O tratamento biológico de ambientes contaminados com HAPs pode ser uma alternativa eficiente, econômica e versátil ao tratamento físico-químico, visto que oferece ótima relação custo-benefício e maior eficiência na degradação de compostos tóxicos e recalcitrantes, além de maior segurança e menor perturbação ambiental (Habe & Omori 2003).

Os tratamentos são basicamente de dois tipos: 1) *ex-situ* (ou *off-site*), realizado fora do local onde ocorreu a contaminação, é portanto um tratamento que requer a escavação e a remoção do solo contaminado para outro local. A adoção deste procedimento pode resultar em um aumento considerável do custo do processo, porém, não obstante a essa desvantagem, é possível controlar com maior facilidade os fatores que influenciam a biodegradabilidade dos contaminantes, tais como o tipo e a quantidade de microrganismos, bem como as condições físicas e químicas do sítio contaminado (pH, umidade, temperatura, salinidade, teor de oxigênio e quantidade dos nutrientes nitrogênio, fósforo e potássio); 2) *in-situ* (ou *on-site*), tratamento feito no próprio local da contaminação. Normalmente, essa opção de bioremediação, a mais empregada no mundo, torna o processo mais atrativo e economicamente viável quando comparado ao tratamento citado anteriormente. Além disso, o tratamento *in-situ* acarreta em menores impactos ambientais advindos da remediação da área contaminada (Andrade et al 2010).

Como exemplo, a bioremediação *in-situ* por bioestimulação, em que a adição de nutrientes permite aumentar a taxa de crescimento de microrganismos autóctones e, logo, as atividades metabólicas, com o consequente aumento da velocidade de biodegradação, foi uma das principais estratégias utilizadas para tratar o ambiente afetado pelo vazamento de 41 milhões de litros óleo do navio petroleiro *Exxon Valdez* em 1989 na costa do Alasca (Atlas & Hazen 2011).

Embora uma diversidade de microrganismos possa degradar com eficiência muitos poluentes, a mineralização de HAPs no solo é geralmente um processo lento, em razão da limitada biodisponibilidade do contaminante, além de fatores bióticos e abióticos vitais à biodegradação (Samanta et al 2002). Em adição, alguns xenobióticos mais tóxicos e recalcitrantes, como compostos aromáticos halogenados e nitrados, são geralmente estáveis e quimicamente inertes em condições naturais, estando além da capacidade de metabolização bacteriana (Cao et al 2009; Paul et al 2005).

Estas limitações abriram caminho para o desenvolvimento de organismos geneticamente modificados (OGMs) resistentes a condições adversas do ambiente, bem como capazes de realizar a degradação completa de contaminantes. Neste contexto, a otimização racional de OGMs tem sido facilitada pelo crescimento de bancos de dados genômicos de microrganismos associado ao conhecimento da relação estrutura-função de enzimas componentes de tais vias de degradação. Todavia, a aplicação de microrganismos geneticamente modificados no ambiente ainda é limitada devido aos riscos de competição com a microbiota nativa (Paul et al 2005).

Frente a isto, a bioremediação enzimática tem sido apontada como uma alternativa favorável para apoiar as técnicas de biotratamento atualmente disponíveis. Muitas são as vantagens sobre o uso de enzimas em detrimento de microrganismos, dentre as quais se citam (Ahuja et al 2004; Alcalde et al 2006; Sutherland et al 2004):

- a) Em virtude da especificidade das enzimas, a biotransformação por elas catalisada não gera subprodutos tóxicos, como normalmente acontece com o uso de produtos químicos e microrganismos;
- b) Após o tratamento, as enzimas não permanecem no ambiente, uma vez que podem ser degradadas *in situ* pela microbiota local;
- c) A biodisponibilidade das enzimas pode ser aumentada por meio da introdução de co-solventes orgânicos ou surfactantes;
- d) O uso de enzimas não gera biomassa como aquela produzida ao se utilizar microrganismos, principalmente em tratamentos aeróbicos.

Ainda que os benefícios da biocatálise sejam evidentes, os principais obstáculos que dificultam explorar o vasto repertório de processos enzimáticos são, na maioria dos casos,

baixos rendimentos obtidos e elevados custos de produção e comercialização, como o uso de enzimas, por exemplo, que dependam de cofatores (Peixoto et al 2011). Também, é imperativo que as condições ótimas para a atividade enzimática sejam fornecidas e mantidas durante todo o processo (Alcalde et al 2006). Assim, técnicas de biologia molecular, engenharia de proteínas e *design* racional de enzimas vêm sendo utilizadas no sentido de melhorar simultaneamente várias propriedades enzimáticas, como estabilidade e atividade, conferindo características aprimoradas a enzimas que podem ser utilizadas em processos de bioremediação (Ahuja et al 2004).

1.3. BIODEGRADAÇÃO DE NAFTALENO POR *Pseudomonas putida* G7

Espécies de *Pseudomonas* e microrganismos intimamente relacionados têm sido extensamente estudados em virtude de sua versatilidade em degradar uma gama de compostos aromáticos (Cao et al 2009).

Dentre os hidrocarbonetos aromáticos, o naftaleno, devido à sua estrutura mais simples e, logo, maior solubilidade, é frequentemente utilizado como um composto modelo em estudos que visam caracterizar a habilidade de bactérias em degradar HAPs. As informações obtidas sobre a degradação bacteriana do naftaleno são úteis na compreensão de vias catabólicas de HAPs complexos, com três ou mais anéis benzênicos (Seo et al 2009).

O metabolismo do naftaleno é bem caracterizado em *Pseudomonas putida* G7, bactéria gram-negativa cujo plasmídeo metabólico de 83 Kb, NAH7, permite ao hospedeiro utilizar naftaleno ou salicilato como única fonte de carbono e energia (Dunn & Gunsalus 1973).

Os genes do catabolismo do naftaleno (*nah*) em NAH7 organizam-se em dois operons: ***nah*** ou ***superior*** (*nahAaAbAcAdBFCED*), que codifica para enzimas envolvidas na conversão de naftaleno a salicilato, e ***sal*** ou ***inferior*** (*nahGTHINLOMKJXY*), cujos produtos gênicos *nahGTHINLOMKJ* oxidam salicilato a piruvato e acetaldeído pela via de *meta*-clivagem. Também presentes no operon *sal*, o gene *nahY* codifica para um quimiorreceptor sensível ao naftaleno, enquanto *nahX* não possui função conhecida (Grimm & Harwood 1999; Sota et al 2006; Yen & Gunsalus 1982) (Figura 3).

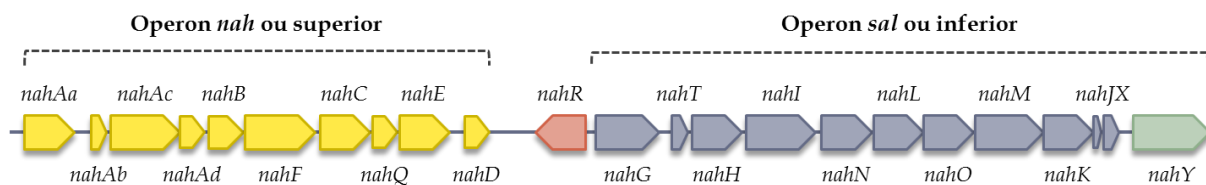


Figura 3: Representação esquemática dos operons *nah* e *sal* de NAH7. Os genes *nah* estão organizados em dois operons, *nah* e *sal*, que codificam para enzimas das vias superior e inferior de degradação do naftaleno. O gene não catabólico *nahY*, presente no operon *sal*, está envolvido na quimiotaxia ao naftaleno, enquanto *nahR* codifica para um ativador transcricional de ambos os operons. Os pentágonos indicam a orientação transcricional dos genes. Adaptado de (Sota et al 2006).

Situado entre os dois operons, o gene *nahR* codifica para um elemento regulatório positivo de ação *trans*, cuja família de reguladores transcricionais do tipo Lys-R (LTTR), a mais abundante entre procariotos, apresenta uma estrutura conservada com um motivo hélice-volta-hélice N-terminal para ligação ao DNA e um domínio C-terminal para ligação ao co-indutor. NahR é essencial para a expressão elevada de ambos os operons induzida por salicilato, o último produto da via superior de degradação do naftaleno (Maddocks & Oyston 2008; Yen & Gunsalus 1985).

Em *P. putida* G7, a via inicial de degradação do naftaleno, dita superior, tem como primeiro passo a incorporação de oxigênio molecular ao núcleo aromático do naftaleno para produzir *cis*-1,2-dihidroxi-1,2-dihidronaftaleno (*cis*-naftaleno dihidrodiol). Esta etapa é catalizada pelo complexo da naftaleno dioxigenase (NDO), compreendido por ferredoxina redutase (**NahAa**), ferredoxina (**NahAb**) e proteínas ferro-enzofre (**NahAc** e **NahAd**). A transferência de dois elétrons, um de cada vez, do NAD(P)H ao FAD é realizada pela ferredoxina redutase. A forma totalmente reduzida do FAD fornece elétrons para os centros 2Fe-2S da ferredoxina, que finalmente transfere para outras proteínas ferro-enzofre, onde serão utilizados no sítio ativo para facilitar a adição do oxigênio molecular ao naftaleno (Davies & Evans 1964; Habe & Omori 2003) (Figura 4).

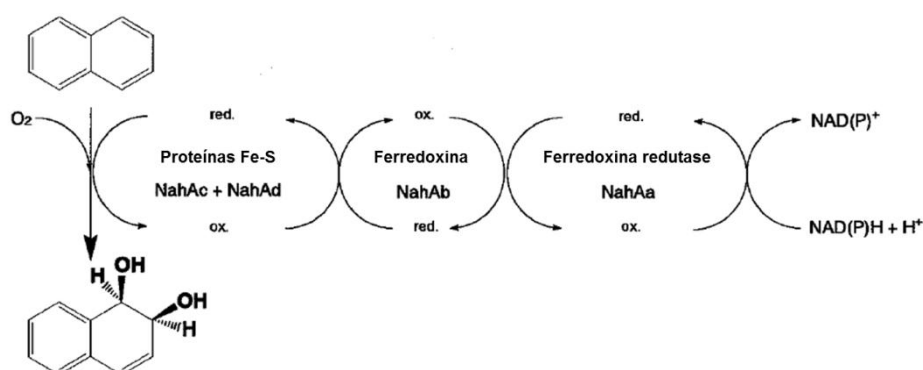


Figura 4: Oxidação inicial do naftaleno a cis-1,2-dihidroxi-1,2-dihidronaftaleno pelo complexo da naftaleno dioxigenase (NahAaAbAcAd). Os elétrons fornecidos pelo NADPH são transferidos através de todo o complexo da naftaleno dioxigenase, a fim de serem utilizados na adição do oxigênio molecular ao naftaleno.

As reações subsequentes são catalisadas pela *cis*-naftaleno dihidrodiol desidrogenase (**NahB**), 1,2-dihidroxi-naftaleno dioxigenase (**NahC**), 2-hidroxicromeno-2-carboxilato isomerase (**NahD**) e 2-hidroxibenzil-piruvato aldolase (**NahE**), resultando na formação de salicilaldeído, que é oxidado a salicilato pela salicilaldeído desidrogenase (**NahF**), a última enzima da via superior (Davies & Evans 1964) (Figura 5 A).

Já na via inferior de degradação do naftaleno, o salicilato é oxidado pela salicilato hidroxilase (**NahG**) a catecol, cujo anel aromático sofre uma clivagem extradiol pela catecol-2,3-dioxigenase (**NahH**) produzindo 2-hidroxi-muconato semialdeído (2-HMS) (Figura 5 B). Esta reação constitui a primeira etapa da via de *meta*-clivagem do catecol que, a partir deste ponto, diverge em dois ramos (Sala-Trepat et al 1972). No primeiro, dito multi-enzimático, uma 2-hidroxi-muconato semialdeído desidrogenase (**NahI**) oxida 2-HMS ao ácido correspondente, 2-hidroxi-muconato, que é subsequentemente convertido ao conjugado cetona pela 4-oxalocrotonato tautomerase (**NahJ**). A descarboxilação deste conjugado é catalisada pela 4-oxalocrotonato descarboxilase (**NahK**) a 2-hidroxi-penta-2,4-dienoato (HPD). Já no outro ramo da via de *meta*-clivagem, dito hidrolítico, uma 2-hidroxi-muconato semialdeído hidrolase (**NahN**) hidrolisa 2-HMS diretamente a HPD. Assim, ambos os ramos convergem para a formação de HPD, ao qual é adicionada água pela vinilpiruvato hidratase (**NahL**) resultando em 2-oxo-4-hidroxi-pentenoato (2,4-HP). Finalmente, este é convertido a piruvato e acetaldeído pela 2-oxo-4-hidroxi-pentenoato aldolase (**NahM**), os quais são canalizados para o ciclo de Krebs (Bosch et al 2000; Harayama et al 1987a; Peng et al 2008; Sala-Trepat et al 1972) (Figura 5 B).

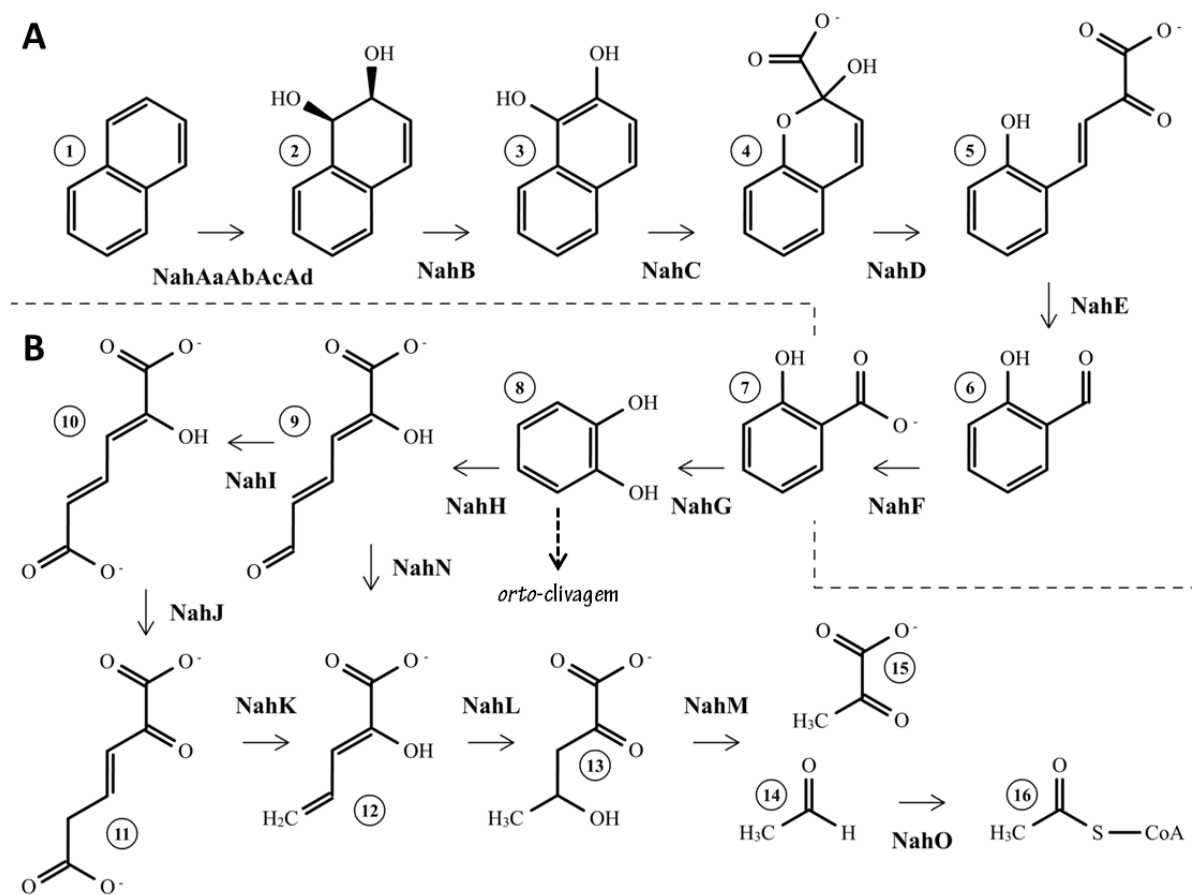


Figura 5: Vias superior e inferior de degradação do naftaleno. A. Via superior. **B.** Via inferior. Os nomes das enzimas estão descritos no texto. Os intermediários das vias são: (1) naftaleno, (2) *cis*-1,2-dihidroxi-1,2-dihidronaftaleno, (3) 1,2-dihidroxi-naftaleno, (4) 2-hidroxicromeno-2-carboxilato, (5) *trans*-o-hidroxi-benzilidenopiruvato, (6) salicilaldeído, (7) salicilato, (8) catecol, (9) 2-hidroxi-muconato semialdeído, (10) 2-hidroxi-muconato, (11) 4-oxalocrotonato, (12) 2-hidroxi-penta-2,4-dienoato, (13) 2-oxo-4-hidroxipentenoato, (14) acetaldeído, (15) piruvato e (16) acetil-CoA.

Outra possibilidade de clivagem do anel aromático do catecol é a intradiol, catalisada pela catecol-1,2-dioxigenase (Harayama & Reikik 1989), que compromete o intermediário na via de *orto*-clivagem, codificada pelo cromossomo bacteriano, originando *cis*-muconato que por passos sucessivos será convertido em succinil-CoA e acetil-CoA.

Cao e colaboradores (2008) sugerem a expressão simultânea das vias de *orto* e *meta*-clivagem após identificarem, por meio de espectrometria de massa, ambas as enzimas catecol-1,2-dioxigenase e catecol-2,3-dioxigenase em células de *P. putida* P8 crescidas em altas concentrações de benzoato, análogo de salicilato (Cao et al 2008).

Ainda assim, o direcionamento para uma ou outra via pode ser observado e reflete o fato de que os grupos substituintes em alguns substratos são incompatíveis com uma ou mais

enzimas de uma rota catabólica particular. Clorocatecóis, por exemplo, são geralmente degradados pela via de *orto*-clivagem, uma vez que, se submetidos à *meta*-fissão, intermediários tóxicos ou parcialmente metabolizados são formados. Por outro lado, a enzima catecol-1,2-dioxigenase apresenta afinidade reduzida para metilcatecóis sendo, portanto, canalizados para a via de *meta*-clivagem (Harayama et al 1987a). Dentro desta via, por sua vez, o produto da *meta*-fissão do 3-metilcatecol é conduzido para o ramo hidrolítico, pois não possui aldeído oxidável para ser metabolizado pela desidrogenase. Em contraste, o produto da *meta*-clivagem do catecol, 2-hidroximuconato semialdeído, pode ser degradado pelos dois ramos (Figura 5 B), ainda que a atividade catalítica da enzima desidrogenase se sobreponha à da hidrolase conforme verificado em células de *P. putida* portadoras de plasmídeo TOL (Harayama et al 1987a; Murray et al 1972).

1.4. A VIA DE META-CLIVAGEM DO CATECOL

A via de *meta*-clivagem do catecol é um tema recorrente no catabolismo aeróbico bacteriano sobre uma variedade de compostos aromáticos. Enquanto tolueno, naftaleno e fenol, a exemplo, são assimilados por diferentes reações enzimáticas a catecol, a conversão deste aos intermediários do ciclo de Krebs é mediada pela mesma rota catabólica, a via de *meta*-clivagem. Tal ocorrência em comum tem sido atribuída à transferência horizontal de uma via ancestral e sua integração em variados operons que degradam diferentes compostos aromáticos (Harayama & Rekik 1993).

Genes isofuncionais aos *nahTHINLOMKJX* de NAH7 são encontrados no plasmídeo TOL pWW0 de *P. putida* mt-2 (*xylTEGFJQKIH*) e em pVI150 de *P. putida* CF600 (*dmpQBCDEFGHI*), que codificam para enzimas envolvidas na degradação, respectivamente, de tolueno/xileno e metilfenol, com considerável identidade de sequência com aquelas codificadas pelos genes *nah* (Harayama & Rekik 1993; Harayama et al 1987b; Sota et al 2006) (Figura 6).

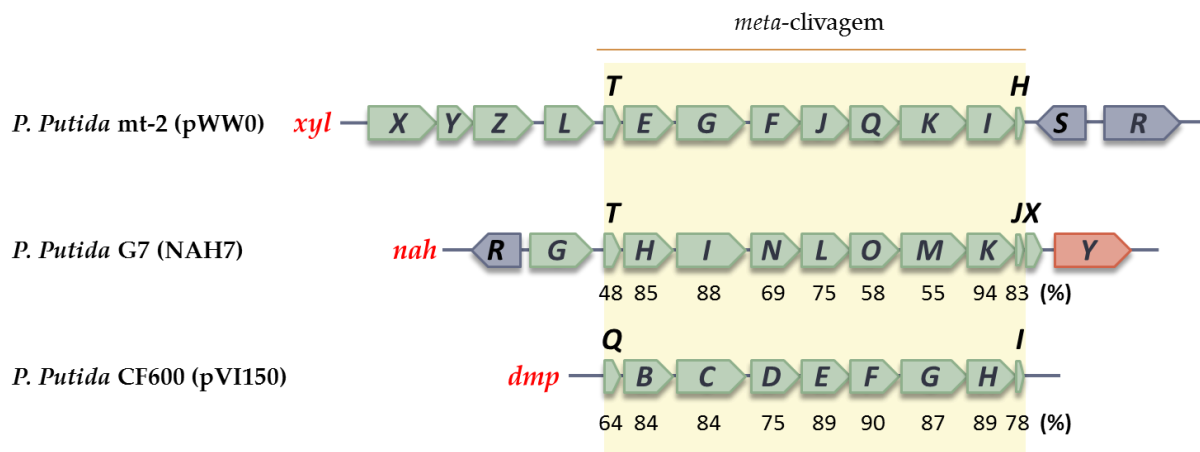


Figura 6: Organização dos genes catabólicos *xyl*, *nah* e *dmp* nos operons de via inferior / meta-clivagem codificados por pWW0, NAH7 e pVI150 para a degradação de tolueno/xileno, naftaleno e dimetilfenol, respectivamente. Os genes isofuncionais estão alinhados. É mostrada a porcentagem de identidade de sequência de aminoácidos entre os produtos gênicos das vias de *meta-clivagem* de pWW0 e os correspondentes em NAH7 e pVI150. Adaptado de (Harayama & Rekik 1993; Sota et al 2006).

O elevado grau de similaridade gênica entre estes operons se dá a partir do gene *nahT* em NAH7 (*xylT* em pWW0 e *dmpQ* em pVI150), situado a montante do gene da catecol 2,3-dioxigenase (*nahH*, *xylE* e *dmpB*) (Williams & Sayers 1994). Há evidências de que a proteína XylT e suas homólogas NahT e DmpQ, caracterizadas como ferredoxinas [2Fe-2S], sejam responsáveis em reativar a catecol 2,3-dioxigenase através da redução do átomo de ferro, de Fe^{3+} para Fe^{2+} , presente como cofator no sítio catalítico desta enzima (Hugo et al 1998; Hugo et al 2000).

A montante do gene *nahT*, a homologia entre as sequências *nah*, *xyl* e *dmp* é perdida e os genes, ainda pertencentes a vias inferiores, diferem. São eles: *nahG* em NAH7, para a conversão de salicilato a catecol, *xylXYZL* em pWW0 para a conversão de benzoato a catecol, e os componentes *dmpLMNOP* da fenol hidroxilase em pVI150 (Williams & Sayers 1994).

1.5. ALDEÍDO DESIDROGENASES

Aldeídos são moléculas altamente reativas, com ampla distribuição na natureza e uma variedade de efeitos em sistemas biológicos. Ainda que alguns dos efeitos mediados por aldeídos sejam benéficos, como a visão, por exemplo, muitos outros são citotóxicos, mutagênicos ou carcinogênicos aos organismos (Lindahl 1992).

Uma vez que aldeídos endógenos são formados durante o metabolismo de aminoácidos, carboidratos, lipídeos, aminas biogênicas, vitaminas e esteroides, os níveis de aldeídos como intermediários metabólicos devem ser cuidadosamente regulados. A oxidação de aldeídos a seus ácidos carboxílicos correspondentes é a via de eliminação mais efetiva, sendo catalisada por uma diversidade de enzimas aldeído desidrogenases (ALDHs), que participam assim de diversos processos fisiológicos (Lindahl 1992).

As ALDHs compõem uma superfamília de enzimas dependentes de NAD ou NADP como cofatores para a oxidação de uma variedade de aldeídos alifáticos e aromáticos. Membros da superfamília das ALDHs compartilham características estruturais e funcionais, ainda que um alinhamento de 145 sequências completas de aldeído desidrogenases tenha revelado apenas quatro resíduos invariantes, além de outros 12 encontrados em mais de 95% das sequências (Perozich et al 1999).

Neste mesmo estudo filogenético, as ALDHs agrupam-se em dois grandes ramos principais, Classe 3 e Classe 1/2, constituídos conjuntamente por no mínimo 13 famílias (Tabela 1). Longe de ser uma distinção absoluta, muitas famílias do ramo “Classe 3” são substrato-específicas, enquanto maior variabilidade sobre a especificidade ao substrato é encontrada entre as famílias da “Classe 1/2”, exceção feita para 10-formiltetrahidrofolato desidrogenase e 2-hidroximuconato semialdeído desidrogenase. Além destas duas, outras famílias de semialdeído desidrogenases (metilmalonil semialdeído desidrogenase, γ -glutamil semialdeído desidrogenase e succinato semialdeído desidrogenase) também apresentam estrita preferência pelo substrato, um indicativo de que a divergência evolutiva tenha ocorrido antecipadamente para estas ALDHs (Hempel et al 1993; Perozich et al 1999).

Membros da superfamília das ALDHs são classificados em ‘famílias’ e ‘subfamílias’ com base na porcentagem de identidade de cada sequência proteica comparada às outras. Sequências que apresentam menos de 40% de identidade entre si são consideradas como pertencentes a famílias diferentes, enquanto aquelas com identidade superior a 60% são agrupadas em uma mesma subfamília (Sophos & Vasiliou 2003).

Em relação à diversidade de estruturas quaternárias e à preferência pelos cofatores, não há distinção aparente entre as Classes 3 e 1/2 de ALDHs, pois famílias de ambos os ramos apresentam proteínas homotetraméricas e homodiméricas que parecem ligar NAD ou NADP (Perozich et al 1999).

Tabela 1: As 13 famílias de ALDHs. Adaptado de (Perozich et al 1999).

Ramo	Família	Especificidade ao substrato	Estrutura quaternária	Coenzima
Classe 3	Betaína ALDH	Específica	Tetrâmero	NAD
	Metilmalonil semialdeído desidrogenase (MMSALDH)	Específica	Tetrâmero	NAD e CoA
	γ -Glutamil semialdeído desidrogenase (GGSALDH)	Específica	Dímero	NAD
	ALDH de resposta ao turgor	n.d.	n.d.	n.d.
	ALDH classe 3	Variável	Dímero	NAD ou NADP
	ALDH de aldeídos aromáticos	Específica	Tetrâmero	NAD
	Succinato semialdeído desidrogenase	Específica	Tetrâmero	NAD (animais) e NADP (bactérias)
Classe 1/2	Gliceraldeído-3-fosfato desidrogenase (não fosforilativa)	Específica	Tetrâmero	NADP
	ALDH classe 1	Variável	Tetrâmero	NAD
	ALDH classe 2	Variável	Tetrâmero	NAD
	ALDH de fungos	Variável	Tetrâmero	NAD
	10-Formiltetrahidrofolato desidrogenase	Específica	Tetrâmero	NADP
	2-Hidroximuconato semialdeído desidrogenase (HMSALDH)	Específica	Dímero	NAD

ALDH – aldeído desidrogenase; n.d. – não determinado; CoA – Coenzima A.

A maioria das famílias de ALDHs é específica para NAD; poucas têm preferência por NADP, enquanto outras ainda precisam ser caracterizadas nesse sentido. Notavelmente, a família ALDH classe 3 é a única cuja habilidade em utilizar NAD ou NADP *in vitro* é bem estabelecida, embora, *in vivo*, dadas as concentrações fisiológicas de cada coenzima, estas enzimas sejam mais susceptíveis a funcionar apenas com NAD (Perozich et al 2001; Perozich et al 2000).

Pertencente à família ALDH classe 3, uma aldeído desidrogenase citosólica de fígado de rato foi a primeira destas enzimas a ter a estrutura tridimensional determinada (ALDH3; PDB id 1AD3), resolvida por cristalografia de raios-X a 2,6 Å de resolução (Figura 7) (Liu et

al 1997). Cada subunidade da enzima dimérica contém um domínio de ligação ao NAD(P), um domínio de ligação ao substrato e outro de oligomerização, também chamado de conector. A estabilização do dímero é dada por ligações de hidrogênio estabelecidas entre o domínio conector de uma subunidade e o domínio de ligação ao substrato da subunidade vizinha.

Em toda a estrutura, a cadeia polipeptídica adquire um enovelamento α/β . Como a maioria das desidrogenases que utilizam NAD ou NADP, o domínio de ligação ao cofator apresenta um motivo proteico chamado de dobra de *Rossmann*, que na ALDH é formado por cinco fitas- β associadas a quatro α -hélices, ao invés das seis fitas- β tipicamente comuns em outras desidrogenases dependentes de NAD, como é observado em uma álcool desidrogenase (ADH; PDB id 2OHX), por exemplo.

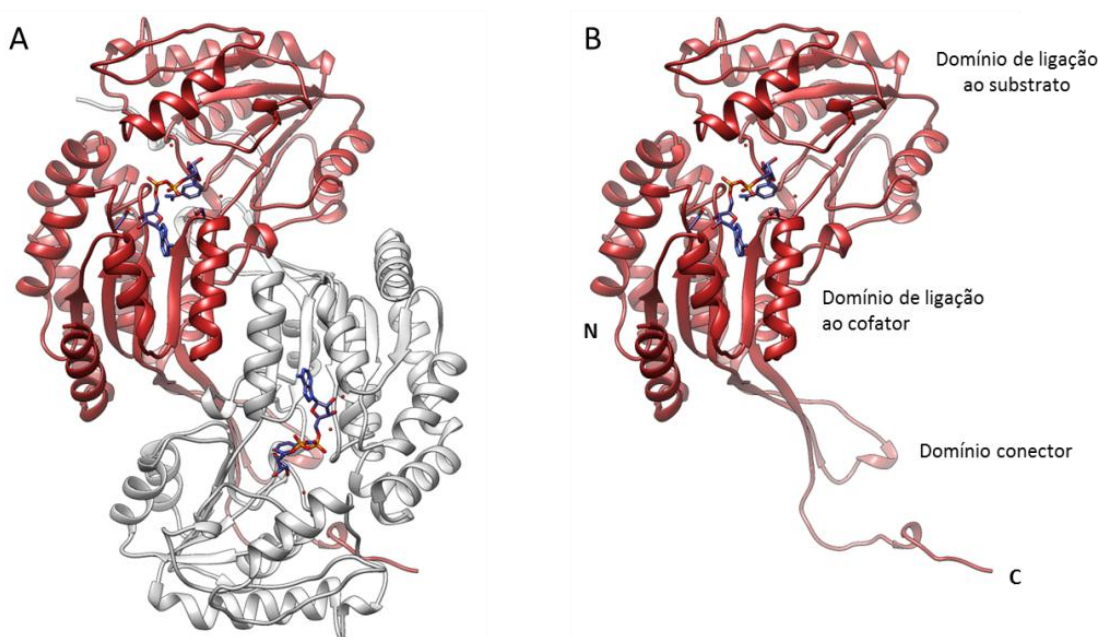


Figura 7: Estrutura tridimensional de uma ALDH classe 3 resolvida a 2,6 Å de resolução. A. Representação de fita do dímero da ALDH3, em que se evidencia um enovelamento α/β . Uma das subunidades está colorida de vermelho, e a outra de cinza. Destaque para a localização do NAD (estrutura em vareta). B. Representação de fita do monômero da ALDH. Os três domínios, de ligação ao cofator, ao substrato e conector, estão assinalados. As extremidades N e C-terminal estão indicadas. PDB id 1AD3 (Liu et al 1997).

Ainda na estrutura da ALDH classe 3, o sítio catalítico, formado por resíduos pertencentes aos três domínios, se posiciona na base de um funil a 15 Å da superfície, onde se localiza a cisteína catalítica (Liu et al 1997).

No mecanismo de ação proposto para as ALDHs (Figura 8), o NAD^+ interage primeiro com a enzima e, então, o carbono carbonílico do aldeído sofre o primeiro ataque nucleofílico pela cisteína catalítica formando uma ligação covalente tio-hemiacetal entre o substrato e o tiol reativo. Em seguida, o intermediário enzima-substrato é oxidado pelo NAD^+ havendo a transferência de um hidreto e concomitante formação de NADH e uma ligação tioéster. Na etapa limitante do processo, o glutamato catalítico ativa uma molécula de água que fará um ataque nucleofílico na ligação tioéster liberando o produto ácido (Ho et al 2006 ; $\text{Perez-Miller \& Hurley 2003}$; Wymore et al 2004).

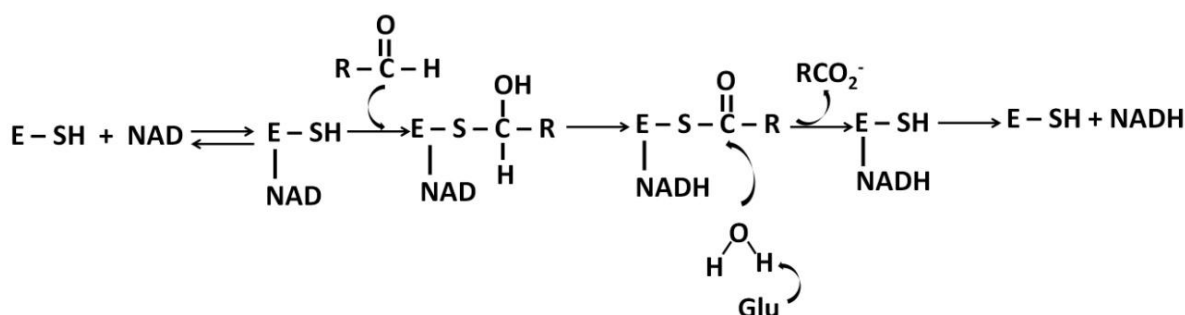


Figura 8: Mecanismo de ação proposto para ALDHs. O NAD interage com a enzima antes do substrato aldeído. A cisteína catalítica realiza um ataque nucleofílico no substrato aldeído formando um intermediário tio-hemiacetal. Ocorre então a transferência de um hidreto do intermediário para o NAD, formando NADH e um tioéster. Uma molécula de água é ativada pelo glutamato catalítico e faz um ataque nucleofílico no tioéster liberando o ácido carboxílico e o NADH. E-SH indica a enzima com a cisteína como o nucleófilo. Adaptado de (Wymore et al 2004).

Uma análise comparativa entre as sequências da ALDH classe 3 de rato e das ALDHs classe 1 e classe 2 de humano sugere que, a propósito da conservação de resíduos-chave no sítio catalítico, em regiões interagentes com o cofator e em porções estabilizadoras de elementos de estrutura secundária, todas as classes de ALDHs devem compartilhar um enovelamento estrutural semelhante, a despeito da baixa identidade de sequência (menor que 30% entre ALDH classe 1, 2 e 3) e de diferentes modos de oligomerização e especificidade ao substrato (Liu et al 1997).

De fato, a estrutura da ALDH classe 2 de boi (ALDH2; PDB id 1AG8), a forma mitocondrial da aldeído desidrogenase que catalisa a oxidação de acetaldeído a acetato durante o metabolismo do etanol e cuja identidade de sequência compartilhada com a isoenzima humana é de 95%, mas inferior a 30% com a ALDH3 de rato (PDB id 1AD3),

mostra claramente a manutenção do arcabouço estrutural, salvas as inúmeras particularidades concernentes ao estado de oligomerização e às interações estabelecidas entre enzima, cofator e substrato (Figura 9). Ao contrário da dimérica ALDH3, ALDH2 é um tetrâmero (Figura 10), também descrita como um “dímero de dímeros” segundo os autores (Steinmetz et al 1997).

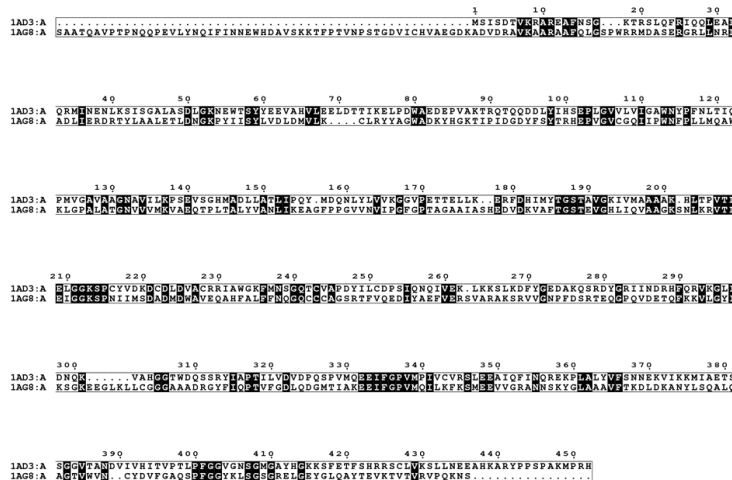


Figura 9: Comparação das estruturas primária e terciária de ALDH2 (1AG8) e ALDH3 (1AD3). ALDH2, em verde, e ALDH3, em roxo, apresentam estrutura terciária semelhante, ainda que a identidade de sequência (blocos em preto) compartilhada entre as duas seja de 28% (similaridade de 44%). A título de comparação, estão sobrepostas apenas as cadeias A de ambas.

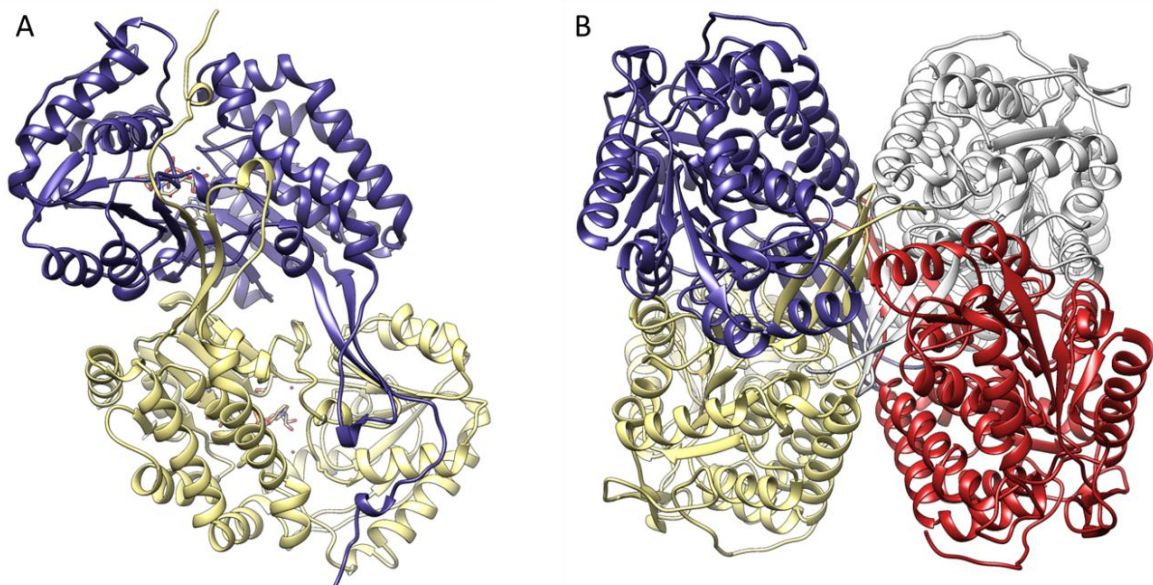


Figura 10: Estrutura quaternária de ALDH2 (1AG8) e ALDH3 (1AD3). Ao contrário da dimérica ALDH3 (A) (Liu et al 1997), ALDH2 (B) é um tetrâmero, ou um “dímero de dímeros” (Steinmetz et al 1997).

A primeira observação de uma ALDH hexamérica em solução se deu recentemente com Δ^1 -pirrolina-5-carboxilato desidrogenase (P5CDH) de *Thermus thermophilus* (Figura 11) e *Deinococcus radiodurans* (Luo et al 2013) e *Saccharomyces cerevisiae* (Pemberton et al 2014). Responsável por catalisar a oxidação de γ -glutamato semialdeído a L-glutamato na presença de NAD^+ , etapa final do catabolismo da prolina, a enzima P5CDH destes organismos teve seu estado oligomérico caracterizado como um trímero de dímeros, ao contrário das P5CDHs de humano, camundongo e *Bacillus* que chegam a formar dímeros, mas não se organizam em oligômeros maiores. Segundo os autores, o ponto crítico para a hexamerização de P5CDH reside no trio Ala100, Asp/Glu166 e Arg461, que estabelece interações nas interfaces do hexâmero, mas não se mantém conservado nas P5CDHs diméricas. Estados oligoméricos variados entre as P5CDHs podem sugerir ainda consideráveis diferenças em mecanismos de cooperatividade (Luo et al 2013; Pemberton et al 2014).

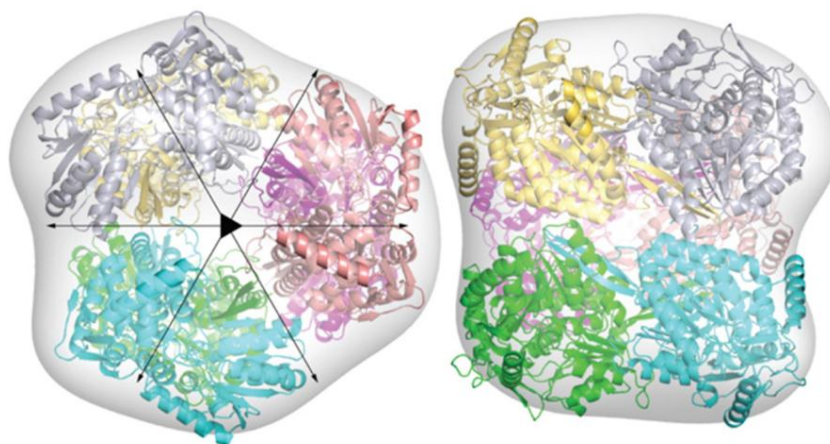


Figura 11: A ALDH hexamérica P5CDH de *Thermus thermophilus*. Sobreposição do envelope de SAXS e do hexâmero cristalográfico de TtP5CDH (PDB id 2BHQ) obtido a 1,54 Å de resolução. Extraído de (Luo et al 2013).

De volta à via de degradação do naftaleno de *P. putida*, a já mencionada salicilaldeído desidrogenase, NahF (PDB id 4JZE), cujo enovelamento estrutural se assemelha ao das outras ALDHs citadas, exibe clara preferência por aldeídos aromáticos ao catalisar a sua oxidação com elevados valores de k_{cat}/K_M próximos a $10^6 \text{ M}^{-1}\text{s}^{-1}$. Para aldeídos alifáticos, por sua vez, ainda que os valores de k_{cat}/K_M sejam inferiores aos determinados para os aromáticos, há um

aumento significativo na eficiência catalítica de NahF à medida que a cadeia alifática do substrato aldeído aumenta (Figura 12) (Coitinho 2013).

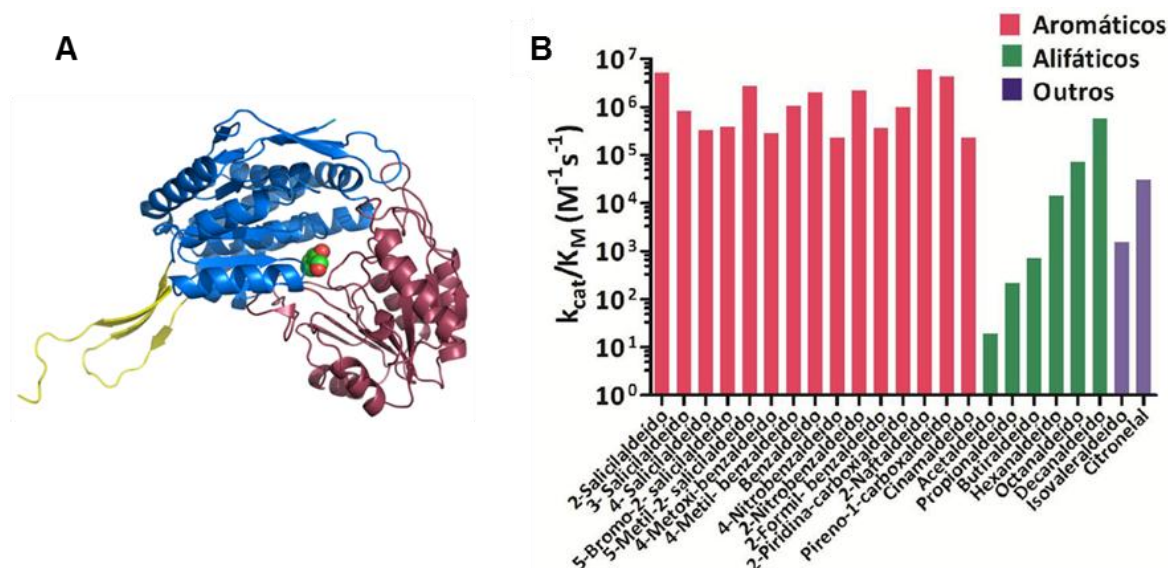


Figura 12: A enzima NahF, uma salicilaldeído desidrogenase da via de degradação do naftaleno de *Pseudomonas putida* G7. A. Estrutura tridimensional do monômero de NahF determinada por cristalografia de raios-X a 2,42 Å de resolução. Destaque para os três domínios: em azul, domínio de ligação ao dinucleotídeo, em vermelho, domínio catalítico, e em amarelo, domínio de oligomerização. A unidade biológica de NahF é dimérica (não mostrado). PDB id 4JZE. B. Atividade de NahF sobre diferentes aldeídos, em que se observa a preferência da enzima por aldeídos aromáticos e alifáticos de cadeia longa. Extraído de (Coitinho 2013).

Também presente em *P. putida* G7, a enzima NahI, uma 2-hidroxiuconato semialdeído desidrogenase atuante na via inferior de degradação do naftaleno, catalisa a conversão de 2-hidroxiuconato semialdeído (2-HMS) a 2-hidroxiuconato (2-HM) (Figura 13).

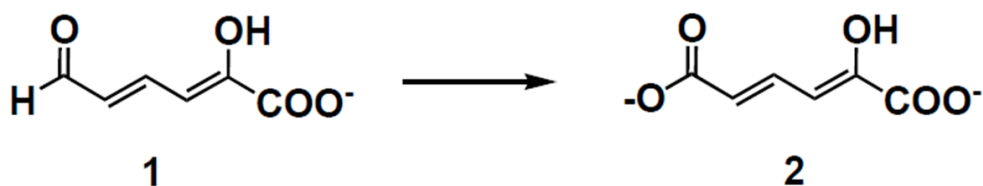


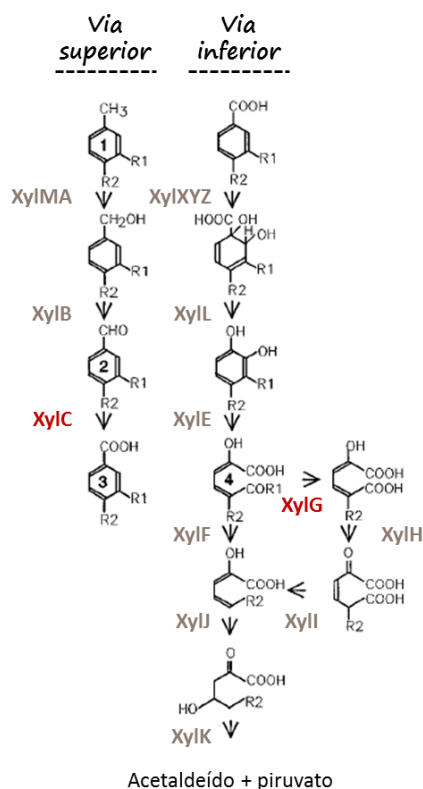
Figura 13: 2-hidroxiuconato semialdeído (1) e 2-hidroxiuconato (2). Substrato e produto da reação catalisada por 2-hidroxiuconato semialdeído desidrogenase.

Segundo a classificação filogenética de Perozich e colaboradores (1999) apresentada na tabela 1, membros da família HMSALDH apresentam preferência mais restrita ao substrato, sugerindo que a especialização de tais enzimas seja o reflexo de uma divergência evolutiva de longa data (Hempel et al 1993; Perozich et al 1999).

Curiosamente, de acordo com o modelo de evolução modular proposto para as vias catabólicas de hidrocarbonetos aromáticos em *Pseudomonas* (Williams & Sayers 1994), a aquisição de uma via de *meta*-clivagem do catecol a piruvato e acetil-CoA teria sido o primeiro passo (Bosch et al 2000), uma vez que se trata da metabolização de um composto mais simples, contendo apenas um anel aromático. As etapas subsequentes seriam a aquisição de um gene adicional, *nahG*, de modo a compor a via inferior, a incorporação do operon da via superior, *nahAaAbAcAdBFCED*, e finalmente a aquisição de um gene regulatório, *nahR* (Bosch et al 2000; Williams & Sayers 1994). Corrobora para isto a observação de que, na degradação aeróbica de compostos aromáticos, enzimas das vias superiores, ditas vias periféricas, reconhecem poluentes estruturalmente diversos e os transformam a um número limitado de intermediários, os quais são posteriormente canalizados ao metabolismo central através de vias de clivagem do anel, conhecidas como inferiores (Cao et al 2009).

Dentro deste contexto, é bem provável que NahI seja mais específica ao substrato quando comparada a NahF pelo fato de pertencer a uma via canalizadora, que recebe um número restrito de intermediários, enquanto a outra atua em uma via captadora e, portanto, mais abrangente.

Entretanto, de forma intrigante, é reportado (Inoue et al 1995) que a enzima XylG, uma 2-HMS desidrogenase homóloga de NahI, envolvida, porém, na via de degradação de xilenos e tolueno codificada pelo plasmídeo pWW0 de *P. putida* mt-2, pode oxidar não somente 2-HMS ($k_{cat}/K_M = 1,6 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$) como também substratos aromáticos, tais como benzaldeído ($k_{cat}/K_M = 3,2 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$) e seus análogos (Figura 14 / Tabela A). Por sua vez, a enzima XylC, uma benzaldeído desidrogenase com papel semelhante ao de NahF, oxida eficientemente benzaldeído ($k_{cat}/K_M = 1,7 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$), contudo não apresenta qualquer atividade sobre substratos alifáticos (Figura 14 / Tabela B). Com base nestes dados, os autores sugerem que a conformação estrutural do sítio ativo de XylG pode aceitar uma ampla gama de substratos ao contrário de XylC (Inoue et al 1995).

Tabela A. Constantes cinéticas de XylG (Inoue *et al.* 1995).

Substrato	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
Benzaldeído	460	17	$3,2 \times 10^4$
3-Metilbenzaldeído	630	26	$3,6 \times 10^4$
4-Metilbenzaldeído	530	6,2	$1,0 \times 10^4$
3-Clorobenzaldeído	1.300	32	$2,2 \times 10^4$
3-Nitrobenzaldeído	810	190	20×10^4
3-Metoxibenzaldeído	940	9	$0,8 \times 10^4$
3-Fluorobenzaldeído	170	64	33×10^4
NAD ⁺	330 ± 13	$21 \pm 1,2$	$5,6 \times 10^4$
2-Hidroxiuconato semialdeído	17 ± 6	27 ± 9	$1,6 \times 10^6$
2-Hidroxi-5-metil-6-oxohexa-2,4-dienoato	$9,3 \pm 2$	22 ± 3	$2,1 \times 10^6$

Tabela B. Constantes cinéticas de XylC (Inoue *et al.* 1995).

Substrato	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
Benzaldeído	$2,5 \pm 1,0$	48 ± 10	$1,7 \times 10^7$
3-Metilbenzaldeído	$1,9 \pm 0,7$	31 ± 4	$1,4 \times 10^7$
4-Metilbenzaldeído	$1,5 \pm 0,2$	28 ± 1	$1,6 \times 10^7$
3-Clorobenzaldeído	$2,1 \pm 0,5$	48 ± 4	$1,9 \times 10^7$
3-Nitrobenzaldeído	$22 \pm 6,0$	121 ± 14	$0,5 \times 10^7$
3-Metoxibenzaldeído	$2,6 \pm 0,5$	38 ± 3	$1,2 \times 10^7$
3-Fluorobenzaldeído	$2,9 \pm 0,8$	53 ± 6	$1,6 \times 10^7$
NAD ⁺	320 ± 59	47 ± 6	$1,3 \times 10^5$
2-Hidroxiuconato semialdeído	-	-	-
2-Hidroxi-5-metil-6-oxohexa-2,4-dienoato	-	-	-

Figura 14: Aldeído desidrogenases, XylC e XylG, envolvidas na degradação de tolueno e xilenos por *Pseudomonas putida* mt-2. À esquerda, vias superior e inferior de degradação de tolueno e xilenos codificadas pelo plasmídeo TOL pWW0. A partir da reação catalisada por XylE, uma catecol-2-3 dioxigenase, as enzimas da via de *meta*-clivagem do catecol são homólogas àquelas codificadas pelo plasmídeo NAH7, conforme reportado na subseção 1.4. Composto 1: tolueno ($R_1 = \text{H}$, $R_2 = \text{H}$), *m*-xileno ($R_1 = \text{CH}_3$, $R_2 = \text{H}$), *p*-xileno ($R_1 = \text{H}$, $R_2 = \text{CH}_3$). Composto 2: benzaldeído ($R_1 = \text{H}$, $R_2 = \text{H}$), *m*-metilbenzaldeído ($R_1 = \text{CH}_3$, $R_2 = \text{H}$), *p*-metilbenzaldeído ($R_1 = \text{H}$, $R_2 = \text{CH}_3$). Composto 3: benzoato ($R_1 = \text{H}$, $R_2 = \text{H}$), *m*-toluato ($R_1 = \text{CH}_3$, $R_2 = \text{H}$), *p*-toluato ($R_1 = \text{H}$, $R_2 = \text{CH}_3$). Composto 4: 2-hidroxiuconato semialdeído ($R_1 = \text{H}$, $R_2 = \text{H}$), 2-hidroxi-6-oxohepta-2,4-dienoato ($R_1 = \text{CH}_3$, $R_2 = \text{H}$), 2-hidroxi-5-metil-6-oxohexa-2,4-dienoato ($R_1 = \text{H}$, $R_2 = \text{CH}_3$). À direita, tabelas das constantes cinéticas de (A) XylG, uma 2-hidroxiuconato semialdeído desidrogenase, e (B) XylC, uma benzaldeído desidrogenase. Extraído e adaptado de (Inoue *et al.* 1995).

Assim, ainda que membros da superfamília ALDH apresentem um enovelamento estrutural comum, a elucidação da estrutura tridimensional de NahI e sua comparação com NahF e outras ALDHs permitirá inferir distinções no sítio catalítico que podem ser responsáveis por mudanças sistemáticas nas propriedades cinéticas destas enzimas sobre uma

variedade de substratos aldeídos. A caracterização de NahI irá expandir o entendimento de aldeído desidrogenases em geral e, em particular, a categoria das semialdeído desidrogenases. Por sua vez, o acúmulo de estudos de caracterização cinética e estrutural de enzimas e a compreensão das diferenças estruturais que as fazem únicas em suas reações podem permitir, por exemplo, a criação de enzimas degradadoras de naftaleno mais efetivas através de modificações nas já existentes, sendo assim útil em processos de bioremediação (Habe & Omori 2003).

2. OBJETIVOS

2.1. GERAL

Caracterizar cinética e estruturalmente a proteína recombinante NahI, uma 2-hidroxiomuconato semialdeído desidrogenases de *Pseudomonas putida* G7, bem como formas mutantes recombinantes da enzima.

2.2. ESPECÍFICOS

2.2.1. *NahI*

- Determinar a estrutura cristalográfica da enzima NahI recombinante em continuidade a um trabalho prévio desenvolvido em nosso laboratório;
- Elucidar o estado oligomérico da proteína em solução;
- Caracterizar cineticamente a NahI frente a aldeídos alifáticos e aromáticos.

2.2.2. *Mutantes de NahI*

- Obter as formas mutantes recombinantes de NahI em estado puro;
- Realizar ensaios de cristalização das enzimas recombinantes;
- Coletar dados de difração de raios-X de cristais obtidos;
- Determinar as estruturas cristalográficas;
- Caracterizar cineticamente os mutantes de NahI frente a aldeídos alifáticos e aromáticos.

3. MATERIAIS E MÉTODOS

3.1. PLASMÍDEOS

As construções plasmidiais utilizadas neste trabalho estão relacionadas no Quadro 1. Formas mutantes da enzima NahI de *Pseudomonas putida* G7 foram obtidas por mutagênese sítio-dirigida a partir do gene *nahI* (Gene Id: 3974218) previamente clonado no vetor pET28a-TEV em trabalho anterior (Neves 2013). Foram enviados 4 µg do vetor pET28a-TEV-*nahI* em solução para a *Genscript USA Inc.*, empresa responsável pelos serviços.

Quadro 1: Construções Plasmidiais

Construções Plasmidiais	Descrição
pET28a-TEV- <i>nahI</i> *	Gene <i>nahI</i> de <i>P. putida</i> G7 clonado no vetor pET28a-TEV.
pET28a-TEV- <i>nahI</i> -L156G	Mutante de NahI, no qual a Leucina 156 foi substituída por Glicina.
pET28a-TEV- <i>nahI</i> -F448S	Mutante de NahI, no qual a Fenilalanina 448 foi substituída por Serina.
pET28a-TEV- <i>nahI</i> -L156G/F448S	Duplo mutante de NahI: L156→G e F448→S.

*(Neves 2013).

O vetor pET28a-TEV (Figura 15), construído por Carneiro e colaboradores a partir do vetor pET28a (Novagen), adiciona uma sequência de seis resíduos de histidina e sítio de clivagem pela protease do vírus *etch* do tabaco (*tobacco etch virus* - TEV) (Carneiro et al 2006) na extremidade N-terminal da proteína de interesse (MGHHHHHHENLYFQGH – proteína).

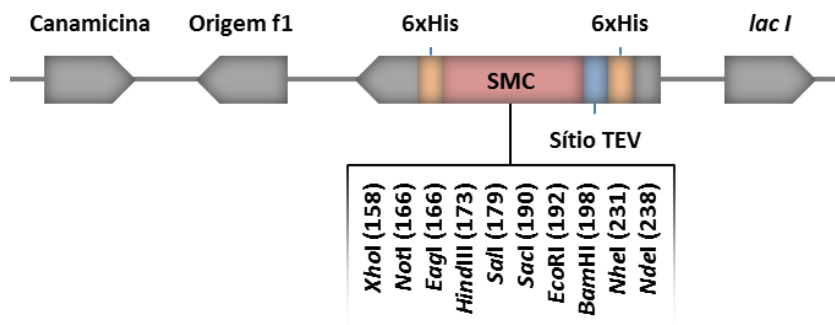


Figura 15. Desenho esquemático do vetor de expressão pET28a-TEV (5357 pb). A cauda de histidinas (6xHis) seguida do sítio de clivagem para a protease TEV são adicionados à sequência da proteína recombinante na extremidade N-terminal. SMC: sítio múltiplo de clonagem.

3.2. ANTIBIÓTICOS

Os antibióticos utilizados, bem como suas respectivas concentrações de estoque e de uso no meio de cultura, além de plasmídeo e/ou cepa que conferem resistência para tais, são listados no Quadro 2.

Quadro 2: Antibióticos

Antibiótico	Plasmídeo / Cepa	Concentração estoque	Concentração final
Canamicina (sulfato)	pET28a-TEV	50 mg/mL	50 µg/mL
Gentamicina (sulfato)	ArcticExpress™	40 mg/mL	20 µg/mL

Em condições de assepsia, as soluções de estoque foram preparadas com água ultrapura estéril e esterilizadas por filtração (0,22 µm), exceto quando compradas em ampolas, e então armazenadas a -20 °C.

3.3. CEPAS BACTERIANAS

Cepas bacterianas de *Escherichia coli* DH5α™ (Life Technologies) e ArcticExpress™ (DE3) (Stratagene) foram utilizadas no preparo de células eletrocompetentes para transformação.

Bactérias da linhagem DH5 α foram utilizadas para replicação das construções plasmidiais. Esta cepa de *E. coli* possui uma mutação no gene *endA*, representada no genótipo por *endA1*. O gene *endA* selvagem codifica para a endonuclease I, capaz de degradar DNA de dupla fita. A linhagem mutante, referida como EndA⁻, produz uma forma inativa da nuclease, o que permite um significativo aumento em qualidade e rendimento de DNA plasmidial usado para sequenciamentos ou outros ensaios (Taylor et al 1993).

Para expressão das proteínas recombinantes foi utilizada a linhagem ArcticExpress (DE3), derivada de *E. coli* B e deficiente das proteases Lon e OmpT, as quais podem degradar proteínas durante a purificação. Como um lisógeno λ -DE3, esta cepa contém integrado ao seu cromossomo bacteriano, e sob controle do promotor *lacUV5*, o gene que codifica para a enzima T7 RNA polimerase, cuja indução da expressão com IPTG resulta na expressão aumentada de genes clonados em vetores que possuem o promotor T7. Mais notavelmente, a cepa ArcticExpress apresenta um plasmídeo que codifica para as chaperoninas Cpn10 (10 kDa) e Cpn60 (57 kDa) de *Oleispira antarctica*, as quais possuem, respectivamente, 74% e 54% de identidade de sequência de aminoácidos com GroEL e GroES de *E. coli*. Ao demonstrarem elevada atividade de enovelamento proteico em temperaturas de 4 a 12 °C, Cpn10 e Cpn60, quando co-expressas em ArcticExpress, auxiliam no processamento de proteínas expressas em baixas temperaturas e podem, assim, elevar o rendimento de proteínas recombinantes solúveis e ativas (Agilent Technologies 2010).

3.4. PREPARO DE CÉLULAS ELETROCOMPETENTES

Células eletrocompetentes de *E. coli* foram preparadas de acordo com protocolo presente no manual do eletroporador *MicroPulser*TM (Bio-Rad), conforme descrito a seguir.

Inicialmente, o pré-inóculo da cepa de interesse foi feito em 5 mL de meio LB (*triptona 10 g/L; extrato de levedura 5 g/L; cloreto de sódio 5 g/L*) sem antibiótico e incubado a 37 °C por 16 horas sob agitação constante de 150 rpm. Exceção feita para a cepa ArcticExpress (DE3), que teve o antibiótico gentamicina adicionado ao meio de cultura. Após o crescimento, a cultura foi inoculada em 500 mL de meio LB sem antibiótico e cultivada a 37 °C sob agitação de 200 rpm até ser atingida a DO₆₀₀ de 0,5 a 0,7. Uma vez alcançado o nível de crescimento desejado, a cultura foi resfriada em gelo por 20 minutos ao fim dos quais as células foram coletadas por centrifugação a 4.000 *g*, a 4 °C por 15 minutos. Feito o descarte

do sobrenadante, ressuspendeu-se as células em 500 mL de glicerol 10% estéril e gelado. Outros três ciclos de centrifugação e ressuspensão foram realizados, respectivamente, em 250 mL, 20 mL e 2 mL de glicerol 10%, conforme descrito acima. Finalmente, as células ressuspendidas em 2 mL de glicerol 10% gelado foram aliquotadas em microtubos. As alíquotas de 50 µL foram imediatamente incubadas em gelo e armazenadas em freezer -80 °C, estando assim prontas para uso.

3.5. TRANSFORMAÇÃO DE BACTÉRIAS POR ELETROPORAÇÃO

Em alíquotas de 50 µL de bactérias *E. coli* eletrocompetentes (DH5α e ArcticExpress) mantidas em gelo por cerca de 5 minutos, adicionou-se 2 µL de vetor. As células foram transferidas para uma cubeta apropriada, tamanho de 0,2 cm. Após a eletroporação (2,50 kV, 1 pulso, em eletroporador *MicroPulser*TM da Bio-Rad), foram adicionados 300 µL de meio 2xYT (*triptona 16 g/L; extrato de levedura 10 g/L; cloreto de sódio 5 g/L*) líquido, e incubou-se a 37 °C por 1 hora sob agitação de 200 rpm para expressão do gene de resistência. Posteriormente, as bactérias foram plaqueadas em meio seletivo 2xYT - ágar 1,5% (p/v) contendo o antibiótico adequado. As placas foram incubadas a 37 °C em estufa bacteriológica por 16 horas para obtenção de colônias. Os antibióticos utilizados, bem como suas respectivas concentrações de estoque e de uso estão descritos no item 3.2.

3.6. EXPRESSÃO HETERÓLOGA EM *E. coli*

Colônias únicas de *E. coli* ArcticExpress (DE3) previamente transformadas e plaqueadas (item 3.5) foram inoculadas separadamente em meio de cultura 2xYT na presença dos antibióticos gentamicina e canamicina e, então, incubadas a 37 °C sob agitação de 200 rpm. Alternativamente, fez-se o pré-inóculo a partir de alíquotas de culturas, conservadas em glicerol e armazenadas em freezer (-80 °C), diluídas 1000 vezes. Após 16 horas de crescimento, as culturas iniciais foram diluídas 1:50 em meio 2xYT sem antibiótico e cultivadas por 3 horas a 30 °C e 200 rpm, quando fez-se a indução com IPTG a uma concentração final de 0,5 mM. A expressão prosseguiu a 12 °C e em agitação constante de 200 rpm ao longo de 24 horas. Alíquotas no início e no término da expressão foram coletadas e submetidas à eletroforese em gel de poliacrilamida-SDS.

Após o período de expressão, as células foram coletadas por centrifugação (5.000 *g*, a 4 °C por 15 minutos), o sobrenadante descartado e o precipitado celular congelado (-20 °C) ou imediatamente submetido à lise.

3.7. LISE CELULAR

O precipitado celular proveniente de 1,5 litros de cultura foi ressuspensionado em 50 mL de tampão A (*Fosfato de Sódio 0,02 M pH 7,4; NaCl 0,5 M*) e lisado em homogeneizador *Emulsiflex*[®]-C3 (Avestin) sob pressão de 15.000 a 20.000 psi, realizando-se 3 passagens da suspensão celular pelo aparelho. O sobrenadante foi separado da fração insolúvel através de centrifugação a 10.000 *g*, a 4 °C por 20 minutos.

3.8. PURIFICAÇÃO DAS PROTEÍNAS RECOMBINANTES

Cada proteína recombinante, fusionada à cauda de histidinas N-terminal, foi inicialmente purificada através de cromatografia de afinidade a níquel em sistema *ÄKTAprime plus* (GE Healthcare). O extrato foi aplicado em coluna *HisTrap HP* de 5 mL (GE Healthcare) previamente ambientada com tampão A. Após a aplicação da amostra, a coluna foi lavada com tampão A acrescido de imidazol 30 mM até que nenhuma leitura de absorvância fosse observada a 280 nm. A eluição das proteínas foi feita em concentração crescente de imidazol utilizando tampão B (*Fosfato de Sódio 0,02 M pH 7,4; NaCl 0,5 M; Imidazol 0,5 M*).

Para reduzir a concentração de cloreto de sódio e remover o imidazol, frações contendo proteína purificada foram reunidas e aplicadas em coluna de dessalinização *HiPrep 26/10* (GE Healthcare) previamente ambientada com tampão *Tris-HCl 0,05 M pH 7,4; NaCl 0,05 M*, sendo a eluição feita em mesmo tampão.

Para cada proteína recombinante estudada, metade da amostra foi submetida à proteólise da cauda de histidinas pela protease TEV a 5% (p/p) em tampão *Tris-HCl 0,05 M pH 7,4; NaCl 0,05 M; DTT 1 mM*, por 16 horas a 4 °C.

A protease TEV é produzida no próprio laboratório BioEst (ICB/UFMG) a partir do vetor pMHT238Δ. Sua atividade é sítio-específica e reconhece a sequência de resíduos

ENLYFQS, ou ENLYFQG, presente entre a etiqueta de afinidade e o início da sequência da proteína recombinante. A clivagem ocorre após o resíduo de glutamina (Taylor et al 1993).

A amostra foi novamente purificada por cromatografia de afinidade a níquel conforme protocolo descrito acima, objetivando retirar a protease TEV que, por apresentar em sua extremidade N-terminal uma cauda de histidinas, permanece retida na coluna juntamente com contaminantes provenientes da primeira etapa de purificação. Entretanto, a proteína recombinante clivada, por não ter a etiqueta de afinidade, não interage com a coluna e é recolhida já nas primeiras frações.

Finalmente, amostras de proteína completa e clivada foram concentradas separadamente por ultrafiltração em *Vivaspin* (GE Healthcare) com limite de exclusão molecular de 10.000 Da e, então, submetidas à cromatografia de exclusão molecular em coluna *HiLoad 16/60 Superdex 200 prep grade* (GE Healthcare) (faixa de separação de 10.000 a 600.000 Da). Para tal, a coluna foi ambientada em tampão *Fosfato de Sódio 0,05 M pH 7,4; NaCl 0,05 M*, no qual a proteína foi eluída mantendo-se sempre fluxo constante de 1,0 mL/min. Armazenou-se a proteína purificada em geladeira a 4 °C.

O volume vazio da coluna de exclusão molecular foi calculado por meio da aplicação de *Blue Dextran*, enquanto a calibração da coluna foi realizada aplicando-se individualmente as seguintes proteínas estabelecidas como padrões de massa molecular: citocromo C (12,4 kDa), albumina (66 kDa), álcool desidrogenase (150 kDa) e β -amilase (200 kDa).

Amostras recolhidas ao longo de todo o processo foram submetidas à eletroforese em gel de poliacrilamida-SDS.

Todos os ensaios descritos neste trabalho sobre NahI referem-se à proteína recombinante clivada, sem a cauda de histidinas e apenas com os dois resíduos remanescentes de glicina e histidina (GH), exceção feita para a etapa de cristalização, que foi conduzida apenas com 6xHis-NahI.

3.9. DETERMINAÇÃO DA CONCENTRAÇÃO PROTEICA

Fez-se a dosagem das proteínas através da medida da absorvância a 280 nm em espectrofotômetro *Nanodrop*[®]. Uma vez conhecidas as sequências de resíduos de

aminoácidos das proteínas em estudo, é possível prever o coeficiente de extinção molar de cada e aplicar a fórmula de *Lambert-Beer*,

$$A = \epsilon \times l \times c$$

na qual A é a absorvância a 280 nm; ϵ , o coeficiente de extinção molar dado em $M^{-1}cm^{-1}$; l , o caminho óptico da cubeta em cm, e c , a concentração molar da proteína. O coeficiente de extinção molar a 280 nm para proteínas em água, segundo Pace e colaboradores (1995), pode ser predito através da equação:

$$\epsilon (280) (M^{-1}cm^{-1}) = (\#Trp \times 5500) + (\#Tyr \times 1490) + (\#cistina \times 125)$$

em que $\#$ corresponde, na devida ordem, ao número de resíduos de triptofano e de tirosina, bem como pontes dissulfeto presentes na proteína (Pace et al 1995).

Os valores de ϵ para cada proteína, fusionada ou não à cauda de histidinas, foram obtidos através do programa *ProtParam* (Wilkins et al 1999) hospedado no servidor proteômico *ExPASy* (Gasteiger et al 2003). Para 6xHis-NahI e formas mutantes, $\epsilon = 57410 M^{-1}cm^{-1}$ e MM = 53,5 kDa. Após clivagem da cauda de histidinas, $\epsilon = 55920 M^{-1}cm^{-1}$ e MM = 51,7 kDa.

3.10. ESPALHAMENTO DINÂMICO DE LUZ

Ensaio de espalhamento dinâmico de luz (DLS) foram realizados à temperatura de 25°C em aparelho Zetasizer Nano ZS-90 (Malvern). Anterior a isto, a amostra proteica proveniente de cromatografia de exclusão molecular foi centrifugada a 13.000 g por 10 minutos a 4 °C para retirar impurezas ou partículas insolúveis presentes em suspensão. Para cada amostra, dependendo da concentração de proteína, medições foram realizadas por 10 a

100 vezes (10 segundos cada) em cubeta de quartzo de 10 mm de caminho óptico. Os dados foram coletados e analisados utilizando o programa Malvern Zetasizer v. 6.01, através do qual foram obtidas estimativas de massa molecular baseadas no raio hidrodinâmico (R_h) da partícula.

3.11. DICROÍSMO CIRCULAR

O conteúdo de estrutura secundária de NahI, na presença ou ausência de NAD^+ , foi avaliado por espectroscopia de dicroísmo circular (CD, do inglês *Circular Dichroism*) usando um espectropolarímetro JASCO J-815 (Jasco Corporation). As amostras, a uma concentração de 6 μM em tampão *Fosfato de Sódio 0,05 M pH 7,4; NaCl 0,05 M*, foram centrifugadas (13.000 g por 10 minutos a 4 °C) antes de cada ensaio. Para as amostras com NAD^+ , o cofator foi adicionado a uma concentração molar final 10 vezes superior à concentração de proteína e, então, realizada incubação em gelo por 1 hora antes da centrifugação. Espectros de CD no UV distante (190 - 260 nm) foram registrados ao longo do intervalo de 20 a 95 °C com incrementos de temperatura a cada 5 °C, velocidade de varredura de 100 nm/min, passo de 1,0 nm e 5 acumulações por digitalização. A mesma cubeta de quartzo com 0,5 milímetro de caminho óptico foi usada ao longo dos ensaios. Os espectros foram corrigidos à linha de base subtraindo o espectro da solução tampão (*Fosfato de Sódio 0,05 M pH 7,4; NaCl 0,05 M*) utilizada nas análises.

O conteúdo de elementos de estrutura secundária de NahI foi calculado por meio do programa CDNN v.2.1 (Bohm et al 1992). A identificação dos elementos de estrutura secundária a partir de estruturas tridimensionais de proteínas homólogas depositadas no PDB foi realizada por meio do servidor *STRIDE* (Frishman & Argos 1995; Heinig & Frishman 2004).

3.12. CRISTALIZAÇÃO DE PROTEÍNAS E DIFRAÇÃO DE RAIOS-X

Ensaio de cristalização foram realizados segundo a técnica de difusão de vapor em gota suspensa utilizando placas de cristalização *Limbbox* de 24 poços (Hampton Research), lamínulas silanizadas e graxa de silicone.

Nessa técnica, a solução de proteína, juntamente com uma solução de cristalização adequada, é adicionada a uma lamínula na forma de uma gota. A lamínula é então invertida sobre um recipiente contendo 1 mL da mesma solução de cristalização, vedando-o hermeticamente por meio da graxa de silicone. Lentamente, por difusão de vapor, a concentração de precipitante das gotas, inicialmente menor, vai se igualando à concentração da solução presente no poço. Devido ao aumento de concentração, a gota pode chegar lentamente à condição de supersaturação, na qual a proteína não é mais solúvel e começa a precipitar. Caso ocorra precipitação, ela pode ser de forma desordenada, formando precipitados amorfos ou mudanças para fase oleosa (não desejáveis), ou acontecer de forma sistemática pela formação de núcleos cristalinos que correspondem a proteínas em uma estrutura organizada. Dependendo da condição de cristalização e das particularidades da macromolécula, após a nucleação, mais moléculas de proteína podem agregar-se aos núcleos de forma ordenada fazendo com que o cristal cresça.

Os testes de cristalização das formas mutantes de NahI foram realizados com as proteínas clivadas e não clivadas na proporção 1:1 de proteína / solução de cristalização, a 18°C e em concentrações variadas do agente precipitante Formato de Sódio e diferentes valores de pH, por meio da adição de tampão *Hepes-Na* 0,1 M, conforme indicam os mapas das caixas de cristalização esquematizados na Figura 16.

Mapa 1						
Hepes-Na	Formato de Sódio					
	3,0 M	3,2 M	3,4 M	3,6 M	3,8 M	4,0 M
-						
pH 7,0						
pH 7,8						
pH 8,2						

Mapa 2						
Hepes-Na	Formato de Sódio					
	1,8 M	2,0 M	2,2 M	2,4 M	2,6 M	2,8 M
pH 7,2						
pH 7,8						

Figura 16. Mapas das condições de cristalização testadas para os mutantes de NahI. Os testes foram realizados com as proteínas (com e sem 6xHis) na proporção 1:1 de proteína / solução de cristalização, em concentrações variadas de Formato de Sódio e diferentes pHs.

Os cristais formados foram submetidos à difração de raios-X na linha MX2 do Laboratório Nacional de Luz Síncrotron (LNLS) em Campinas, SP. Para isto, foram coletados com *loop* de *nylon* adequado, transferidos brevemente para solução criogênica contendo

etilenoglicol na concentração final de 10% v/v, congelados em vapor de nitrogênio líquido na temperatura de 100 K e então expostos ao feixe de raios-X.

As imagens dos dados de difração, obtidas por meio de detector MARCCD, foram processadas no programa *HKL2000* (Otwinowski & Minor 1997).

3.13. RESOLUÇÃO DA ESTRUTURA CRISTALOGRÁFICA

Inicialmente, a fim de se estimar o número de moléculas que se espera encontrar na unidade assimétrica do cristal, foi realizado o cálculo do coeficiente de *Matthews* (Matthews 1968) dado por:

$$V_m = V/M \times x \times Z$$

em que V é o volume da célula unitária em $\text{\AA}^3$, M é a massa da proteína em Daltons, x é o número de unidades assimétricas por célula unitária, enquanto Z corresponde ao número de moléculas na unidade assimétrica.

A substituição molecular foi realizada pelo programa *Phaser* (McCoy et al 2007) e a construção automática do modelo foi obtida através do *Bucaneer* (Cowtan 2006). O refinamento das fases encontradas procedeu-se, inicialmente, pelo programa *Refmac5* (Murshudov et al 2011; Murshudov et al 1997) disponível no pacote CCP4 (Winn et al 2011). Nesse programa, foi utilizada a opção “*Restraigned Refinement*”, em que o modelo é refinado para melhor se ajustar à densidade experimental, mantendo assim a estereoquímica adequada.

Para inspeção e correção mais apurada dos parâmetros, como ângulos e comprimento de ligações, os ciclos finais de refinamento foram realizados pelo programa *Phenix* (Adams et al 2010) intercalados com ajustes manuais no modelo quando visualizado no programa *Coot* (Emsley et al 2010) juntamente com os mapas de densidade eletrônica $2F_o - F_c$ e $F_o - F_c$.

As figuras dos modelos tridimensionais foram feitas no programa *UCSF Chimera* (Pettersen et al 2004).

3.14. ESPALHAMENTO DE RAIOS-X A BAIXO ÂNGULO

Curvas de espalhamento de raios-X a baixo ângulo (SAXS, do inglês *small-angle x-ray scattering*) para amostras de NahI em tampão *Fosfato de Sódio 0,05 M pH 7,4, NaCl 0,05 M* foram registradas na linha de feixe SAXS-1 do Laboratório Nacional de Luz Síncrotron (LNLS) utilizando um detector *Dectris Pilatus 300k* (84 mm × 107 mm). As amostras foram centrifugadas (13.000 *g* por 10 minutos a 4 °C) antes da coleta de dados. Com o comprimento de onda $\lambda = 1,55 \text{ \AA}$ e a distância amostra-detector de 1617,34 mm, o intervalo do módulo do vetor de espalhamento foi de $0,0069 < q < 0,2692 \text{ \AA}^{-1}$ ($q=4\pi\sin\theta/\lambda$, em que 2θ é o ângulo de espalhamento). As curvas de espalhamento da solução de proteína e do tampão foram coletadas em etapas consecutivas com tempos de exposição de 5, 20 e 60 segundos para monitoramento dos danos causados pela radiação à amostra, além da estabilidade do feixe, sendo as amostras mantidas em um porta-amostras com 1,0 mm de espessura de janelas de mica à temperatura constante (25 °C).

A integração de dados, normalização da intensidade do feixe transmitido e atenuação da amostra, bem como subtração do espalhamento do tampão, foram efetuadas com Fit2D (Hammersley et al 1996). A análise dos dados e a plotagem foram realizadas com ATSAS 2.5.1.1 (Petoukhov et al 2012) e GNU PLOT (<http://www.gnuplot.info>). Os valores para o raio de giro (R_g) e a intensidade de espalhamento a ângulo zero ($I(0)$) foram obtidos tanto a partir da aproximação de *Guinier* (Porod 1982) $I(q) = I(0)\exp(-q^2R_g^2/3)$, válida para $qR_g \lesssim 1.3$, quanto pelo método de transformada de *Fourier* indireta implementado no programa GNOM (Svergun 1992). O último também foi utilizado para calcular a função de distribuição de pares de distância, $P(r)$, a partir da curva de espalhamento. O estado conformacional da proteína em solução foi avaliado pelo gráfico de *Kratky* ($q^2I(q) \times q$) obtido diretamente a partir da curva de espalhamento (Doniach 2001; Rambo & Tainer 2011; Semisotnov et al 1996). Estimativas para a massa molecular das partículas espalhadoras em solução foram obtidas através de um método que é independente de medições da concentração de proteína, com base no invariante de *Porod*, $\int_0^\infty q^2I(q)dq$ (integração da área sob o gráfico de *Kratky*) (Fischer et al 2010).

Simulações de espalhamento com estruturas do banco de dados de proteínas PDB (Berman et al 2000) foram realizadas utilizando o programa CRY SOL (Svergun et al 1995), de forma que os dados experimentais obtidos para NahI puderam ser comparados com estruturas de alta resolução conhecidas. Neste caso, a distância máxima observada em cada

estrutura foi calculada com o programa NCONT (Winn et al 2011). Após determinação do arranjo tetramérico mais provável de NahI (análogo ao PDB id 4I1W), um modelo tridimensional do monômero de NahI foi predito utilizando a sequência de aminoácidos, conforme implementado na plataforma I-TASSER (Roy et al 2010). Feito isto, a estrutura do monômero foi então individualmente posicionada sobre cada uma das quatro cadeias que compõem a unidade biológica, resultando no tetrâmero putativo de NahI. O programa FoXS (Schneidman-Duhovny et al 2010; Schneidman-Duhovny et al 2013) foi utilizado para calcular as curvas de espalhamento teóricas a partir dos modelos de dímero (similar às cadeias A e B de 4I1W) e tetrâmero (similar a 4I1W) de NahI e para estimar a porcentagem de cada oligômero em solução.

3.15. ENSAIO DE ATIVIDADE ENZIMÁTICA

A atividade de 2-hidroximuconato semialdeído desidrogenase foi avaliada seguindo a taxa de desaparecimento do substrato 2-HMS a 375 nm ($\lambda_{\text{máx}} = 22.966 \text{ M}^{-1} \text{ cm}^{-1}$). A produção de NADH a 340 nm não pôde ser observada devido ao substrato absorver fortemente na mesma região espectral. Ao contrário, para a atividade cinética de NahI sobre 2-salicilaldeído, outro intermediário da via de degradação do naftaleno, monitorou-se a taxa de formação de NADH a 340 nm.

A mistura de reação consistiu em tampão *Fosfato de Potássio 50 mM pH 8,5*, *NaCl 50 mM*, contendo diferentes concentrações de 2-hidroximuconato semialdeído, 200 μM de NAD^+ e 0,060 μM de enzima. Tanto o substrato como o cofator foram dissolvidos em tampão *Fosfato de Potássio 50 mM pH 8,5* antes de serem adicionados à mistura de reação. O ensaio foi iniciado pela adição de 2 a 40 μL de 2-hidroximuconato semialdeído a um volume final de 1,0 mL. Todos os ensaios espectrofotométricos foram realizados em cubetas de quartzo de 10 mm de caminho óptico a 25 °C, e os dados foram obtidos utilizando um espectrofotômetro Agilent 8453.

Os parâmetros cinéticos foram ajustados por regressão não linear das velocidades iniciais (V_0) versus concentrações do substrato 2-hidroximuconato semialdeído [S] no programa GraFit (Erathicus Software Ltd., Staines, UK), seguindo a equação de *Michaelis-Menten*,

$$V_0 = \frac{V_{\text{máx}} [S]}{K_M + [S]}$$

na qual se define a relação quantitativa entre a velocidade inicial, V_0 , a velocidade máxima, $V_{\text{máx}}$, e a concentração inicial de substrato $[S]$, todas relacionadas pela constante de *Michaelis*, K_M . Neste trabalho, especificamente, o termo K_M será usado como um indicador da afinidade da enzima pelo seu substrato.

A grandeza $V_{\text{máx}}$ equivale à etapa limitante da velocidade da reação enzimática e varia muito entre enzimas diferentes. Uma constante de velocidade mais geral, k_{cat} , é útil para descrever a etapa limitante de qualquer reação catalisada por enzimas na saturação. A constante k_{cat} , cuja unidade é a recíproca do tempo, também é chamada de número de renovação (“*turnover*”), sendo equivalente ao número de moléculas de substrato convertidas em produto, por unidade de tempo, por uma única molécula de enzima quando esta estiver saturada pelo substrato. Assim, calcula-se:

$$k_{\text{cat}} = V_{\text{máx}}/[E_{\text{total}}]$$

em que $V_{\text{máx}}$ é a velocidade máxima calculada para a reação enzimática, enquanto E_{total} equivale à concentração total de enzima utilizada no ensaio.

Para avaliar as eficiências catalíticas de enzimas diferentes, ou o número de vezes que diferentes substratos são catalisados por uma mesma enzima, compara-se a relação k_{cat}/K_M para as duas reações. Este parâmetro, muitas vezes denominado constante de especificidade, indica qual é o substrato “natural” ou preferido pela enzima. Assim, nossa discussão sobre a atividade de NahI e suas formas mutantes, bem como de outras ALDHs, sobre diferentes substratos aldeídos será feita por comparação dos valores de k_{cat}/K_M .

O substrato 2-HMS foi produzido e gentilmente fornecido pelo professor Christian P. Whitman da Universidade do Texas em Austin, EUA, conforme descrito em (Araujo et al 2015).

4. RESULTADOS E DISCUSSÃO

Em trabalho anterior iniciado por Neves e colaboradores (Neves 2013) do Grupo de Biologia Estrutural (BioEst / ICB / UFMG), a forma recombinante da enzima NahI foi expressa a 12 °C em *E. coli* BL21 ArcticExpress (DE3) por 24h após indução com IPTG. Com este protocolo de expressão, NahI apresentou-se na fração solúvel conforme observado pela banda com peso molecular aproximado de 55 kDa em gel de poliacrilamida-SDS (Figura 17), consistente com o monômero de 6xHis-NahI (53,518 Da).

Duas etapas de purificação consecutivas, cromatografia de afinidade ao níquel e de exclusão molecular, possibilitaram recuperar, com elevado grau de pureza, aproximadamente 30 mg da proteína 6xHis-NahI para cada litro de meio de cultura (Figura 17). Após remoção da cauda de histidinas com a protease TEV, dois resíduos de aminoácidos adicionais, glicina e histidina, permaneceram na proteína, que passou a apresentar peso molecular teórico de 51,712 Da sem qualquer sinal de agregação ou precipitação.

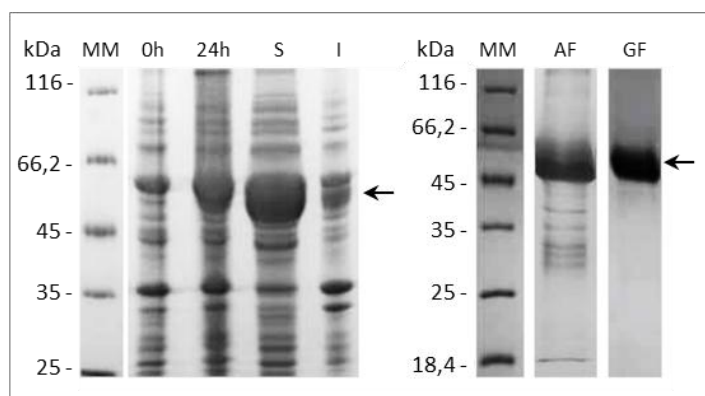


Figura 17: Expressão heteróloga e purificação de 6xHis-NahI. Perfil eletroforético de proteínas totais de bactérias *E. coli* BL-21 ArcticExpress (DE3) transformadas com o vetor de expressão pET28a-TEV-*nahI*, antes (0h) e após (24h) indução da expressão por IPTG. S e I: frações solúvel e insolúvel após a lise celular. AF: Cromatografia de afinidade. GF: Cromatografia de gel filtração ou exclusão molecular. As bandas referentes à NahI recombinante estão indicadas pelas setas. Géis de poliacrilamida-SDS 15% corados com azul de Coomassie. MM (kDa): massa molecular em *kiloDaltons* (*Unstained Protein Molecular Weight Marker* – *Thermo Scientific*). Extraído e adaptado de (Araujo et al 2015).

4.1. ESTRUTURA SECUNDÁRIA DE NahI NA AUSÊNCIA E NA PRESENÇA DE NAD⁺

Os espectros de CD de NahI a 20 °C, na presença e na ausência de NAD⁺, são muito semelhantes um do outro ao longo de toda a gama de comprimento de onda medido (190 a 260 nm) (Figura 18 a), com apenas uma ligeira diferença em comprimentos de onda inferiores a 195 nm. Esta diferença pode ser devida tanto ao maior nível de ruído nos espectros do tampão e da amostra nesta região, como também por uma diferença muito pequena no teor de fitas- β entre as amostras. A deconvolução de qualquer destes dois espectros utilizando o programa CDNN 2.0 (Bohm et al 1992) é insensível a esta pequena diferença, sendo que ambos sugerem uma composição para NahI de 35% de α -hélices, 17% de fitas- β e 46% de outros elementos de estrutura secundária (17% de voltas- β e 29% de *random coil*).

Na prática, estes resultados sugerem que a ligação de NAD⁺ não provoca alterações conformacionais significativas na estrutura secundária da enzima a 20 °C. Além disso, estes dados aproximam-se do conteúdo de estrutura secundária de 2-aminomuconato 6-semialdeído (2-AMS) desidrogenase de *Pseudomonas fluorescens* (PDB id 4I1W) (Huo et al 2015), que compartilha 61% de identidade de sequência com NahI e apresenta 43% de α -hélices e 22% de fitas- β , conforme atribuído pelo servidor STRIDE (Frishman & Argos 1995; Heinig & Frishman 2004). É de se esperar que NahI também apresente um enovelamento semelhante ao de NahF (PDB id 4JZ6) (Coitinho et al 2012), uma ALDH aromática de *P. putida* G7 pertencente à mesma via de degradação do naftaleno. Ambas compartilham apenas 33% de identidade de sequência, contudo NahF é composta de 43% de α -hélices e 20% de fitas- β .

Na medida em que a temperatura é elevada até 95 °C, os espectros de CD são modificados e direcionam-se para um perfil de proteínas parcialmente desenoveladas, resultante da perda de conteúdo de α -hélices (Figura 18 a). Um espectro de CD típico de α -hélices possui contribuições positivas em torno de 194 nm e negativas em torno de 210 e 220 nm.

Curiosamente, as formas apo e holo (NAD⁺ ligado) de NahI se comportam de forma diferente quando submetidas à desnaturação térmica. Esta diferença torna-se mais evidente a partir de 55-60 °C na faixa de 190 a 195 nm, mas se estende a aproximadamente 210 nm ao se atingir 95 °C, como mostrado na Figura 18 a. Estes resultados sugerem que a ligação de NAD⁺ aumenta a estabilidade do complexo devido à manutenção de fitas- β do domínio de ligação ao cofator, mais provavelmente do motivo de *Rossmann*, durante a desnaturação

térmica. Como fitas- β têm uma contribuição positiva sobre a elipticidade molar residual ($[\theta]_{MR}$) na faixa de 195 a 210 nm, a curva de CD de NahI-NAD⁺ é menos negativa do que aquela observada para a enzima sem o cofator.

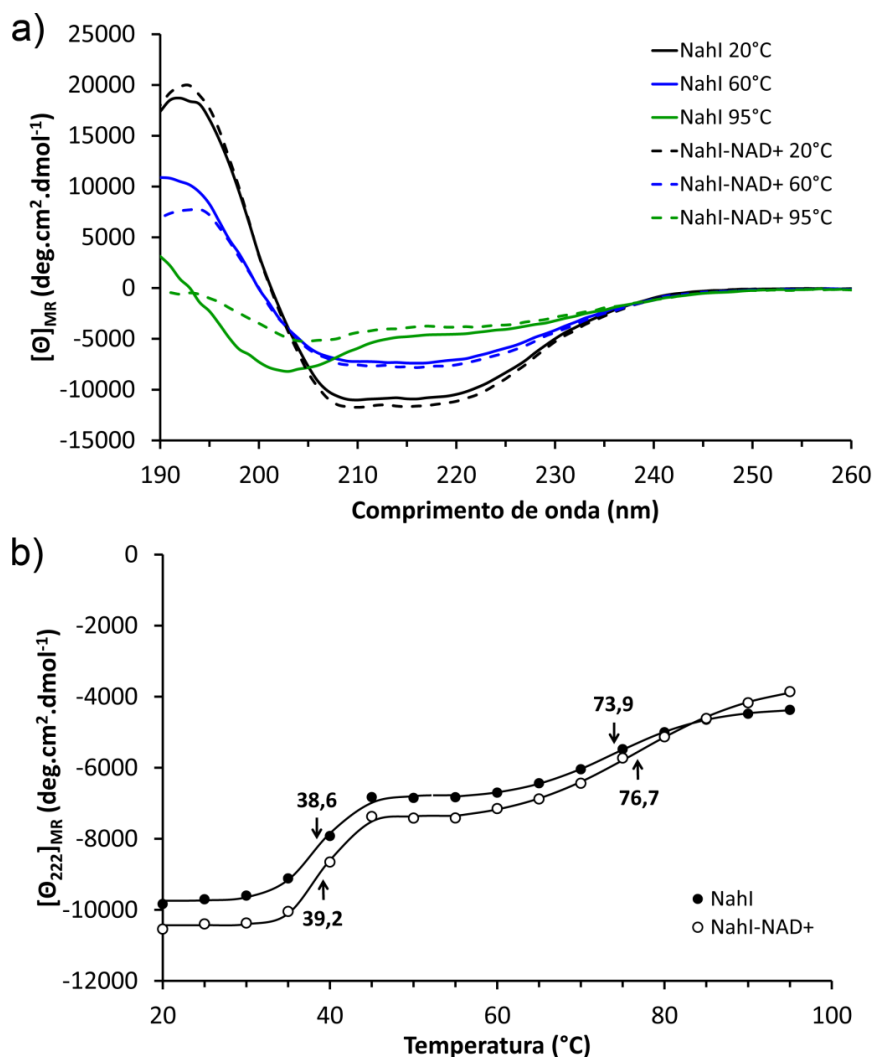


Figura 18. Análises de estrutura secundária e termoestabilidade de NahI na presença e ausência de NAD⁺. **a)** Elipticidade molar ($[\theta]_{MR}$) versus comprimento de onda (190 a 260 nm) nas temperaturas de 20, 60 e 95 °C. **b)** Monitoramento da elipticidade molar residual no comprimento de onda fixo de 222 nm ($[\theta]_{222MR}$) em função da temperatura. A enzima NahI apresenta um processo de desnaturação em duas etapas, contudo a ligação de NAD⁺ parece ter maior impacto apenas sobre a segunda temperatura de transição.

Além disso, quando a elipticidade molar residual é monitorada no comprimento de onda fixo de 222 nm ($[\theta]_{222MR}$) em função da temperatura (Figura 18 b), é possível sugerir que NahI segue um processo de desnaturação em duas etapas, sendo a primeira temperatura

de transição de 38,6 °C e a segunda de 73,9 °C. Quando complexada ao NAD^+ , a enzima também apresenta um processo de desnaturação em duas etapas, contudo a ligação de NAD^+ parece ter maior impacto apenas sobre a segunda temperatura de transição (76,7 °C) (Figura 18 b).

Um aumento de 2 a 3 °C na temperatura de transição de proteínas frente à presença de um ligante é consistente com valores observados por fluorimetria de varredura diferencial (do inglês *Differential Scanning Fluorimetry*, DSF). Neste método, a temperatura à qual se desenovela uma proteína é medida pelo aumento na fluorescência de um corante com afinidade para porções hidrofóbicas expostas durante a desnaturação proteica. A intensidade da fluorescência é dada em função da temperatura e apresenta uma curva sigmoideal, descrita por uma transição entre dois estados, sendo o ponto de inflexão (T_m) calculado por equações simples, como a de *Boltzmann*. Dessa forma, calcula-se, portanto, a temperatura de transição na presença e na ausência de um determinado ligante (Niesen et al 2007).

Segundo Coitinho (2013), a enzima NahF, quando avaliada por DSF, apresenta a temperatura de desenovelamento de 58,5 °C. Na presença de NAD^+ , a temperatura de transição se eleva para 61 °C. Quando comparado ao substrato salicilaldeído, o NAD^+ é capaz de melhor estabilizar a enzima, efeito que pode ser devido ao maior número de interações estabelecidas pelo cofator (Coitinho 2013).

Novamente, o comportamento observado sustenta a ideia de que o domínio de ligação ao cofator é, entre os outros, menos suscetível à desnaturação térmica.

4.2. ESTADO OLIGOMÉRICO DE NahI EM SOLUÇÃO

4.2.1. Por cromatografia de exclusão molecular e espalhamento dinâmico de luz

Ensaio de cromatografia de exclusão molecular e espalhamento dinâmico de luz foram conduzidos para avaliar indiretamente o estado oligomérico de NahI em solução.

Após a obtenção de um lote único de NahI, foram preparadas amostras em três diferentes concentrações (0,5 , 1,5 e 6 mg.mL^{-1}), a partir das quais separaram-se duas alíquotas. A uma delas adicionou-se NAD^+ para uma concentração molar final 10 vezes superior à concentração de proteína. Após 1 hora de incubação em gelo, cada alíquota, com e

sem NAD^+ , foi individualmente conduzida aos ensaios de cromatografia de exclusão molecular e espalhamento dinâmico de luz.

Os perfis cromatográficos de gel filtração obtidos para as alíquotas de proteína, com e sem NAD^+ , revelaram um único pico bem resolvido, porém com volumes de retenção diferentes. NahI e NahI- NAD^+ eluíram entre 76,7 e 79,3 mL, volumes consistentes com massas moleculares aparentes de 184,3 a 135 kDa (Figura 19). Tais observações não sugerem com confiabilidade qual é o estado oligomérico de NahI, que pode estar sendo formada por três (155 kDa) ou quatro monômeros (207 kDa), uma vez que sua massa molecular monomérica é de 51,7 kDa. Por outro lado, o estado oligomérico de NahI parece não ser influenciado pela interação com o cofator NAD^+ .

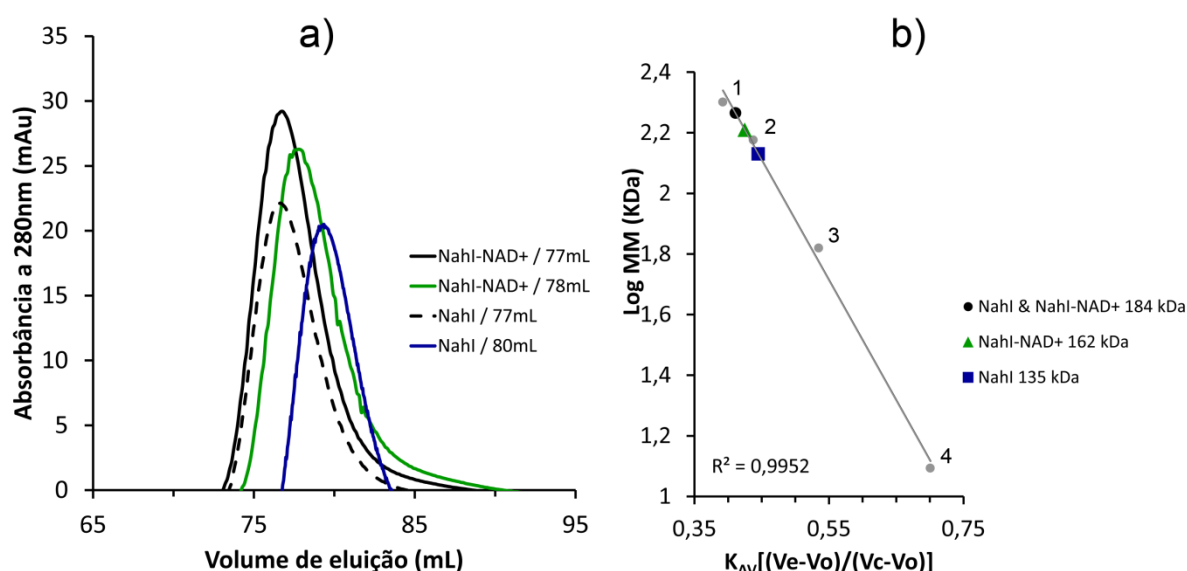


Figura 19. Estado oligomérico de NahI avaliado por cromatografia de exclusão molecular. a) Perfis cromatográficos obtidos para NahI na ausência (linhas cheias azul e pontilhada preta) e na presença de NAD^+ (linhas cheias verde e preta). O volume de eluição do NAD^+ em excesso na solução corresponde a 110 mL (não mostrado). A absorbância a 280 nm (mAu) é dada em função do volume de eluição (mL) da coluna. b) Curva de calibração de log MM versus K_{AV} . K_{AV} , ou coeficiente de partição, é dado pela fórmula $[(Ve-Vo)/(Vc-Vo)]$, em que Ve corresponde ao volume de eluição da proteína, enquanto Vc e Vo são, por esta ordem, o volume total (122 mL) e o volume vazio (45 mL) da coluna. O volume vazio foi calculado por meio da aplicação de *Blue Dextran*, enquanto a calibração da coluna foi realizada aplicando-se individualmente as seguintes proteínas estabelecidas como padrões de massa molecular: 1 - β -amilase (200 kDa), 2 - álcool desidrogenase (150 kDa), 3 - albumina (66 kDa) e 4 - citocromo c (12,4 kDa).

Quanto aos resultados obtidos por espalhamento dinâmico de luz, sugere-se que NahI, quer na presença ou na ausência de NAD^+ , apresenta-se como um trímero. NahI e NahI- NAD^+

comportam-se em solução, respectivamente, como uma partícula com diâmetros hidrodinâmicos (D_h) de 9,8 e 9,9 nm, os quais para proteínas globulares podem corresponder na devida ordem a massas moleculares de 139 e 142 kDa (Tabela 2).

Tabela 2: Espalhamento dinâmico de luz de NahI na ausência e presença de NAD⁺.

	D_h (nm)	Polidispersidade (%) [*]	MM (kDa)	Intensidade (%)	Massa (%)
NahI	9,8	12,6	139 ± 18	100	100
NahI-NAD⁺	9,9	15,7	142 ± 22	100	100

^{*}A técnica de DLS é, sobretudo, utilizada para avaliar o estado de agregação de uma amostra e medir a polidispersidade, que é preditivo de cristalizabilidade (Ferré-D'amaré & Burley 1994; 1997). Índices de polidispersidade inferiores a 20% para NahI indicam que a enzima se comporta como uma partícula monodispersa em solução, suscetível de ser cristalizada.

Ainda que estes resultados proponham o oligômero de NahI como um trímero em solução, há poucas evidências na literatura de que aldeído desidrogenases possam se comportar de tal forma. Até então, ALDHs cujas estruturas tridimensionais são conhecidas organizam-se como um dímero, com os domínios de oligomerização entrelaçados, como um tetrâmero, em que duas cópias de dímeros se abraçam, e até como um hexâmero, caracterizado como um trímero de dímeros (Liu et al 1997; Luo et al 2013; Pemberton et al 2014; Steinmetz et al 1997).

Assim, os valores teóricos esperados para o dímero e o tetrâmero de NahI seriam, respectivamente, de 103,4 e 206,8 kDa, com base na massa molecular monomérica de 51,7 kDa.

Curiosamente, situação semelhante à de NahI foi reportada por He e colaboradores (He et al 1998) em que resultados de exclusão molecular referentes a enzima 2-aminomuconato 6-semialdeído (2-AMS) desidrogenase de *Pseudomonas pseudoalcaligenes* sugerem o estabelecimento de um trímero. Todavia, os autores concluem com ressalvas, pois, segundo os mesmos, mais investigação é necessária para rigorosamente determinar a estrutura nativa da enzima, tendo em vista ser incomum uma estrutura homotrimérica para ALDHs (He et al 1998). Resolvida por cristalografia de raios-X a 2,0 Å de resolução, a estrutura tridimensional de 2-AMS desidrogenase de *Pseudomonas fluorescens* (PDB id 4I1W), já

mencionada na subseção 4.1, revela ser um homotetrâmero a sua unidade biológica (Huo et al 2015).

Cabe ressaltar que a técnica de DLS utiliza as flutuações de intensidade da luz espalhada, induzidas pelo movimento *browniano* entre soluto e solvente, para inferir o tamanho de partículas em solução. Estas flutuações ocorrem em escalas de tempo relacionadas com a velocidade de movimento das moléculas, e, portanto, dependem de seu tamanho e forma (Matte & Cygler 2007).

A taxa das flutuações de intensidade detectadas é utilizada para calcular o coeficiente de difusão translacional que, por sua vez, é convertido em um raio hidrodinâmico da partícula através da equação de *Stoke-Einstein*,

$$D = \frac{\kappa_B T}{6\pi\eta R_h}$$

em que D é a constante de difusão, κ_B (*kappa*) é a constante de *Boltzmann*, T é a temperatura absoluta, η (*eta*) é a viscosidade e R_h é o raio hidrodinâmico da partícula (Harding 1994; Lorber et al 2012).

Por definição, o raio “medido” no DLS é o raio de uma esfera hipotética que se difunde na mesma velocidade que a partícula examinada. Na prática, macromoléculas em solução são não esféricas, dinâmicas e solvatadas. Assim, o raio medido no DLS é apenas um indicativo do tamanho aparente da partícula hidratada e deve, por assim dizer, ser considerado com cautela (Harding 1994; Lorber et al 2012).

Portanto, outra tentativa de identificar corretamente o estado oligomérico e a estrutura quaternária de NahI em solução foi proposta por meio de um experimento de espalhamento de raios-X a baixo ângulo (SAXS).

4.2.2. Por espalhamento de raios-X a baixo ângulo

Para verificar uma possível dependência do estado oligomérico com a concentração de proteína, amostras de NahI analisadas em três concentrações diferentes (1,5 , 5,8 e 11,3 mg.mL⁻¹) resultaram em estimativas de massa igual a 185, 190 e 177 kDa, respectivamente. Estes dados indicam inequivocamente que NahI comporta-se como um tetrâmero em solução, tendo em vista o peso molecular previsto igual a 206,8 kDa para o oligômero. De certa forma, concordam também com um dos valores de massa molecular sugeridos por cromatografia de exclusão molecular (184,3 kDa) reportado na subseção anterior (4.2.1).

Em todos os casos, danos da radiação foram visíveis apenas para os tempos de exposição mais longos e cumulativos. Assim, a primeira curva de espalhamento recolhido com 20 segundos de tempo de exposição a partir da amostra a 5,8 mg.mL⁻¹ (Figura 20 A) foi escolhida para posterior análise devido à sua maior proporção de sinal-ruído e confiabilidade observados para o procedimento de determinação da massa.

O claro comportamento linear da curva de *Guinier* (inserto da Figura 20 A) sugere fortemente uma amostra de proteína monodispersa e permite estimar com confiança o raio de giro (R_g) igual a 38,2 Å. A linearidade da curva de *Guinier* é um indicativo de monodispersidade satisfatória para a amostra de proteína, uma vez que desvios podem sinalizar agregação inespecífica ou interferência interpartículas (Jacques et al 2012). Além disso, este valor está em excelente concordância com aquele obtido pelo GNOM (Svergun 1992), $R_g = 37,4$ Å, durante o cálculo da função de distribuição de distâncias $P(r)$ (Figura 20 B), que já antecipa um conjunto globular em solução com dimensão máxima ($D_{máx}$) igual a 106 Å.

A curva $P(r)$ fornece a distribuição provável dos pares de distâncias entre os centros de espalhamento (átomos) da proteína, além de evidências da qualidade dos dados pelo modo que o perfil se aproxima de zero em $r = 0$ e $r = D_{máx}$, a máxima dimensão linear da partícula espalhadora. Em ambos os limites, a curva deve ser lisa e côncava, quando acima do eixo r (Jacques et al 2012).

Por fim, para completar esta análise preliminar, o gráfico de *Kratky*, apresentado na Figura 20 C com um pico claramente definido e um decaimento em ângulos superiores, típico para domínios bem enovelados, indica que a proteína possui uma estrutura compacta em solução.

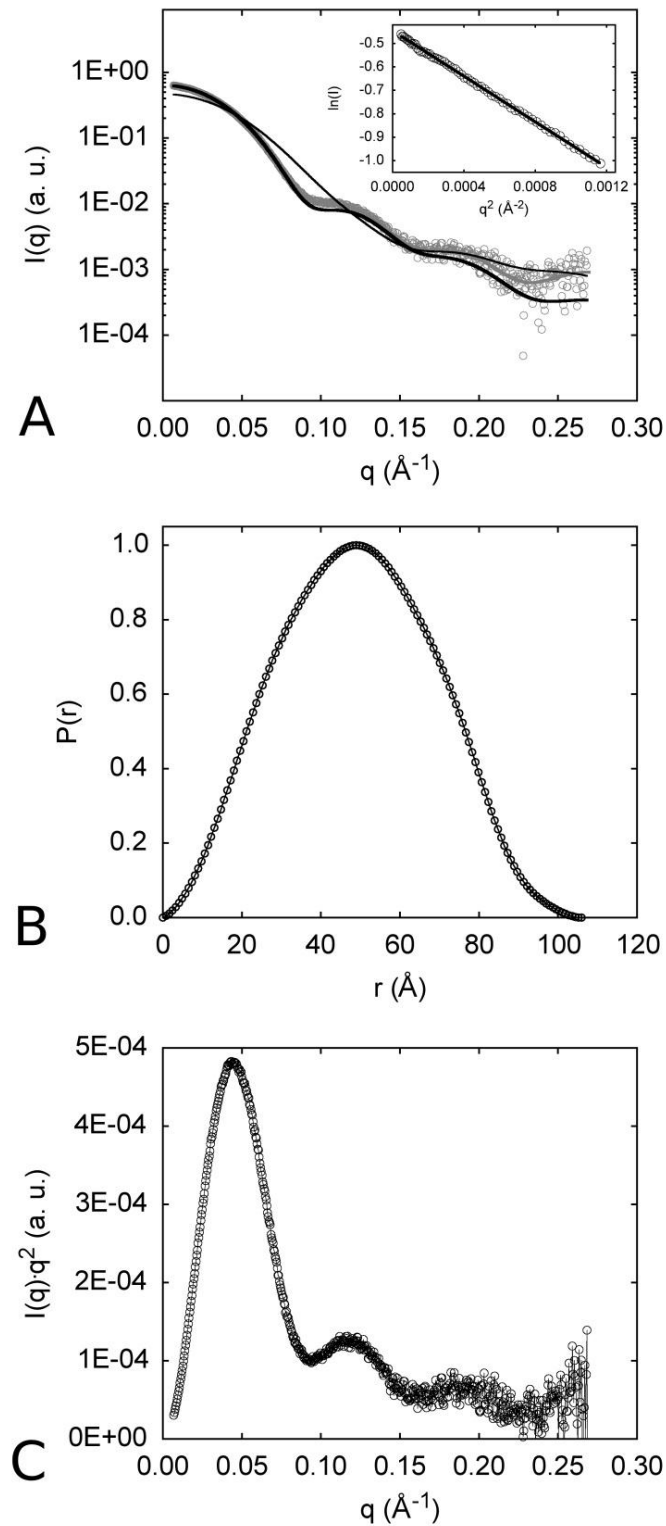


Figura 20. Dados de SAXS e parâmetros globais. **A.** Curva experimental de espalhamento de raios-X (círculos abertos) de NahI em solução, sendo o ajuste (linha sólida em cinza) obtido durante o cálculo da função de distribuição de pares de distâncias, $P(r)$. Gráfico de *Guinier* e ajuste linear são mostrados no inserto. Curvas teóricas de SAXS para dímero (linha fina em preto) e tetrâmero (linha grossa em preto) de NahI calculadas com o programa FoXS. **B.** Função de distribuição de pares de distâncias derivada da curva experimental. O perfil em forma de sino, com um máximo centrado, é típico de um espalhamento de partículas globulares em solução. **C.** Gráfico de *Kratky* calculado a partir dos dados experimentais, típico de proteínas compactas e com domínios bem enovelados.

Para verificar a presença de simetria no oligômero de NahI, possíveis arranjos tetraméricos foram primeiramente investigados utilizando CRY SOL (Svergun et al 1995). Curvas de espalhamento teóricas para os 12 homólogos mais próximos, cujas estruturas estão disponíveis no PDB (Berman et al 2000), foram ajustados aos dados de SAXS experimentais medidos para NahI. Embora estas estruturas tenham sido determinadas por cristalografia de raios-X, é importante ter em mente que o conteúdo da unidade assimétrica do cristal não corresponde necessariamente ao estado oligomérico da proteína em solução. No caso de NahI, entretanto, o tetrâmero contido na unidade assimétrica dos homólogos mais próximos exibiu um bom ajuste em relação aos dados experimentais. Além disso, o R_g calculado pelo CRY SOL (Svergun et al 1995) e a dimensão máxima obtida com NCONT (Winn et al 2011) para as estruturas homólogas mais próximas apresentaram excelente concordância com os valores experimentais correspondentes para NahI.

Outra confirmação do estado oligomérico e da monodispersidade da amostra foi obtida a partir de um cálculo do conjunto mínimo de modelos moleculares necessários para melhor atender aos dados experimentais (Figura 20 A), cuja pesquisa foi realizada utilizando o programa FoxS (Schneidman-Duhovny et al 2010; Schneidman-Duhovny et al 2013) e supondo a existência de uma combinação de dímeros e tetrâmeros de NahI em solução. Uma vez que a porcentagem de tetrâmeros em solução foi estimada em 99%, esta análise indica ao todo que NahI está presente como um tetrâmero, assumindo um arranjo compatível com o tetrâmero da homóloga 2-aminomuconato 6-semialdeído desidrogenase de *Pseudomonas fluorescens* (PDB id 4I1W).

4.3. RESOLUÇÃO DA ESTRUTURA CRISTALOGRÁFICA DE NahI

A forma recombinante da enzima NahI foi cristalizada por Neves (Neves 2013) em condições variadas, dentre as quais a mais promissora para o crescimento de cristais adequados à difração continha Formato de Sódio como agente precipitante (Figura 21 A-E). Dados de difração foram coletados nas linhas de luz MX1 e MX2 do Laboratório Nacional de Luz Síncrotron em Campinas - SP a partir de cristais previamente imersos, ou não, em solução contendo o cofator NAD^+ (Figura 21 F) (Neves 2013).

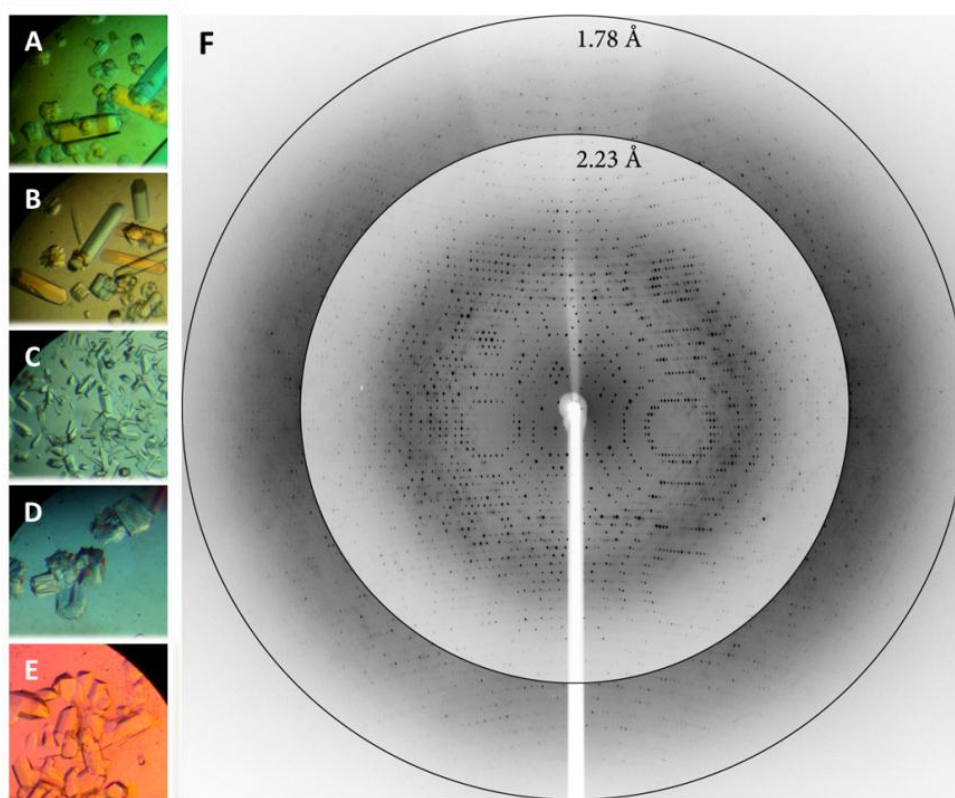


Figura 21: Cristais de 6xHis-NahI, uma 2-hidroxiuconato semialdeído desidrogenase de *Pseudomonas putida* G7, crescidos em variadas condições de cristalização. (A) formato de sódio 3,8 M, Hepes 0,1 M pH 7,0; (B) formato de sódio 2,8 M; (C) fosfato de sódio e potássio 1M pH 7,0; (D) BIS-TRIS propano 0,1 M pH 9,0 e acetato de sódio 2,5 M; (E) fosfato de amônio dibásico 1,1 M, cloreto de sódio 0,2 M, citrato de sódio 0,1 M pH 5,6. (F) Padrão de difração de raios-X de um dos cristais de 6xHis-NahI evidenciando duas faixas de resoluções diferentes (Neves 2013).

Cristais de 6xHis-NahI crescidos em condições de formato de sódio pertencem ao grupo espacial $P6_422$ e apresentam célula unitária com parâmetros $a = b = 189,47$ e $c = 79,28 \text{ \AA}$. As estatísticas dos dados de difração de dois cristais, com e sem NAD^+ , são

apresentadas na Tabela 3 (Neves 2013). A resolução da estrutura tridimensional de NahI faz parte do escopo desta tese.

Tabela 3: Estatísticas dos dados de difração dos cristais de NahI e NahI-NAD⁺. Os valores estatísticos para a faixa de maior resolução são dados entre parênteses (Neves 2013).

	NahI	NahI-NAD ⁺
Comprimento de onda (Å)	1,459	1,459
Distância cristal-detector (mm)	110	125
Rotação por imagem (°)	0,3	0,3
Variação total de rotação (°)	120	105,3
Grupo espacial	P 6 ₄ 2 2	P 6 ₄ 2 2
Parâmetros da célula unitária (Å, °)	a = b = 189,47; c = 79,28 $\alpha = \beta = 90; \gamma = 120$	a = b = 189,13; c = 79,56 $\alpha = \beta = 90; \gamma = 120$
Faixa de resolução (Å)	50,0 – 1,85 (1,89 – 1,85)	50,0 – 2,15 (2,21 – 2,15)
Número de observações	1.1018.744	509.450
Número de reflexões únicas	68.981 (4.402)	45.797 (3.723)
Completeza (%)	96,4 (94,0)	99,8 (99,4)
<I/σ (I)>	42,3 (6,7)	29,4 (4,1)
Redundância	14,8 (14,8)	11,1 (7,6)
$R_{merge}^{\dagger}$ (%)	5,8 (45,0)	7,6 (48,3)
Mosaicidade média (°)	0,262	0,561

[†] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, em que $I_i(hkl)$ é a intensidade medida de cada reflexão hkl e $\langle I(hkl) \rangle$ é sua média.

A determinação da estrutura cristalográfica de 6xHis-NahI foi realizada pela técnica de substituição molecular tendo como modelo a cadeia A de NahF identificada no PDB por 4JZ6 (Coitinho 2013), que possui 33% de identidade e 52% de similaridade de sequência com NahI. A ideia foi utilizar as fases calculadas para NahF como uma estimativa inicial para as fases dos fatores de estrutura de NahI.

Simplificadamente, na substituição molecular proposta, o modelo da cadeia A de NahF é posicionado na célula unitária obtida para NahI seguindo uma função de rotação e outra de translação. Dessa forma, calculam-se os fatores de estrutura de NahF, já corretamente posicionada, de modo a se obter fases iniciais para NahI. Estas fases emprestadas de NahF são então utilizadas em conjunto com as intensidades medidas experimentalmente a fim de se obter os fatores de estrutura de NahI que permitam o cálculo da densidade eletrônica inicial.

Um dos primeiros passos na análise de uma estrutura cristalina é estimar o número de moléculas de proteína contidas na unidade assimétrica, que depende, por sua vez, do conteúdo de solvente, da densidade do cristal, do grupo espacial e da massa molecular da proteína.

O cálculo do coeficiente de *Matthews* (V_M) (Matthews 1968) utilizando o peso molecular de 53.518 Da para 6xHis-NahI resultou nas possibilidades apresentadas na Tabela 4.

Tabela 4: Número de moléculas de proteína presentes na unidade assimétrica, coeficiente de Matthews e conteúdo de solvente calculados para cristais de NahI e NahI-NAD⁺.

Número de moléculas	Probabilidade total	Coeficiente de <i>Matthews</i> - V_M ($\text{\AA}^3 \text{Da}^{-1}$)	Conteúdo de solvente (%)	MM (Da)
1	0,06	3,84	68,0	53.518
2	0,94	1,92	35,9	107.036

Sugere-se, com probabilidade de 94% para os conjuntos ‘NahI’ e ‘NahI-NAD⁺’, duas moléculas de proteína na unidade assimétrica ($V_M = 1,92 \text{ \AA}^3 \text{Da}^{-1}$) e conteúdo de solvente de 35,9%. Entretanto, a substituição molecular em busca de duas moléculas por unidade assimétrica não teve êxito. Uma boa solução somente foi possível considerando-se uma única molécula na unidade assimétrica ($V_M = 3,84 \text{ \AA}^3 \text{Da}^{-1}$) e teor de solvente de 68,0%.

Para realizar a substituição molecular pelo programa *Phaser* (McCoy et al 2007), o grupo espacial e o número de moléculas por unidade assimétrica devem ser inicialmente fornecidos. Feito isto, é possível ao programa efetuar operações de rotação e translação com o modelo (PDB id 4JZ6) a fim de orientá-lo e posicioná-lo corretamente na unidade assimétrica correspondente ao cristal de NahI. A partir daí, a densidade eletrônica pode ser calculada usando as fases deste modelo corretamente posicionado e as intensidades experimentais do

padrão de difração. No caso de NahI, fez-se a procura utilizando o grupo espacial $P6_422$ e primeiramente duas, mas depois somente uma molécula na unidade assimétrica.

Prosseguiu-se aos ciclos iniciais de refinamento do modelo de NahI utilizando o programa *Refmac5* (Murshudov et al 2011; Murshudov et al 1997), em que as posições dos átomos são ajustadas à densidade eletrônica com o consequente melhoramento das fases. Para inspeção e correção mais apurada dos parâmetros, como geometria das ligações químicas dos aminoácidos e rotâmeros permitidos, os ciclos finais de refinamento foram realizados pelo programa *Phenix* (Adams et al 2010).

O refinamento dos modelos de NahI e NahI-NAD⁺ a 1,85 e 2,15 Å de resolução, respectivamente, foi acompanhado pelo decréscimo significativo nos valores de *R-factor* e *R_{free}*, que indicaram êxito na resolução das estruturas cristalográficas e encontram-se listados na Tabela 5, bem como as estatísticas de refinamento e de validação dos modelos.

A concordância global do modelo da unidade assimétrica com os dados experimentais é medida pelo *R-factor*,

$$R = \Sigma ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \Sigma |F_{obs}(hkl)|$$

em que $F_{obs}(hkl)$ são as amplitudes observadas e $F_{calc}(hkl)$ corresponde às amplitudes calculadas a partir do modelo. Os valores de *R* variam de zero, em uma perfeita convergência entre amplitudes observadas e calculadas, para cerca de 0,60, quando um conjunto de amplitudes observadas é comparado com um conjunto de amplitudes aleatórias. Neste último caso, valores de *R-factor* superiores a 0,50 sugerem uma baixa concordância entre o modelo tridimensional proposto e aquele de fato presente no cristal em estudo. Um modelo inicial com *R-factor* próximo de 0,40 é promissor e tende a melhorar ao longo das etapas de refinamento (Rhodes 2000). Assim, *R-factor* é um parâmetro utilizado para avaliar o melhoramento das fases na determinação da estrutura final, sendo 0,20 um valor desejável para um modelo de proteína já refinado e com dados a 2,5 Å. Para os modelos finais de NahI e NahI-NAD⁺, os valores de *R-factor* obtidos foram 0,18 e 0,19, respectivamente.

Tabela 5: Estatísticas de refinamento e de validação dos modelos cristalográficos de NahI e NahI-NAD⁺.

	NahI	NahI-NAD ⁺
Limites de resolução (Å)	50,0 – 1,85 (1,89 – 1,85)	50,0 – 2,15 (2,21 – 2,15)
R -factor [†] (%)	18,4	19,2
R_{free} [†] (%)	21,5	22,4
R_{iso} ^{††} (%)	-	15,8
Fatores de temperatura médios (Å ²):		
Macromolécula	29,76	45,91
Solvente	40,68	46,77
$R.m.s.D.$ para o comprimento de ligação (Å)	0,015	0,012
$R.m.s.D.$ para o ângulo de ligação (°)	1,506	1,423
Diagrama de <i>Ramachandran</i> (resíduos em regiões):		
Favoráveis (%)	96,9	96,3
Permitidas (%)	3,1	3,5
Proibidas (%)	0,0	0,2

[†] R -factor = $\sum ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \sum |F_{obs}(hkl)|$. R_{free} é o valor de R -factor calculado para 5% do conjunto de dados não inclusos no refinamento.

^{††} R_{iso} = $\sum (|F_{obs}(hkl) \text{ isomórfico} - F_{obs}(hkl) \text{ nativo}|) / 0,5 \times \sum (|F_{obs}(hkl) \text{ isomórfico} + F_{obs}(hkl) \text{ nativo}|)$. Neste caso, o valor de R_{iso} indica ter havido incorporação do NAD⁺.

Um critério mais exigente e revelador da acurácia do modelo e de melhorias durante o refinamento é o R -factor “livre”, R_{free} , computado com um pequeno conjunto de amplitudes escolhidas ao acaso que são retiradas desde o início e, portanto, omitidas no refinamento. Este conjunto “teste” é usado como controle de qualidade no processo de avaliação da concordância entre os dados observados e aqueles calculados a partir do modelo (Brunger 1992). Basicamente, em qualquer fase do refinamento, R_{free} mede quão bem o modelo atômico atual prediz um subconjunto das intensidades medidas que não foram incluídas no refinamento, enquanto R -factor mede quão bem o modelo atual prevê todo o conjunto de dados que produziu o modelo. Em geral, já em estágios finais de refinamento, os valores de R -factor e R_{free} tornam-se próximos (Rhodes 2000).

Além dos indicadores de convergência R -factor e R_{free} , outros parâmetros estruturais mostraram-se favoráveis ao modelo final proposto para NahI. Os valores de $R.m.s.D.$ (do

inglês *Root-mean-square Deviation*) obtidos para comprimentos e ângulos de ligação (0,015Å e 1,506°, respectivamente) enquadram-se dentre aqueles aceitos com base na geometria de moléculas orgânicas pequenas, o que corrobora a validade geométrica da estrutura. Valores de *R.m.s.D.* para comprimentos e ângulos de ligação indicam o desvio que a estrutura apresenta em relação aos valores esperados teoricamente em termos de geometria das ligações químicas. Um modelo bem refinado e quimicamente razoável apresenta *R.m.s.D.* de não mais que 0,02Å para comprimentos de ligação e 3° para ângulos de ligação (Engh & Huber 1991).

Mais ainda, os fatores de temperatura médios são um indicativo da mobilidade de diferentes partes da estrutura. Conforme o esperado, este valor é inferior para os átomos da proteína (29,76 Å²), que apresenta uma estrutura mais rígida e bem ordenada quando comparada às moléculas de água (40,68 Å²), que apresentam uma desordem bem mais significativa.

Por fim, outro parâmetro estrutural que confere maior confiabilidade ao modelo proposto para NahI é a conformação da cadeia polipeptídica, definida pelos pares de ângulos diedros Φ (*phi*) e ψ (*psi*) para cada aminoácido, cujos valores enquadram-se nas regiões favoráveis (96,9 %) e permitidas (3,1 %) no diagrama de *Ramachandran*.

4.4. MODELO CRISTALOGRÁFICO PROPOSTO PARA NahI

O modelo cristalográfico de 6xHis-NahI contém todos os 486 resíduos de aminoácidos da proteína nativa, enquanto a densidade eletrônica correspondente à cauda de histidinas N-terminal e o sítio de clivagem da protease TEV adicionados pelo vetor de expressão não foi observada no modelo obtido, o que indica se tratar de uma região flexível, altamente desordenada. A estrutura apresenta o enovelamento α/β típico da superfamília das ALDHs, que se organiza em três domínios (Figura 22 A): domínio de oligomerização e domínios de ligação ao dinucleotídeo e ao substrato, este último também conhecido como catalítico.

Os resíduos N-terminais que se estendem de Lys2 a His36 formam um conjunto de cinco fitas- β ($\beta 1 - \beta 5$). O enovelamento da cadeia polipeptídica prossegue por cinco α -hélices, inicia a composição do domínio de oligomerização por meio de duas fitas- β antiparalelas ($\beta 6$ e $\beta 7$) e retorna ao domínio de ligação ao dinucleotídeo para finalmente formar o característico enovelamento de *Rossmann*, usualmente observado em desidrogenases que utilizam NAD ou NADP como coenzima (Rossmann et al 1974). As cinco fitas- β paralelas $\beta 8$, $\beta 9$, $\beta 10$, $\beta 11$ e $\beta 12$ associadas às quatro α -hélices $\alpha 6$, $\alpha 7$, $\alpha 8$ e $\alpha 9$, duas dispostas em cada lado da folha- β , compõem um enovelamento similar ao de *Rossmann*, ou “*Rossmann-like*”, que, em outras ALDHs e conforme observado também em NahI, apresenta cinco fitas- β ao invés das seis tipicamente comuns em outras desidrogenases dependentes de NAD (Liu et al 1997) (Figura 22 A, C e D). A cadeia polipeptídica prossegue para o domínio de ligação ao substrato, ou catalítico, que também exhibe enovelamento α/β consistindo de seis fitas- β paralelas ($\beta 13$, $\beta 14$, $\beta 16$, $\beta 17$, $\beta 18$ e $\beta 19$), uma antiparalela ($\beta 15$) e seis α -hélices ($\alpha 10 - \alpha 15$). A cadeia então retorna se estendendo abaixo do domínio de ligação ao NAD ($\alpha 16$ e $\alpha 17$) e termina após formar a terceira fita- β ($\beta 20$) que completa o domínio de oligomerização. Este, portanto, formado por três fitas- β antiparalelas ($\beta 6$, $\beta 7$ e $\beta 20$), é responsável pela união dos monômeros em dímero formando uma alça sobre a entrada do domínio de ligação ao substrato da outra subunidade (Figura 22 A e B).

Em relação ao modelo de NahI-NAD⁺, foi identificada uma densidade eletrônica consistente com o cofator NAD no mapa de diferença $F_o - F_c$ (Figura 22 C e D). Por meio do mapa de diferença, podem-se visualizar porções positivas de densidade, correspondentes a regiões ainda não construídas do modelo, bem como áreas negativas de densidade, que apresentam demasiada população eletrônica em relação ao observado pelos dados experimentais.

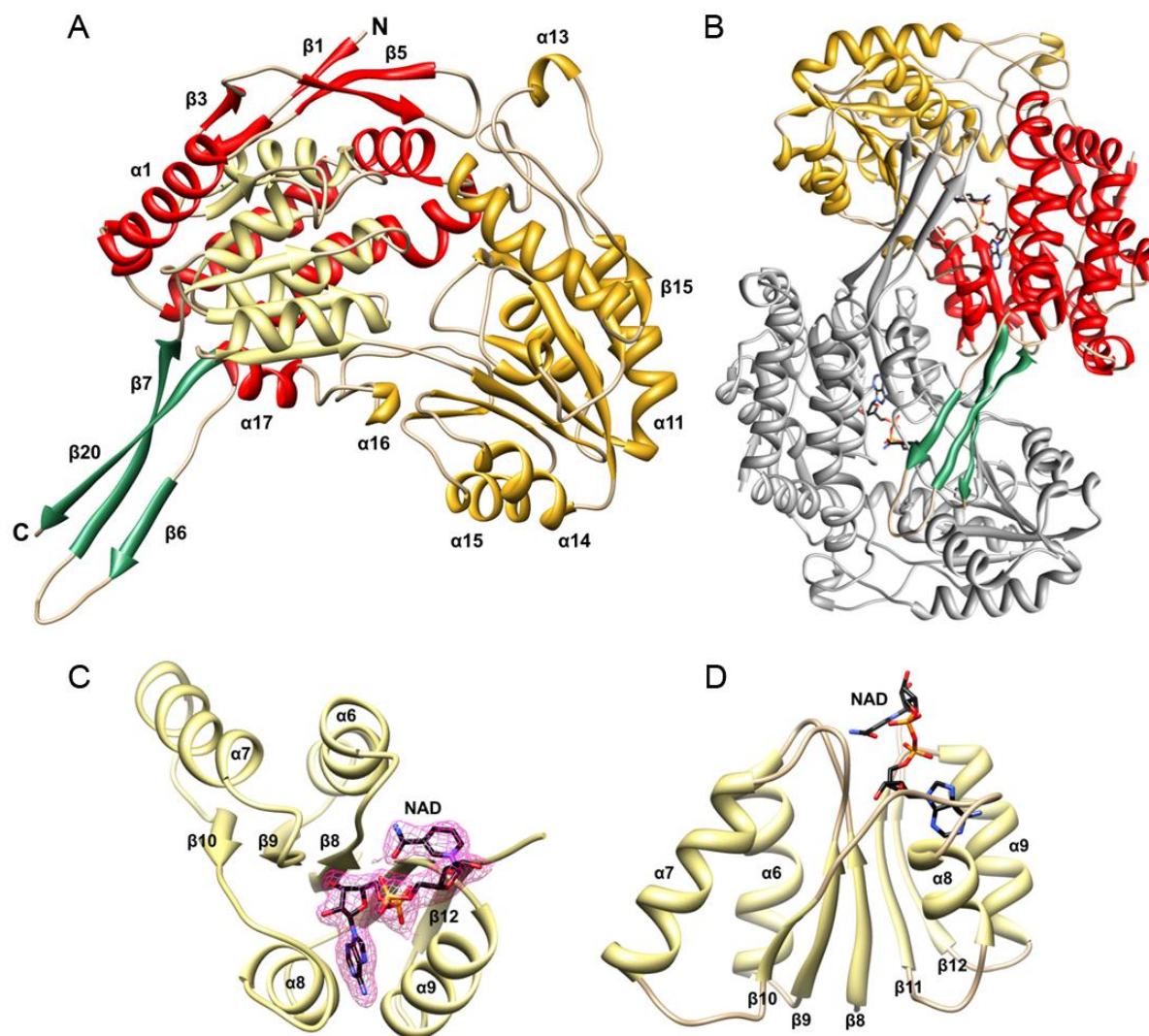


Figura 22. Estrutura cristalográfica da enzima NahI recombinante, uma 2-hidroxiuconato semialdeído desidrogenase de *Pseudomonas putida* G7, resolvida por difração de raios-X a 1,85 e 2,15 Å de resolução. **A.** Representação de fita do esqueleto polipeptídico evidenciando o enovelamento α/β típico da superfamília das ALDHs com destaque para os três domínios: domínio de oligomerização (em verde), domínio de ligação ao dinucleotídeo (em vermelho), também composto pelo enovelamento de *Rossmann* (em creme), e domínio de ligação ao substrato (em amarelo). Algumas hélices e fitas estão identificadas. **B.** Dímero cristalográfico de NahI a 2,15 Å de resolução, em que se observa o domínio de oligomerização se sobrepor à entrada do domínio de ligação ao substrato da outra subunidade, bem como duas moléculas do cofator NAD incorporadas. **C e D.** O enovelamento de *Rossmann* presente em NahI (“*Rossmann-like*”) apresenta cinco fitas- β paralelas ($\beta 8$ a $\beta 12$), ao invés das seis tipicamente comuns em outras desidrogenases dependentes de NAD, e quatro α -hélices ($\alpha 6$ a $\alpha 9$), duas dispostas em cada lado da folha- β . Destaque para a localização do NAD (estrutura em vareta) que, em C, apresenta o mapa de densidade eletrônica $F_o - F_c$ a 3σ (em magenta). As hélices $\alpha 8$ a $\alpha 9$ se distanciam em suas extremidades N-terminais permitindo acomodar o anel da adenina.

Quando superpostos os modelos de apo-NahI e NahI-NAD⁺, rearranjos conformacionais induzidos pela ligação do cofator não são perceptíveis, observação confirmada pelo baixo valor de *R.m.s.D.* global da cadeia polipeptídica igual a 0,17 Å. Estes dados apoiam aqueles obtidos por CD (item 4.1), em que os espectros de NahI a 20 °C, na

presença e na ausência de NAD^+ , são muito semelhantes um do outro ao longo de toda a gama de comprimento de onda medido (190 a 260 nm).

Ainda que uma única molécula tenha sido encontrada na unidade assimétrica do cristal, através de operações de simetria em torno de dois eixos cristalográficos observa-se a composição de um dímero que, através de outras duas operações cristalográficas, origina um tetrâmero de NahI (Figura 23), consistente com o estado oligomérico inferido por SAXS (item 4.2.2).

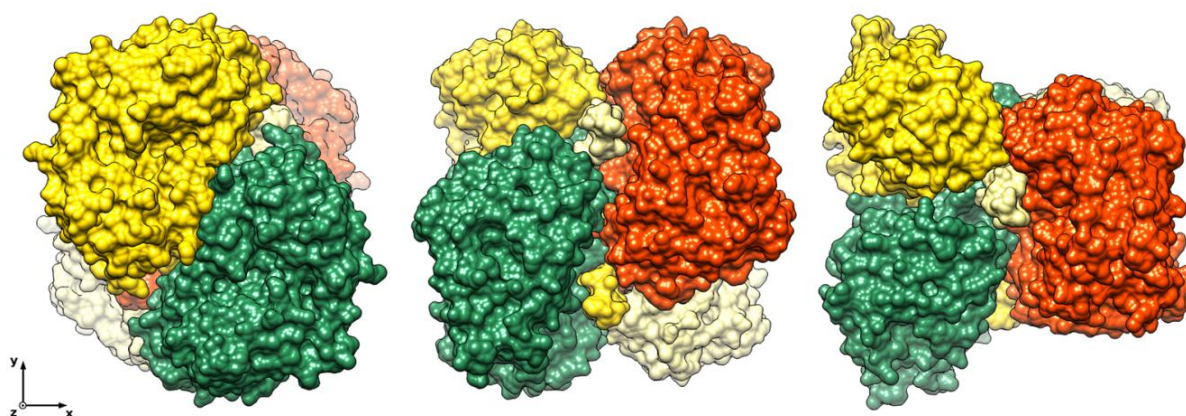


Figura 23. Modelo cristalográfico do tetrâmero da enzima NahI resolvido por difração de raios-X a 1,85 Å de resolução. O tetrâmero é composto por dois homodímeros (verde e amarelo / laranja e creme). Os domínios de oligomerização escondem-se no centro do oligômero, enquanto os sítios de ligação ao cofator se expõem na periferia. Quando comparados ao modelo central, o modelo da esquerda está rotacionado 90° em relação ao eixo y, enquanto o da direita está rotacionado aproximadamente 90° em relação ao eixo x.

Tal observação a respeito da estrutura quaternária de NahI não se enquadra à classificação sugerida por Perozich e colaboradores (Perozich et al 1999), segundo os quais a família HMSALDH (hidroximuconato semialdeído desidrogenase) é dimérica, ao contrário das outras famílias tetraméricas que integram a “Classe 1/2”. Todavia, a estrutura de uma γ -hidroximuconato semialdeído desidrogenase, também chamada de PnpE, enzima da via de degradação de *para*-nitrofenol em *Pseudomonas sp.* WBC-3, apresenta um tetrâmero cristalográfico similar ao de NahI (PDB id 4GO3 e 4GO4) (Su et al 2013). A enzima HpcC, uma 5-carboximetil-2-hidroximuconato semialdeído desidrogenase de *Thermus Thermophilus* HB8 (PDB id 2D4E), também apresenta um tetrâmero como sua unidade biológica.

4.5. SÍTIO DE LIGAÇÃO AO DINUCLEOTÍDEO E SUAS INTERAÇÕES COM O NAD^+

Conjuntos de dados de difração foram coletados de cristais previamente imersos em solução de NAD^+ e NADP^+ (Neves 2013), entretanto não se observou densidade eletrônica correspondente a este último.

O domínio de ligação ao cofator em NahI apresenta uma cavidade de volume igual a $1.066 \text{ \AA}^3$ para acomodação do NAD, conforme calculado pelo servidor CASTp (<http://cast.engr.uic.edu>) (Binkowski et al 2003), e é compreendida ao todo por trinta e quatro resíduos de aminoácidos, dos quais oito são provenientes do domínio de ligação ao substrato (Figura 24).

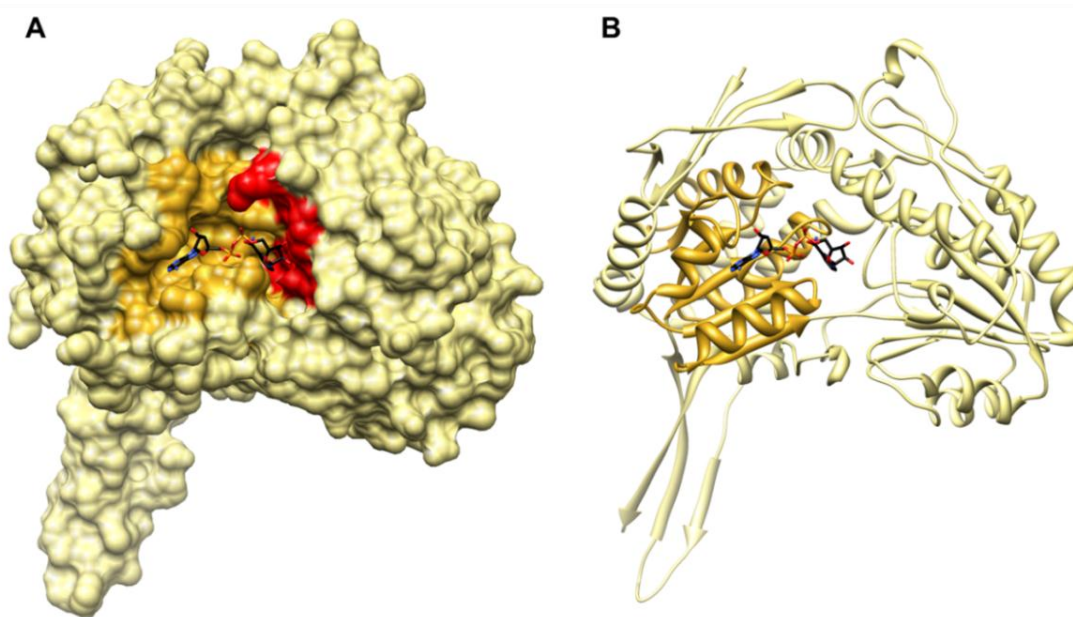


Figura 24. Localização do cofator NAD^+ na estrutura terciária de NahI resolvida por difração de raios-X a 2,15 Å de resolução. **A.** Imagem da superfície da proteína em que se evidencia a molécula de NAD acomodada na cavidade de $1.066 \text{ \AA}^3$ revestida por 34 resíduos de aminoácidos pertencentes aos domínios de ligação ao cofator (em amarelo dourado) e ao substrato (em vermelho). **B.** Representação de fita do esqueleto polipeptídico na mesma orientação de (A), com destaque para o enovelamento de *Rossmann* (em amarelo dourado), motivo estrutural essencial para a ligação do cofator.

Dentre os resíduos que compõem esta cavidade, é curioso destacar os papéis influentes de Glu181 e Lys178 na preferência pelo cofator NAD em NahI (Figura 25).

O resíduo Glu181 estabelece ligação de hidrogênio (2,70 Å) entre o grupo ácido de sua cadeia lateral e o grupo hidroxila situado na posição 2' da ribose da adenina (Figura 25). Em NADP, este grupo hidroxila é esterificado com fosfato, que pode ser repelido pelo resíduo ácido. Tal resíduo de glutamato é conservado entre ALDHs de classe 2, as quais são específicas ao NAD. Na ALDH classe 3 de rato (1AD3) (Liu et al 1997), cuja especificidade para NAD e NADP é ambígua, mutações sítio-dirigidas no resíduo de glutamato equivalente (Glu140) revelaram a contribuição deste para uma ligação mais forte ao NAD na enzima nativa, evidenciada pelo aumento nos valores de $K_{M(NAD)}$ após substituições de Glu140. Já o efeito destas mutações na afinidade ao NADP foi o oposto, visto pelo decréscimo nos valores de $K_{M(NADP)}$ (Perozich et al 2001; Perozich et al 2000).

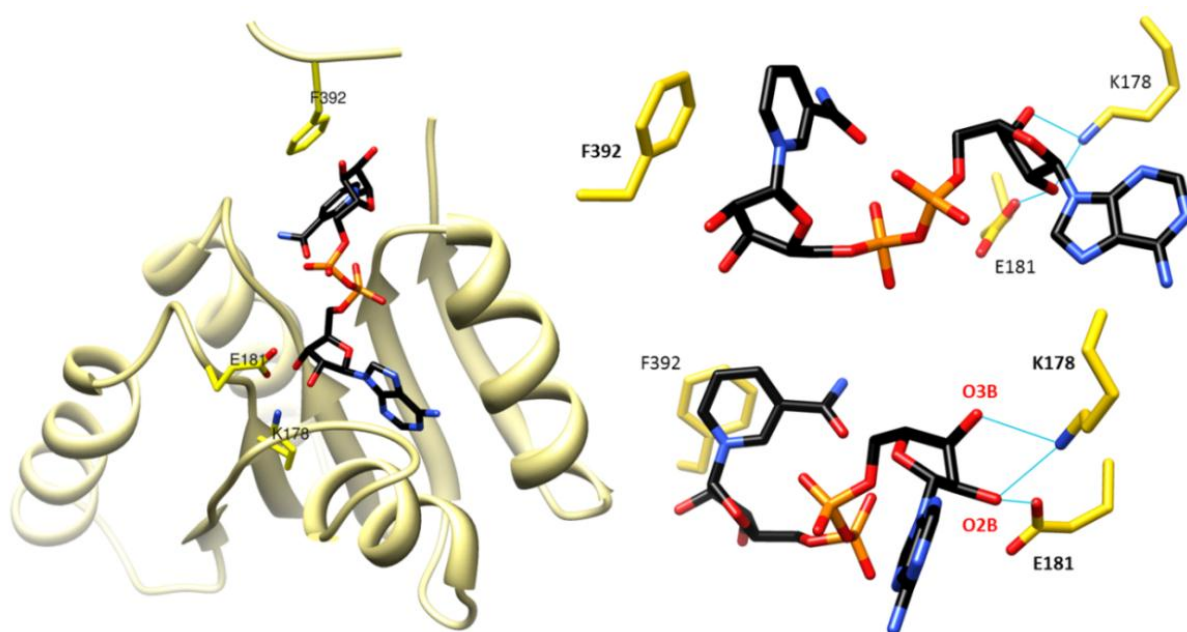


Figura 25. Interações entre o NAD e resíduos do sítio de ligação ao cofator em NahI. Os resíduos Lys178 e Glu181 estabelecem ligações de hidrogênio com grupos hidroxila da ribose da adenina e parecem definir a preferência da ALDH por NAD, ao invés de NADP, enquanto Phe392 contribui na orientação e estabilização do cofator através de interações do tipo π -stacking com o anel de nicotinamida. As distâncias das ligações de hidrogênio desenhadas na figura (traço em azul) são: Glu181 OE1 - O2B (2,70 Å), Lys178 NZ - O2B (2,83 Å) e Lys178 NZ - O3B (2,99 Å).

Segundo um alinhamento de sequências de ALDHs cuja especificidade pelo cofator foi estabelecida, dentre aquelas preferentes ao NAD, mais de 75% apresentaram o resíduo de glutamato frequente na posição crítica. Em contraste, dentre as ALDHs com preferência ao NADP como cofator, 60% e 30% contêm os resíduos de serina e treonina, respectivamente.

(Yuan et al 2013). Como exemplo, na ALDH de *Streptococcus mutans* (Sm-ALDH) responsável por catalisar a oxidação de gliceraldeído-3-fosfato a 3-fosfoglicerato com a concomitante redução de NADP^+ a NADPH, o resíduo de glutamato equivalente dá lugar a uma treonina, o que proporciona uma cavidade para acomodar e ligar o grupo fosfato do NADP (Cobessi et al 1999).

De fato, a adoção de resíduos de cadeia lateral mais curta nesta posição em ALDHs com preferência ao NADP parece evitar impedimento espacial e repulsão eletrostática do grupo fosfato. Em uma ALDH dependente de NADP^+ da bactéria *Vibrio harveyi*, a substituição de Thr175 por glutamato elevou dramaticamente o K_M para NADP^+ e reduziu o K_M para NAD^+ (Ahvazi et al 2000; Zhang et al 1999). Todavia, exceção é descrita para a enzima betaína aldeído desidrogenase (BADH) de *Pseudomonas aeruginosa* (PDB id 2WME), que apresenta o resíduo de glutamato (E179) em questão, ao contrário do esperado para ALDHs dependentes de NADP^+ . Entretanto, por meio de uma forte interação iônica com um resíduo adjacente de arginina (R40), a cadeia lateral do glutamato é mantida afastada do local de ligação do grupo fosfato 2', uma diferente solução encontrada para evitar confrontos estéricos e/ou repulsão eletrostática (González-Segura et al 2009).

Tais dados, portanto, corroboram o papel relevante do resíduo de Glu181 em NahI no reconhecimento seletivo do cofator NAD^+ . Ainda assim, determinar o K_M para NAD^+ e NADP^+ de NahI é uma perspectiva, uma vez que nos estudos citados para outras ALDHs, nenhuma mutação foi capaz de abolir a utilização de qualquer das coenzimas.

Outro resíduo em destaque na interação com o NAD é a Lys178, que estabelece ligação de hidrogênio com os grupos hidroxila 2' (2,83 Å) e 3' (2,99 Å) da adenosina (Figura 25). Tal resíduo é estritamente conservado em 97% das sequências de ALDHs até então conhecidas e sua presença parece ser crítica para a manutenção da estrutura local (Perozich et al 2000). Em uma ALDH classe 2, a forma mutante K192Q, cujo resíduo de lisina correspondente foi trocado por glutamina, reteve apenas 20% da atividade apresentada pela enzima nativa, além de um aumento de 100 vezes no K_M para NAD (Ni et al 1997). Novamente na ALDH classe 3, a perda do resíduo de lisina equivalente (Lys137) afetou a capacidade de ligação aos cofatores, NAD ou NADP, contribuindo para a inatividade dos mutantes (Perozich et al 2000).

Verifica-se também a participação da Phe392 em orientar e estabilizar o NAD por meio de interações do tipo π -stacking, em que se observa o empilhamento dos anéis aromáticos da cadeia lateral e da nicotinamida através de um arranjo praticamente paralelo (Figura 25).

Muitas outras interações estabelecidas com o cofator NAD ainda podem ser citadas, como as ligações de hidrogênio entre a porção de nicotinamida e os resíduos Trp154, Asn155, Lys338 e Glu390, bem como as interações hidrofóbicas entre o anel da adenina e os resíduos que o cercam (Figura 26). Como já descrito por Liu e colaboradores (1997), o sítio de ligação à adenina é de natureza hidrofóbica.

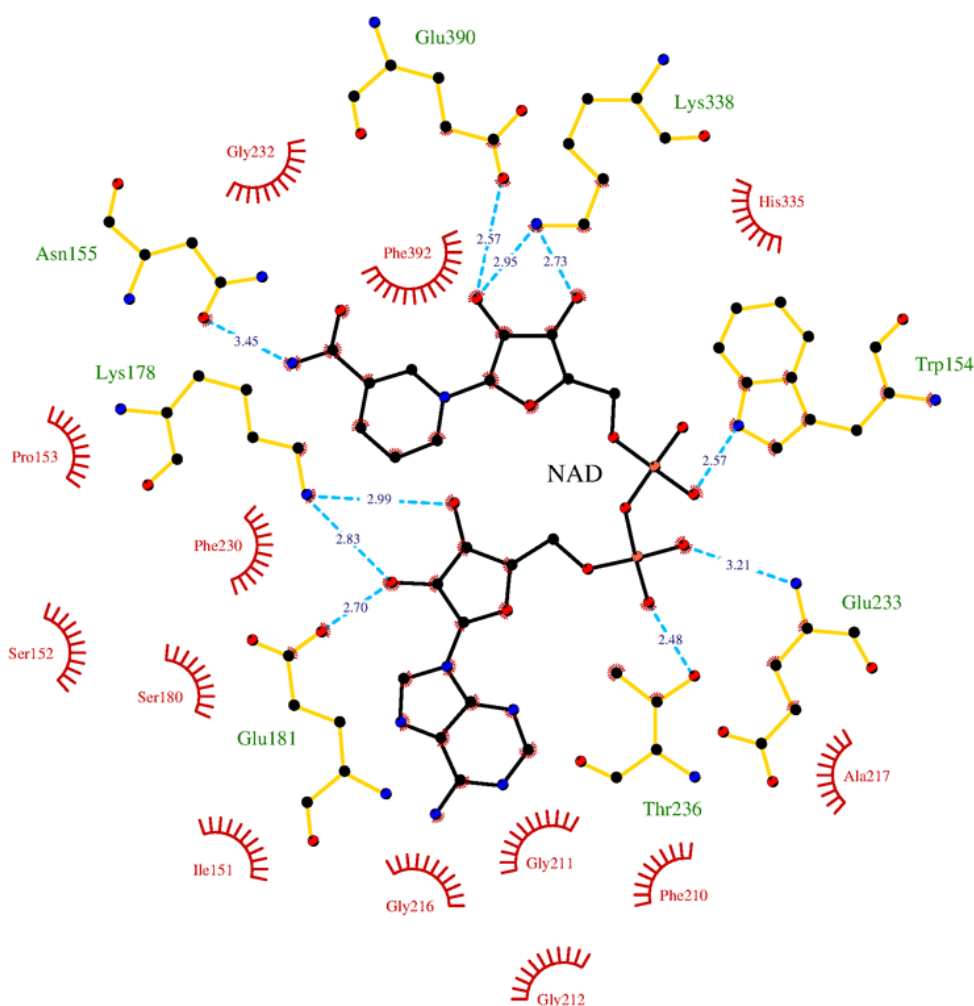


Figura 26: Ligações de hidrogênio e interações hidrofóbicas observadas no complexo NahI-NAD⁺. Resíduos em amarelo estabelecem ligações de hidrogênio, e em vermelho, interações hidrofóbicas. A distância de cada ligação de hidrogênio (linha pontilhada em azul) é dada em Å. Figura gerada no programa *LigPlot* (Wallace et al 1995).

4.6. SÍTIO DE LIGAÇÃO AO SUBSTRATO E ATIVIDADE CINÉTICA DE NahI

No que se refere ao domínio de ligação ao substrato, este apresenta uma cavidade essencialmente hidrofóbica de volume igual a $167 \text{ \AA}^3$ e em formato de funil revestida por seis resíduos provenientes do domínio de ligação ao dinucleotídeo (Arg106, Leu156, Leu159, Leu160, Trp163 e Glu254) e seis do domínio catalítico propriamente dito (Val287, Cys288, Leu289, Phe448, Arg450, Phe456) (Figura 27 A).

O resíduo de cisteína catalítico (Cys288), que forma o tio-éster intermediário com o aldeído, está localizado na base inferior do funil, que compreende a interface das duas cavidades de NahI. Assim, a Cys288 está situada a uma distância de $4,57 \text{ \AA}$ do anel de nicotinamida do NAD (Figura 27 B).

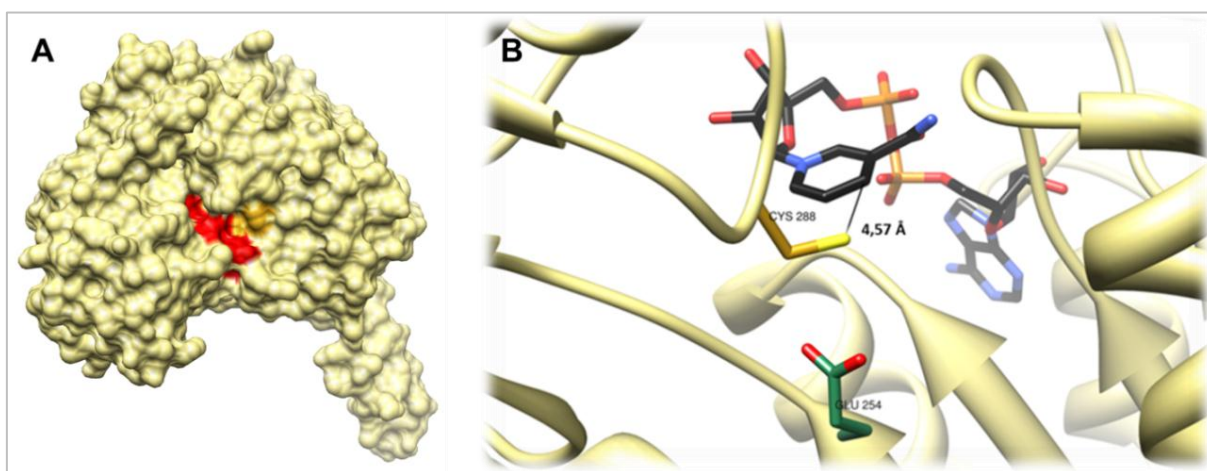


Figura 27: Sítio de ligação ao substrato. **A.** Imagem da superfície da proteína em que se evidencia a cavidade de $167 \text{ \AA}^3$ revestida por 12 resíduos de aminoácidos pertencentes aos domínios de ligação ao substrato (em vermelho) e ao cofator (em amarelo dourado). **B.** Visualizada através do sítio de ligação ao substrato, a Cys288 catalítica se posiciona entre as duas cavidades de NahI e se situa a uma distância de $4,57 \text{ \AA}$ do anel de nicotinamida (C4N) do NAD. Destaque também para o Glu254, outro resíduo catalítico.

A atividade cinética da enzima recombinante NahI complexada ao cofator NAD^+ foi monitorada por meio da diminuição da absorvância a 375 nm , condizente com o decréscimo da concentração do substrato 2-hidroxiuconato semialdeído.

Um gráfico das velocidades iniciais (V_0) da reação catalisada pela enzima sobre diferentes concentrações de substrato $[S]$ revela que NahI segue o modelo cinético de *Michaelis-Menten* (Figura 28), com valores de K_M e k_{cat} de $1,3 \pm 0,3 \text{ \mu M}$ e $0,9 \text{ s}^{-1}$,

respectivamente. O valor resultante de k_{cat}/K_M corresponde a $0,66 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ e é ligeiramente menor quando comparado ao observado para a enzima homóloga XylG do plasmídeo TOL pWW0 de *P. putida* mt-2 em relação ao 2-HMS ($k_{cat}/K_M = 1,6 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$) (Inoue et al 1995).

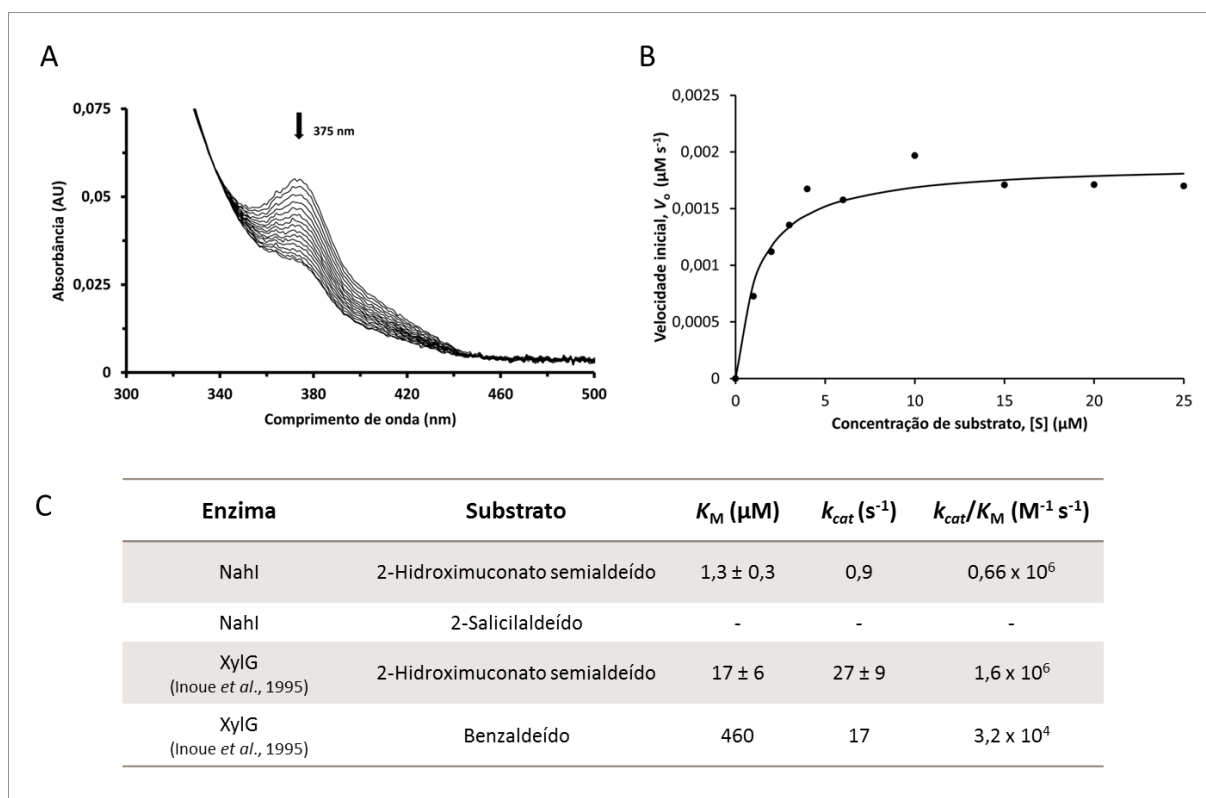


Figura 28: Atividade cinética de NahI. **A.** Atividade de NahI complexada ao cofator NAD^+ monitorada por meio da diminuição da absorvância a 375 nm, condizente com o decréscimo da concentração do substrato 2-hidroxiuconato semialdeído. **B.** Efeito do substrato 2-HMS sobre a velocidade inicial da reação catalisada por NahI. A atividade de NahI exibe uma dependência hiperbólica de V_0 sobre [S] e, portanto, segue o modelo cinético de *Michaelis-Menten*. O valor de K_M para o substrato 2-hidroxiuconato semialdeído é de aproximadamente $1,3 \mu\text{M}$. **C.** Constantes cinéticas de NahI e XylG. Para esta última, os dados foram extraídos de (Inoue et al 1995).

De acordo com as condições experimentais testadas, NahI não apresentou qualquer atividade sobre *orto*-salicilaldeído, substrato aromático natural de NahF. XylG, por sua vez, aparenta ser capaz de oxidar uma ampla gama de compostos, inclusive aromáticos como benzaldeído e seus análogos mono-substituídos com grupos metil-, metoxi-, nitro, cloro- e fluoro- em várias posições do anel, embora com menor especificidade em comparação ao alifático 2-HMS (Inoue et al 1995) (Figura 14, item 1.5).

Segundo Wang (2003), não é surpreendente que benzaldeído seja um substrato fraco para uma 2-hidroxiomuconato semialdeído desidrogenase como XylG, conforme evidenciado por seu elevado K_M (460 μM). Enquanto o substrato 2-HMS é uma molécula linear longa com substituintes polares, o benzaldeído é um composto cíclico, carente dos grupos carboxilato e hidroxila, ambos encontrados em 2-HMS e suscetíveis de serem importantes para a interação entre enzima e substrato. Com efeito, a presença de um grupo carboxilato em 2-HMS e em 4-carboxibenzaldeído resulta em um incremento na interação com a enzima em torno de 260 e 540 vezes, respectivamente, quando comparado ao benzaldeído (Wang 2003). É provável que o grupo carboxilato, mais adequado na posição *para*, mimetize melhor a conformação do substrato linear 2-HMS requerida pela enzima (Figura 29).

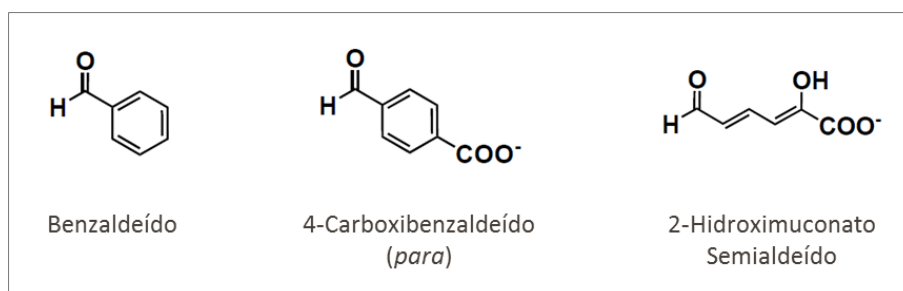


Figura 29: Substratos de XylG. A enzima XylG de *P. putida* mt-2, uma 2-HMSD homóloga de NahI, é capaz de oxidar compostos aromáticos tais como benzaldeído, embora com menor afinidade em comparação ao alifático 2-hidroxiomuconato semialdeído. A presença de um grupo carboxilato na posição *para* em 4-carboxibenzaldeído favorece a interação com a enzima, provavelmente por melhor mimetizar a conformação do substrato linear 2-HMS (Wang 2003).

Uma vez que o substrato 2-HMS possui os grupos carboxilato e hidroxila na extremidade oposta ao aldeído e, portanto, posicionados mais próximos à abertura da cavidade catalítica, sugere-se, com base na estrutura de NahI, que os dois resíduos de arginina R106 e R450 possam estabilizar o substrato por ligações de hidrogênio.

Na ALDH Sp2771 (PDB id 3VZ3) (Yuan et al 2013), uma succinato semialdeído desidrogenase da cianobactéria *Synechococcus*, o par de resíduos hidrofílicos Lys86 e Ser419, em posições equivalentes a Arg106 e Arg450 de NahI, contribuem para a estabilização do substrato linear succinato semialdeído através de ligações de hidrogênio (Figura 30). Corrobora para isto o efeito negativo que a substituição de Ser419 por alanina apresenta (mutante

Sp2771 S419A), ao ser capaz de reduzir em 80% a atividade catalítica de Sp2771 (Yuan et al 2013).

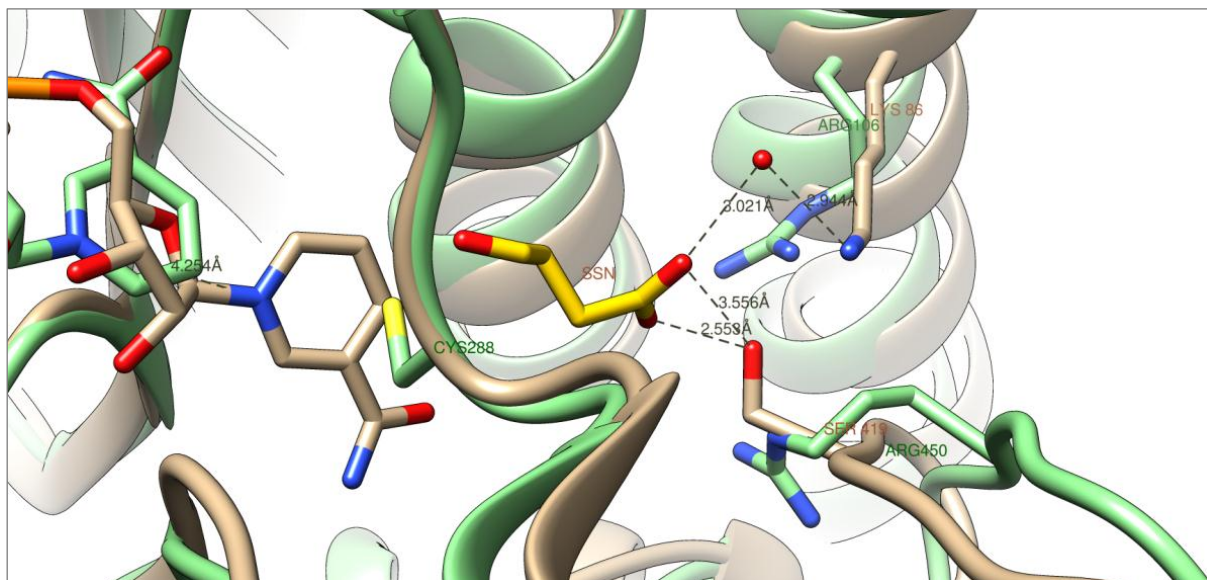


Figura 30: Comparação estrutural de possíveis resíduos interagentes com o substrato em NahI e Sp2771. Segundo Yuan e colaboradores (2013), na ALDH Sp2771 (marrom), uma succinato semialdeído desidrogenase da cianobactéria *Synechococcus*, o par hidrofílico K86 e S419 é responsável por estabilizar o substrato SSN (succinato semialdeído) através de ligações de hidrogênio, sendo uma intermediada por molécula de água. As distâncias em Å entre os átomos envolvidos nas possíveis ligações de hidrogênio estão apresentadas. Em NahI (verde), sugere-se que os resíduos de arginina em posições equivalentes, R106 e R450, realizem interações específicas com os grupos carboxilato e hidroxila de 2-HMS (não mostrado). *R.m.s.D.* global de 1,1 Å para 366 resíduos sobrepostos. Detalhe para o cofator NAD que, em Sp2771, apresenta o anel de nicotinamida mais próximo do substrato quando comparado ao ligado a NahI, cristalizada sem o 2-HMS. Os dois anéis de nicotinamida distam aproximadamente 4,3 Å entre si. Sp2771 C262A: PDB id 3ZV3 (Yuan et al 2013).

Um alinhamento das sequências de resíduos de aminoácidos de NahI e XylG, além de outras duas enzimas 2-HMSD, pode sugerir uma explicação razoável para o diferente comportamento mencionado anteriormente em torno da interação com substratos aromáticos (Figura 31).

Ainda que a identidade de sequência entre NahI e XylG de pWW0 seja elevada (87%) ao longo de toda a cadeia polipeptídica, há uma diferença notável que se estende por uma porção de 21 resíduos de aminoácidos consecutivos (Glu88 a Ala108) (Figura 31). Enquanto NahI, bem como as homólogas XylG do plasmídeo pDK1 de *P. putida* e DmpC de pVI150 de *Pseudomonas* sp. CF600 possuem a mesma sequência 88-ECLDTGKPKSLASHIDIPRGA-

108, XylG de pWW0 apresenta 88-RMPGHRQAEVAGQPHRHSARR-108, havendo apenas dois resíduos similares (sublinhados) entre os segmentos.



Figura 31: Alinhamento de sequencias de aminoácidos de NahI e três outras 2-HMSDs. As homólogas XylG [pDK1], DmpC [pVII50] e XylG [pWW0] compartilham, respectivamente, 91, 90 e 87% de identidade de sequencia com NahI. Os blocos em vermelho correspondem a 100% de identidade entre as quatro sequencias. Alinhamento realizado no servidor ESPript (<http://esprict.ibcp.fr>) (Robert & Gouet 2014).

Até o momento não há estrutura tridimensional disponível para XylG (pWW0) no PDB, embora Inoue e colaboradores tenham obtido cristais que difrataram a 2,5 Å de resolução (Inoue et al 1995). Curiosamente, a localização deste segmento na estrutura cristalográfica de NahI revela sua proximidade à entrada do sítio de ligação ao substrato (Figura 32), o que sugere um possível papel na especificidade destas enzimas por seus ligantes.

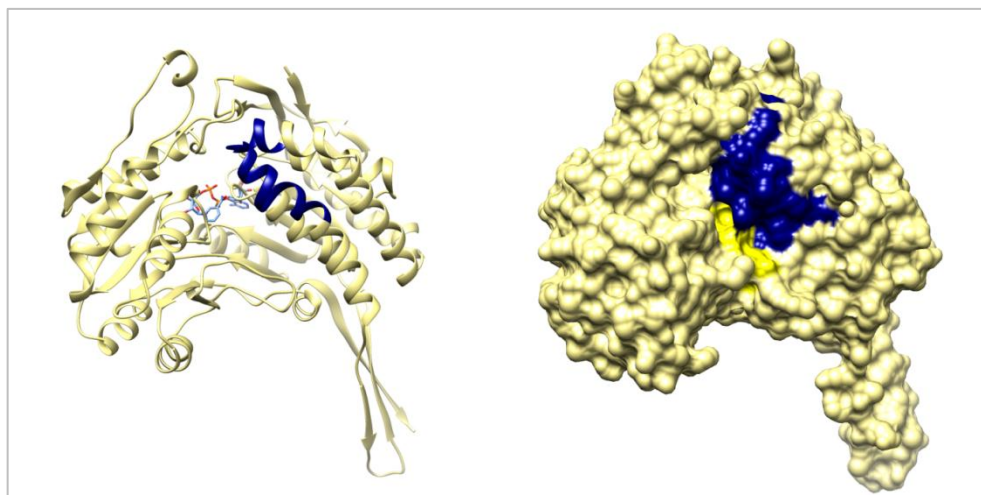


Figura 32: Localização da porção compreendida por Glu88 e Ala108 na estrutura cristalográfica de NahI. Em azul, destaca-se o segmento de 21 resíduos de aminoácidos consecutivos, diferença principal entre as estruturas primárias de NahI e XylG [pWW0], que se localiza em uma região helicoidal circunvizinha à entrada do sítio de ligação ao substrato (em amarelo) em NahI. Da esquerda para a direita: Representação de fita do esqueleto polipeptídico e da superfície da proteína, ambas na mesma orientação. NAD: estrutura em vareta.

De fato, diferenças na acessibilidade ao sítio ativo parecem influenciar a especificidade ao substrato entre ALDHs. RalDH2, por exemplo, uma retinaldeído desidrogenase tipo II (PDB id 1BI9) (Lamb & Newcomer 1999), eficientemente oxida uma variedade de aldeídos alifáticos como octanal, decanal e retinal, mas não acetaldeído. Ao contrário do que foi descrito para ALDH2 (Steinmetz et al 1997) e ALDH3 (Liu et al 1997) em que o canal de acesso ao substrato é uma cavidade pré-formada por meio da qual o acetaldeído pode se difundir, RalDH2 utiliza uma alça desordenada de vinte aminoácidos para limitar o acesso ao sítio catalítico por aldeídos de cadeia curta. Por meio de um ajuste do tipo “mão na luva” (*hand-in-glove-like fit*), a ligação do retinal estabiliza a alça bem como a maquinaria catalítica. Assim, através deste mecanismo, a enzima pode discriminar entre aldeídos de cadeia hidrocarbônica longa e curta, uma vez que estes últimos não possuem a larga superfície hidrofóbica necessária para se enterrar no interior do canal e consequentemente estabilizar os grupos catalíticos (Bordelon et al 2004; Lamb & Newcomer 1999).

Harayama e Rekik (Harayama & Rekik 1993) propuseram que a via de *meta*-clivagem do plasmídeo TOL pWW0 de *P. putida* mt-2 seja composta de genes híbridos que evoluíram a partir de um produto de recombinação entre sequências homólogas ancestrais *nah* (NAH7 de *P. putida* G7) e *dmp* (pVI150 de *P. putida* CF600). Em organismos que contêm tais plasmídeos, a chance de recombinação entre genes não idênticos das vias de *meta*-clivagem

pode ser significativa (Harayama & Rekik 1993). Deste modo, é tentador sugerir que tal fragmento possa ter sido incorporado ao gene *xylG* conduzindo a uma mudança dramática na especificidade desta enzima.

Por outro lado, quando feito um alinhamento das sequências nucleotídicas de *nahI* e *xylG*, observa-se um desvio de enquadramento da janela de leitura do segmento de 63 pares de bases correspondente à sequência de 21 resíduos de aminoácidos, provavelmente originado por eventos de inserção e deleção nas posições flangeadoras deste segmento (Figura 33). Portanto, é possível que a sequência *xylG* tenha sido selecionada positivamente, uma vez que o produto possui algumas propriedades catalíticas vantajosas sobre substratos aromáticos.

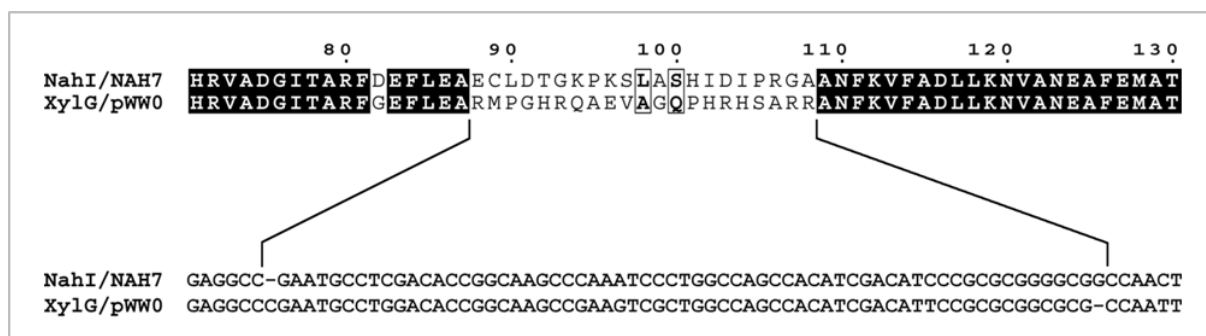


Figura 33. Alinhamento das sequências nucleotídicas de NahI (NAH7) e XylG (pWW0). O desvio de enquadramento de janela de leitura do segmento de 63 pb em XylG correspondente à sequência de 21 resíduos de aminoácidos, diferença principal entre as estruturas primárias das duas 2-HMSDs. A inserção de uma citosina e a deleção de uma guanossina nas extremidades do segmento de 63 pb que codifica para uma sequência de 21 resíduos de aminoácidos pode ser uma explicação para a origem das propriedades catalíticas vantajosas de XylG sobre substratos aromáticos. Alinhamento realizado no servidor ESPript (<http://esprpt.ibcp.fr>) (Robert & Gouet 2014).

4.7. ANÁLISE COMPARATIVA DOS SÍTIOS DE LIGAÇÃO AO SUBSTRATO DE NahI E NahF

Uma vez conhecida a estrutura tridimensional de NahI, é sem dúvida notável a semelhança de seu arcabouço estrutural com o da enzima NahF, a outra ALDH da via de degradação do naftaleno de *P. putida* (Figura 34), embora as duas compartilhem somente 33% de identidade de sequência (Figura 35).

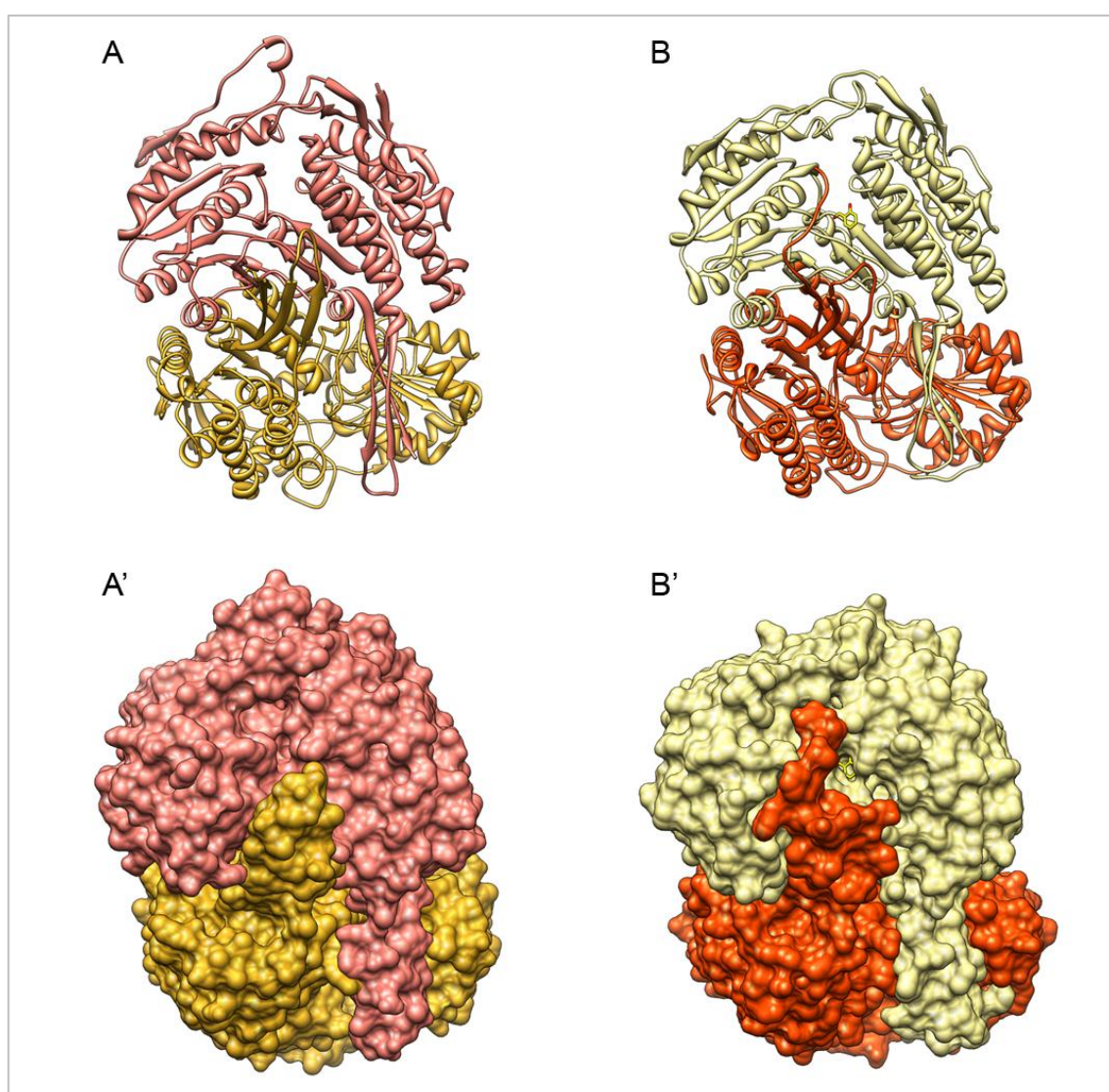


Figura 34: Estruturas diméricas das enzimas NahI e NahF de *P. putida* G7. A e B. Representação de fita das cadeias polipeptídicas do dímero de NahI (A) e de NahF (B), ambos posicionados na mesma orientação. Quando sobrepostas as cadeias polipeptídicas de NahI e NahF, o *R.m.s.D.* global é de 1,12 Å sobre 354 resíduos de aminoácidos. A' e B'. Imagens da superfície das moléculas, NahI (A') e NahF (B'), evidenciando em cada dímero os domínios de oligomerização que se estendem sobre a entrada do sítio catalítico na subunidade vizinha. NahF foi co-cristalizada com o substrato salicilaldeído (PDB id 4JZ6) (Coitinho 2013), visualizado em amarelo no sítio catalítico. Apenas para facilitar a comparação, NahI está representada pelo dímero ao invés do tetrâmero.

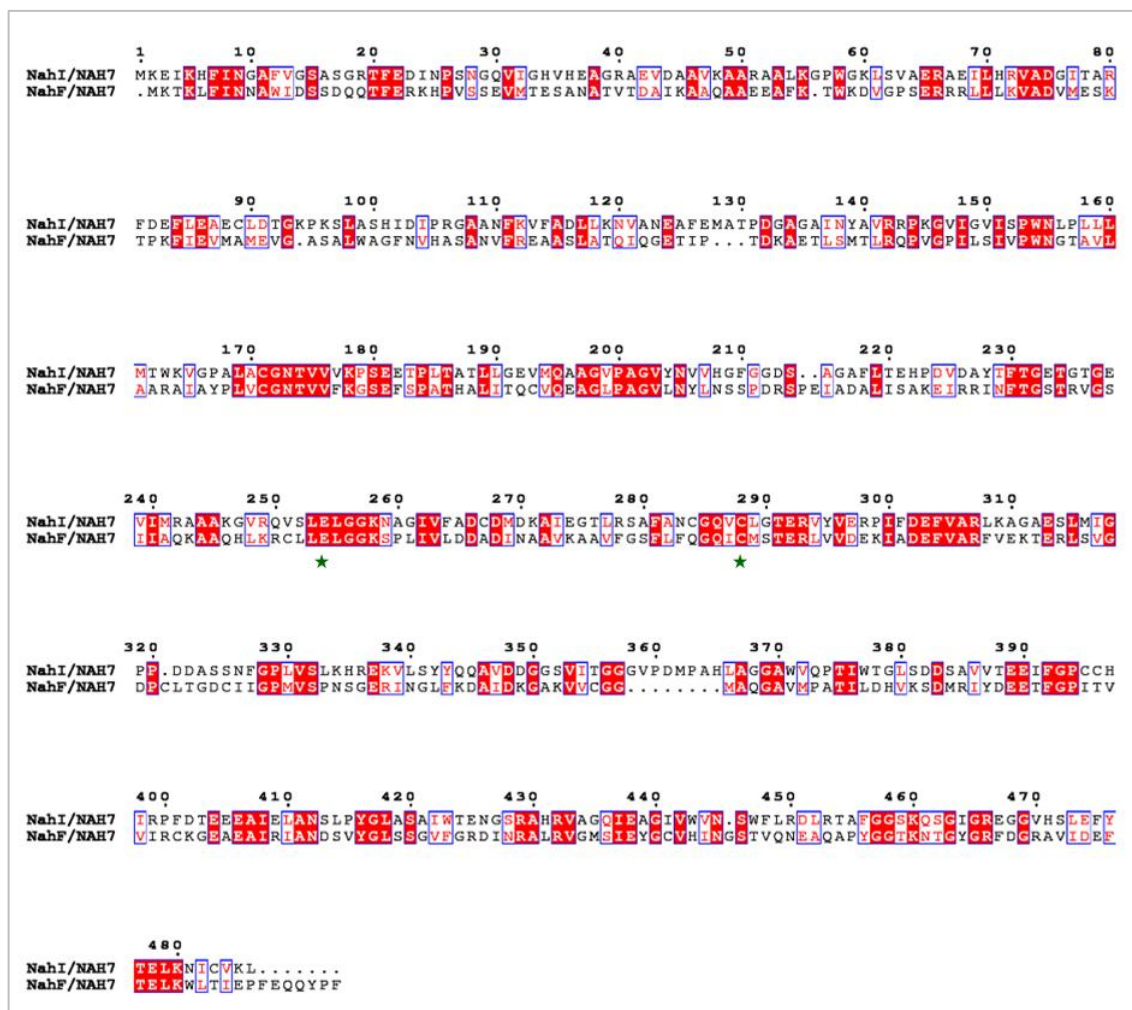


Figura 35. Alinhamento entre as sequencias de NahI e NahF. Os blocos em vermelho correspondem às porções de identidade de sequência - apenas 33% - compartilhada entre as duas ALDHs. As estrelas em verde sinalizam os resíduos catalíticos de cisteína e glutamato. Alinhamento realizado no servidor ESPrpt (<http://esprpt.ibcp.fr>) (Robert & Gouet 2014).

Em NahF, a presença da Arg157 no sítio catalítico a 4,63 Å de distância do substrato pode indicar um papel importante para este aminoácido durante a catálise de substratos aromáticos. Esta mesma posição em NahI é ocupada por um Trp163, resíduo de cadeia lateral volumosa que pode apresentar impedimento espacial para a acomodação de substratos aromáticos (Figura 36 A e A'). Condição semelhante também pode ser sugerida para a Leu156 em NahI, que dá lugar para uma Gly150 em NahF posicionada a uma distância de 3,0 Å em relação ao substrato (Figura 36 A e A'). Assim, pode ser que em NahI, a presença de resíduos com cadeia lateral volumosa no sítio catalítico dificulte a atividade desta enzima sobre substratos aromáticos.

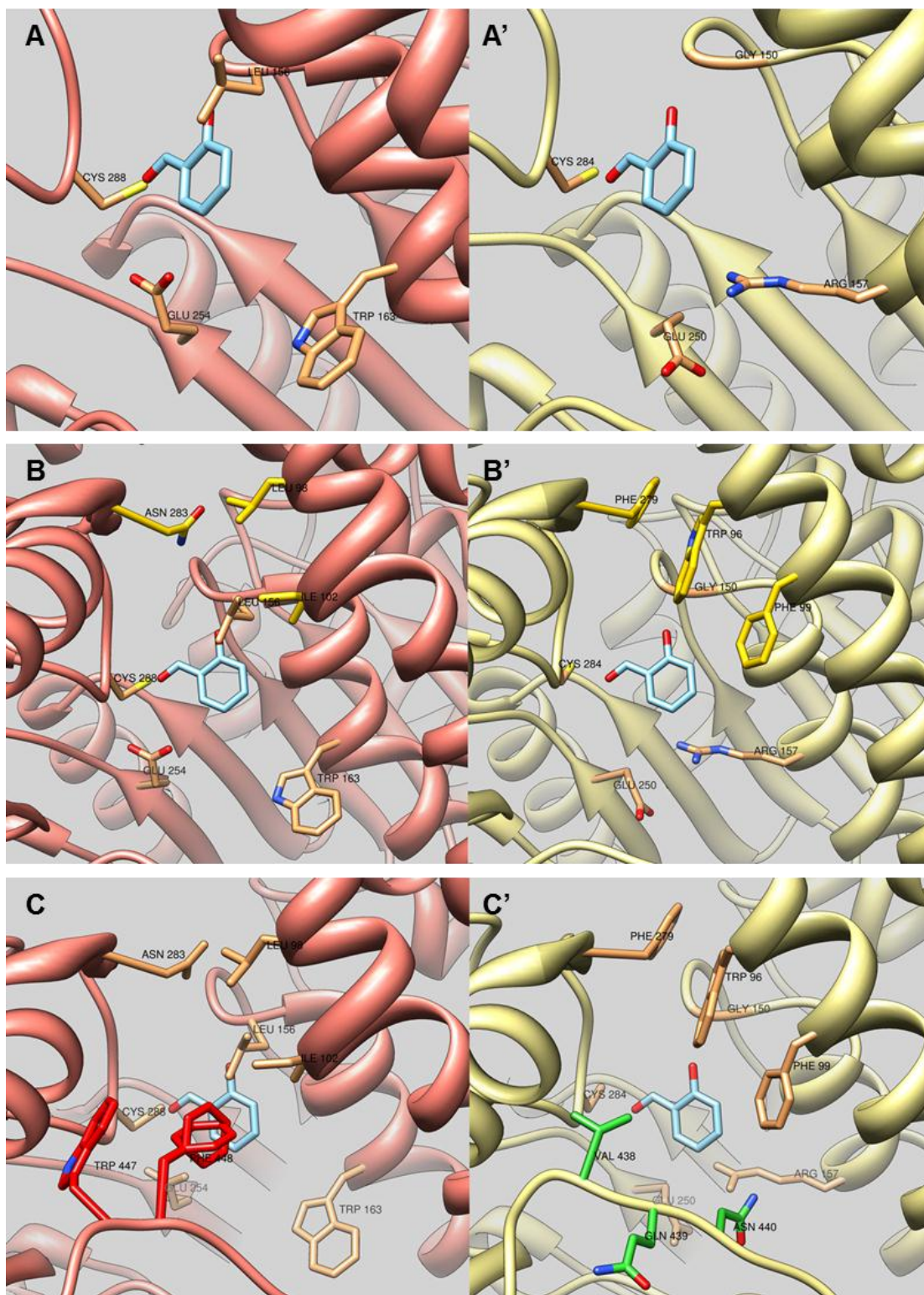


Figura 36: Análise comparativa dos sítios de ligação ao substrato de NahI e NahF. A e A'. No sítio ativo de NahI (A), a presença de resíduos de cadeia lateral volumosa, como Trp163 e Leu156, poderia causar impedimento espacial para a acomodação de substratos aromáticos. Em NahF (A'), posições equivalentes são ocupadas por Arg157 e Gly150, respectivamente. Os resíduos catalíticos Cys e Glu são conservados entre as ALDHs. O ligante em azul corresponde ao salicilaldeído, substrato aromático de NahF. **B e B'.** O trio aromático Trp96, Phe99 e Phe279 em NahF (B') adota um arranjo parcialmente paralelo estabelecendo interações do tipo π -stacking e está ausente em posições equivalentes em NahI (B). **C e C'.** Resíduos hidrofóbicos, como Trp447 e Phe448, se posicionam na entrada do sítio de ligação ao substrato de NahI (C) sugerindo uma barreira à passagem de substratos aromáticos.

Os resíduos aromáticos Trp96, Phe99 e Phe279 dispostos na entrada do sítio ativo de NahF estão ausentes nas posições equivalentes em NahI, ocupadas por Leu98, Ile 102 e Asn283. O trio aromático em NahF forma uma base na entrada do sítio ativo estabilizada por interações do tipo π -stacking, quando há empilhamento dos anéis aromáticos por adotarem um arranjo paralelo, o que parece ser importante para a manutenção da abertura do bolsão catalítico (Figura 36 B e B').

Mais ainda, é interessante notar que a ausência do trio aromático em NahI talvez permita melhor acomodação da Phe448, que se projeta na entrada do sítio da ALDH alifática sugerindo um bloqueio à passagem de substratos aromáticos (Figura 36 C).

Curiosamente, uma análise do volume do sítio ativo, calculado pelo servidor CASTp (Binkowski et al 2003), revelou uma expansão do bolso catalítico de $167 \text{ \AA}^3$ para $635 \text{ \AA}^3$ ao substituir unicamente o resíduo Phe448 de NahI pelo aminoácido equivalente em NahF, sugerindo ao resíduo hidrofóbico, e à própria expansão da cavidade, um papel de seletividade ao substrato, uma vez que o volume disponível à acomodação do substrato torna-se semelhante ao apresentado por NahF (Figura 37 C).

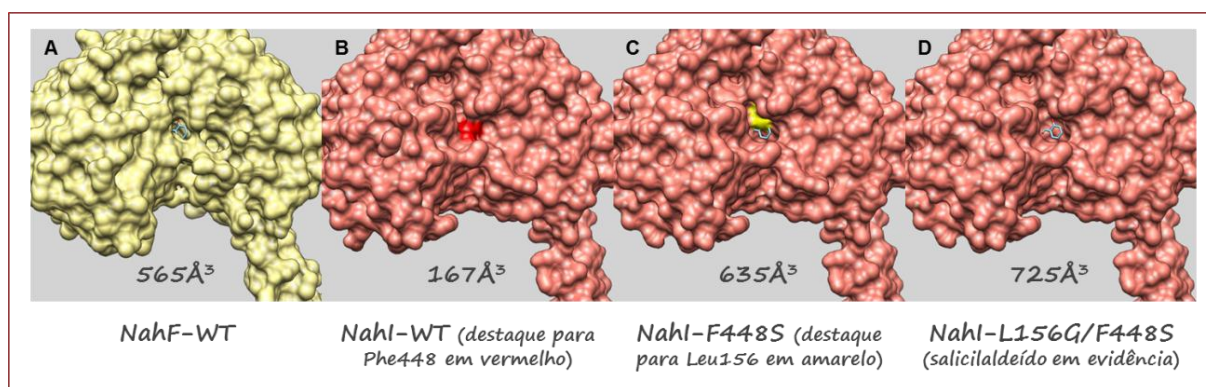


Figura 37: Expansão dos sítios de ligação ao substrato em formas mutantes de NahI. O volume da cavidade catalítica de NahI ($167 \text{ \AA}^3$, em B) sofre expansão para valores próximos ao apresentado por NahF ($565 \text{ \AA}^3$, em A) quando os resíduos de cadeias laterais volumosas Leu156 e Phe448 são substituídos pelos equivalentes na ALDH aromática (C e D).

Expansão maior da cavidade catalítica foi calculada para o duplo mutante de NahI ($725 \text{ \AA}^3$) em que, além da própria Phe448 já referida anteriormente, a Leu156 posicionada na base do funil é substituída pela glicina equivalente em NahF (Figura 37 D). Quando

substituído unicamente o resíduo de Leu156 em NahI por glicina, o sítio de ligação ao substrato apresenta um volume predito de 464 Å³ (omitido na Figura 37).

Com base neste estudo comparativo prévio, a fim de se compreender os determinantes estruturais em NahI para a seleção do substrato, iniciou-se a caracterização estrutural dos três mutantes propostos NahI-L156G, NahI-F448S e NahI-L156G-F448S.

4.8. PURIFICAÇÃO DAS VARIANTES DE NahI

As formas mutantes de NahI foram purificadas por cromatografia de afinidade ao níquel e, em seguida, por gel-filtração (Figura 38, itens A e C). Para obtenção das proteínas recombinantes sem a cauda de seis resíduos de histidina, entre as duas etapas cromatográficas já mencionadas realizou-se incubação com a protease TEV e uma segunda passagem em coluna de níquel (Figura 38, itens B). A protease TEV, por apresentar em sua extremidade N-terminal uma cauda de histidinas, permanece retida na coluna juntamente com contaminantes provenientes da primeira etapa de purificação. A proteína recombinante clivada, entretanto, por não mais apresentar a etiqueta de afinidade, não interage com a coluna sendo recolhida nas primeiras frações.

Após a última etapa cromatográfica, de exclusão molecular, em que se revelou alto grau de pureza das formas mutantes recombinantes de NahI conforme observado em gel de poliacrilamida-SDS (Figura 39), as proteínas foram concentradas para os ensaios de cristalização (Tabela 6). O procedimento de purificação das formas mutantes de NahI possibilitou recuperar aproximadamente de 25 a 30 mg de proteína para cada litro de meio de cultura.

Tabela 6: Concentrações (em mg/mL) das formas mutantes de NahI utilizadas nos ensaios de cristalização.

Mutante NahI	Com 6xHis (mg/mL)	Clivada (mg/mL)
L156G	15	20
F448S	22	25
L156G-F448S	14	20

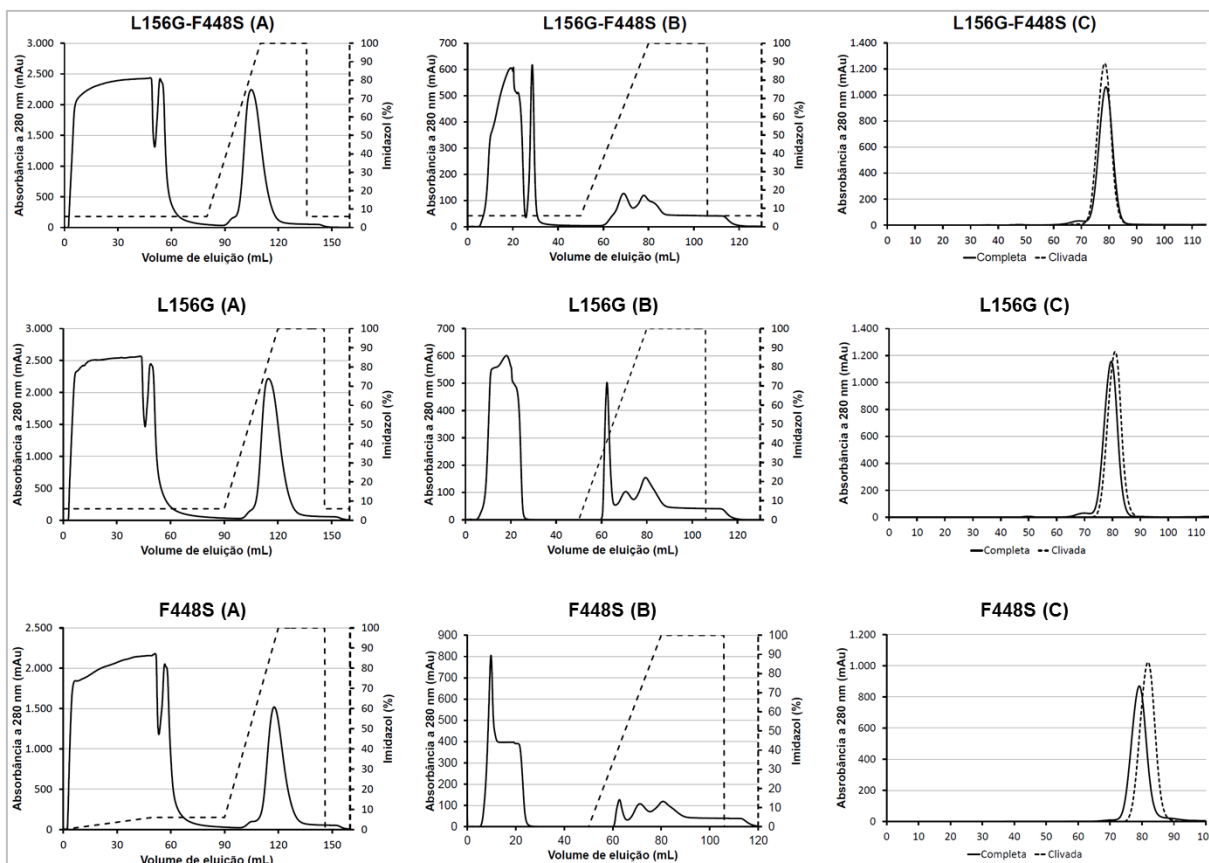


Figura 38: Cromatogramas obtidos durante as etapas de purificação das formas mutantes de NahI. Cromatografia de afinidade em coluna de níquel a partir de extrato bruto celular (A) e após clivagem da cauda de histidinas com a protease TEV (B). Cromatografia de exclusão molecular da proteína completa (linha cheia) e clivada (linha pontilhada) (C). A absorvância a 280 nm em mAu é dada em função do volume de eluição da coluna em mL.

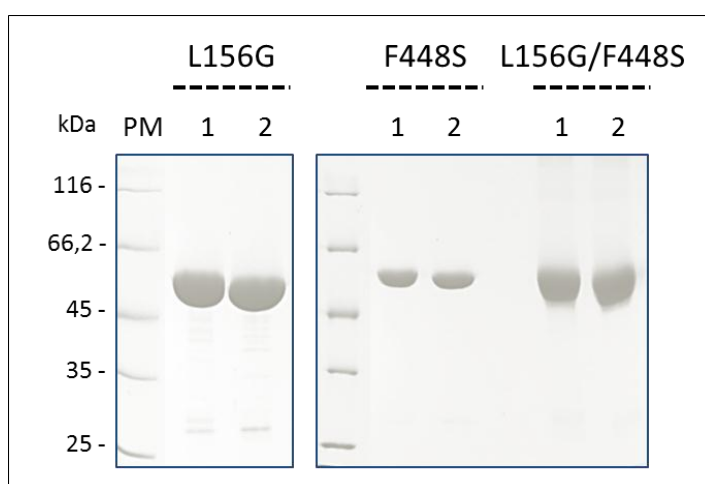


Figura 39: Perfil eletroforético das formas mutantes de NahI após a etapa cromatográfica de exclusão molecular. Canaletas 1 e 2 correspondem respectivamente à proteína recombinante completa e clivada (sem a cauda de afinidade). Géis de poliácridamida-SDS 12% corados com azul de Coomassie. PM (kDa): peso molecular em kiloDaltons (*Unstained Protein Molecular Weight Marker – Thermo Scientific*).

4.9. CRISTALIZAÇÃO, DIFRAÇÃO DE RAIOS-X E ESTRUTURA DAS VARIANTES DE NahI

Conforme previamente estabelecido para a enzima NahI, condições variadas de formato de sódio também mostraram-se favoráveis ao crescimento de cristais das formas mutantes de NahI (Figura 40). Padrões de difração de raios-X foram obtidos para as três variantes na ausência de NAD^+ a partir de cristais crescidos nas condições especificadas na Tabela 7, na qual são listadas as estatísticas dos dados de difração e de processamento.

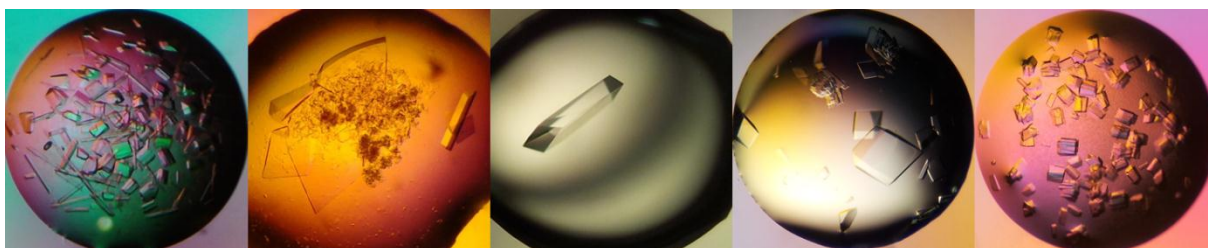


Figura 40: Cristais de formas mutantes de NahI, crescidos em condições variadas de formato de sódio.

Cristais do duplo mutante NahI-L156G-F448S apresentaram os grupos espaciais C2 e C222_1 , tendo o cristal deste último grupo atingido a melhor resolução de 2,9 Å com seis monômeros na unidade assimétrica (Figura 41). Um dos cristais de NahI-F448S pertencente ao grupo C2 difratou a 3,25 Å de resolução com doze monômeros na unidade assimétrica, enquanto o cristal de NahI-L156G, grupo P2_1 , difratou a 2,1 Å. A estrutura cristalina de cada mutante foi resolvida pelo método de substituição molecular tendo a própria NahI como modelo. Apenas NahI-L156G ainda não foi determinada, enquanto NahI-F448S e L156G-F448S estão em fase de refinamento. Em ambas as estruturas, a unidade biológica tetramérica foi observada por meio de operações de simetria da unidade assimétrica.

Tabela 7: Estatísticas dos dados de difração e de processamento dos cristais das formas mutantes de NahI. Os valores estatísticos para as faixas de menor e maior resolução são dadas entre parênteses.

L156G		F448S		L156G-F448S		
Condição de crescimento do cristal	Formato de Sódio 4M Hepes-Na pH 7,0	Formato de Sódio 2,8M Hepes-Na pH 7,8	Formato de Sódio 2,8M Hepes-Na pH 7,8	Formato de Sódio 3,8M Hepes-Na pH 8,2	Formato de Sódio 2,4M Hepes-Na pH 7,2	Formato de Sódio 3,6M Hepes-Na pH 8,2
Comprimento de onda (Å)	1,459	1,459	1,459	1,459	1,459	1,459
Distância cristal-detector (mm)	110	200	210	110	210	210
Número de imagens	952	419	310	490	540	500
Número de imagens processadas	952	340	310	490	540	500
Rotação por imagem (°)	0,2	0,4	0,5	0,3	0,3	0,3
Variação total de rotação (°)	190,4	136	155	147	162	150
Grupo espacial	P2 ₁	C2	C2	C2 2 2 ₁	C2	C2
Parâmetros da célula unitária (Å, °)	a = 188,86 b = 157,67 c = 188,70 $\alpha = \gamma = 90^\circ$ $\beta = 119,93^\circ$	a = 223,92 b = 310,66 c = 160,52 $\alpha = \gamma = 90^\circ$ $\beta = 96,52^\circ$	a = 223,44 b = 310,57 c = 160,03 $\alpha = \gamma = 90^\circ$ $\beta = 96,48^\circ$	a = 188,43 b = 327,21 c = 158,21 $\alpha = \beta = \gamma = 90^\circ$	a = 221,23 b = 310,78 c = 160,99 $\alpha = \gamma = 90^\circ$ $\beta = 96,3^\circ$	a = 222,09 b = 308,79 c = 159,30 $\alpha = \gamma = 90^\circ$ $\beta = 96,3^\circ$
Faixa de resolução (Å)	50,0 – 2,1 (50,0 – 5,7 / 2,14 – 2,10)	50,0 – 3,4 (50,0 – 9,21 / 3,46 – 3,40)	50,0 – 3,25 (50,0 – 8,81 / 3,31 – 3,25)	50,0 – 2,9 (50,0 – 7,86 / 2,95 – 2,90)	50,0 – 3,16 (50,0 – 8,56 / 3,21 – 3,16)	50,0 – 3,4 (50,0 – 9,21 / 3,46 – 3,40)
Número de observações	2.111.032	405.356	541.046	632.215	611.772	422.569
Número de reflexões únicas	558.026 (28.209 / 27.775)	133.440 (6.457 / 6.982)	169.047 (8.446 / 8.442)	104.113 (5.112 / 5.187)	183.782 (9.216 / 9.130)	138.493 (6.636 / 7.102)
Completeza (%)	100 (99,4 / 100)	90,7 (86,6 / 96)	99,8 (98,4 / 100)	96,9 (90,4 / 97,5)	99,9 (98,8 / 100)	95,1 (89,9 / 98)
$\langle I/\sigma(I) \rangle$	12,8 (24,8/3,1)	10,7 (31,6/2,4)	11,9 (27,4/2,4)	11,5 (19,2/2,9)	12,9 (25,7/2,5)	12,8 (35,1/2,2)
Redundância	3,8 (4,0 / 3,9)	3,0 (3,3 / 3,0)	3,2 (3,3 / 3,3)	6,1 (6,4 / 5,6)	3,3 (3,4 / 3,4)	3,1 (3,3 / 3,1)
$R_{merge}^\dagger$	0,089 (0,038 / 0,462)	0,103 (0,028 / 0,526)	0,088 (0,028 / 0,527)	0,112 (0,056 / 0,555)	0,078 (0,035 / 0,509)	0,078 (0,025 / 0,536)
Mosaicidade média (°)	-	0,458	0,364	0,298	0,364	-
$R\text{-factor}^{\dagger\dagger}$ (%)	-	-	20,7	18,9	-	-
$R_{free}^{\dagger\dagger}$ (%)	-	-	26,6	25,3	-	-

$^\dagger R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, em que $I_i(hkl)$ é a intensidade medida de cada reflexão hkl e $\langle I(hkl) \rangle$ é sua média.

$^{\dagger\dagger} R\text{-factor} = \sum ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \sum |F_{obs}(hkl)|$. R_{free} é o valor de $R\text{-factor}$ calculado para 5% do conjunto de dados não inclusos no refinamento.



Figura 41: Célula unitária do cristal de NahI-L156G-F448S. Pertencente ao grupo espacial $C222_1$, o cristal difratou a 2,9 Å de resolução apresentando seis monômeros na unidade assimétrica. Por operações de simetria, é representado todo o conteúdo da célula unitária. Ao centro, observa-se bem claramente a composição de um tetrâmero.

Quando superpostos o modelo de apo-NahI com cada um dos variantes NahI-F448S e L156G-F448S, rearranjos conformacionais da cadeia carbônica induzidos pelas mutações não são perceptíveis, observação confirmada pelos baixos valores de *R.m.s.D.* iguais a 0,403 Å e 0,421 Å, respectivamente.

Todavia, no duplo mutante NahI-L156G-F448S, uma sequência de doze resíduos de aminoácidos compreendida por Leu98 e Ala109 apresenta *R.m.s.D.* igual a 1,109 Å. Curiosamente, tal segmento, que parece sofrer uma retração quando comparado ao equivalente em NahI, está situado na α -hélice adjacente à entrada do sítio catalítico (Figura 42 A). Os resíduos que mais se deslocam são Leu98 (1,117 Å), His101 (1,364 Å), Ile102 (1,638 Å), Arg106 (1,235 Å) e Ala109 (1,456 Å) (Figura 42 C) contribuindo para a expansão do bolso catalítico cujo volume agora abrange 1.336 Å³, conforme calculado pelo servidor CASTp (Binkowski et al 2003).

Em NahI-F448S, o afastamento de tal α -hélice é menos pronunciado, dado o valor de *R.m.s.D.* igual a 0,709 Å para o mesmo segmento, quando comparado ao correspondente em NahI (Figura 42 B). Enquanto Arg106 (1,033 Å) apresenta o maior deslocamento, Leu98 (0,564 Å) e Ile102 (0,366 Å) praticamente mantêm a mesma posição equivalente em NahI (Figura 42 C) e o volume do sítio atinge 883 Å³.

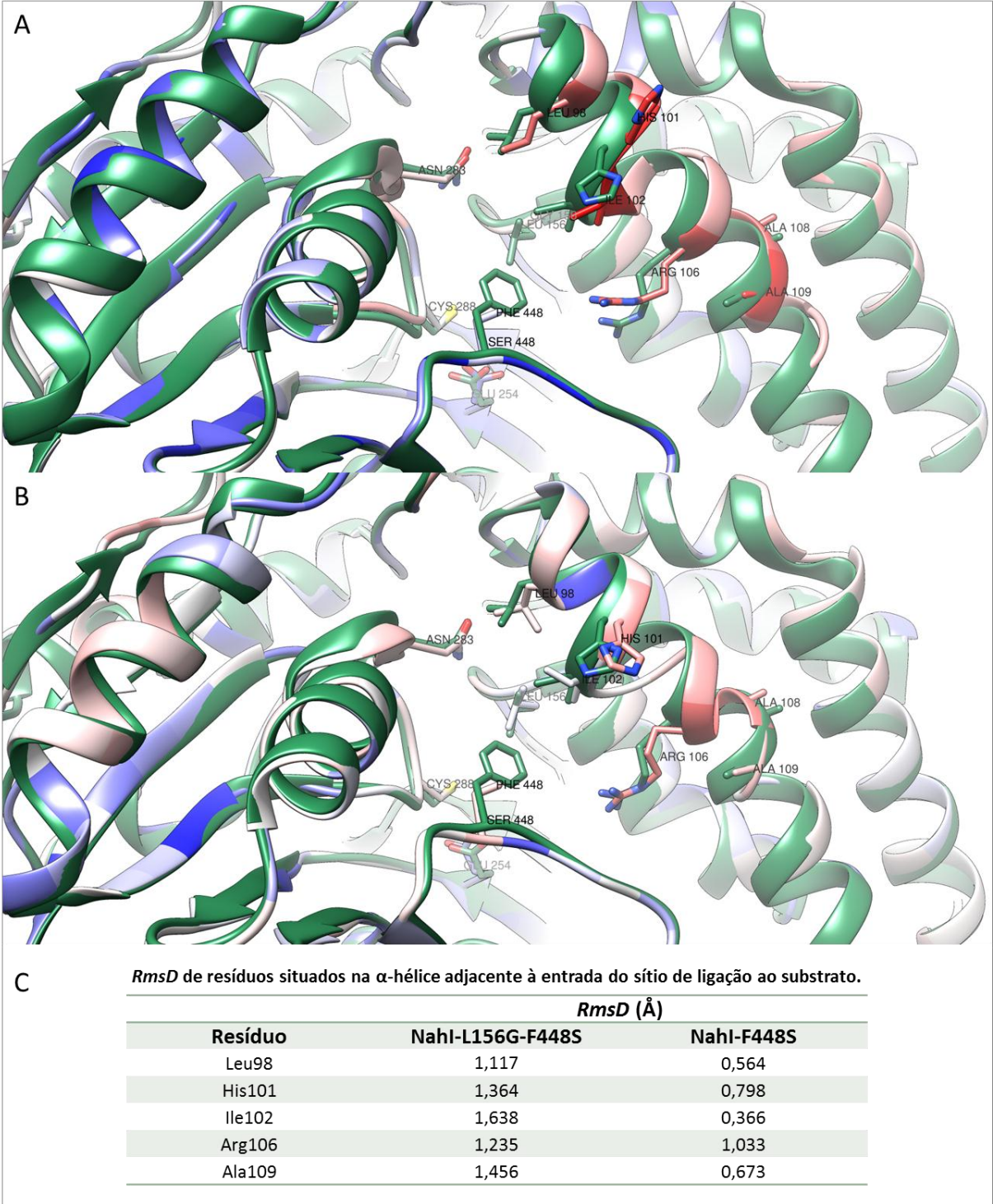


Figura 42: Entrada do sítio catalítico de NahI e de dois mutantes. **A.** No duplo mutante NahI-L156G-F448S, uma sequência de doze resíduos de aminoácidos compreendida por Leu98 e Ala109 apresenta *R.m.s.D.* igual a 1,109 Å em comparação a NahI (em verde). O deslocamento da α -hélice composta por este segmento contribui para a expansão do bolso catalítico. NahI-L156G-F448S está colorida segundo *R.m.s.D.* mínimo (azul) a máximo (vermelho). **B.** Em NahI-F448S, o afastamento de tal α -hélice é menos pronunciado, dado o valor de *R.m.s.D.* igual a 0,709 Å para o mesmo segmento. **C.** Valores de *R.m.s.D.* de alguns resíduos situados na α -hélice em questão.

Tais observações sugerem que Leu156 e Phe448, com suas cadeias laterais volumosas, não somente ocupam a cavidade catalítica reduzindo o seu volume e bloqueando a passagem de substratos aromáticos, mas mantêm a integridade da abertura do bolsão ao aproximar a α -hélice circunvizinha ao sítio que, de certo modo, acaba por limitar o espaço da cavidade. Ao que parece, em ambos os casos, a presença destes dois aminoácidos, Leu156 e Phe448, favorece a construção de um sítio mais fechado capaz de melhor acomodar substratos alifáticos.

Análise semelhante deverá ser feita ainda para NahI-L156G. É de se supor que a presença de Phe448, a despeito da ausência de Leu156, apresente resultado semelhante ao de NahI-F448S no sentido de manter mais próxima a α -hélice adjacente ao sítio limitando assim a sua abertura. Tanto por isto quanto pelo próprio bloqueio que Phe448 representa na entrada do bolsão, é de se esperar que o volume da cavidade seja inferior ao obtido para os outros dois mutantes.

Talvez não por mera coincidência, novamente vemos aqui a participação da α -hélice circunvizinha ao sítio, já mencionada antes ao ser analisada a sequência de 21 resíduos de aminoácidos consecutivos em que NahI e XylG se diferem mais pronunciadamente. Ao contrário dos mutantes propostos para NahI, os resíduos L156 e F448 estão conservados em XylG, ainda que esta apresente atividade sobre aldeídos aromáticos, volumosos. Pode ser que as modificações ocorridas na sequência desta α -hélice tenham promovido novas interações com resíduos adjacentes resultando em uma abertura da entrada do sítio através do afastamento da α -hélice em questão.

4.10. ATIVIDADE CINÉTICA DAS VARIANTES DE NahI

Segundo a hipótese inicialmente sugerida, a expansão do sítio catalítico de NahI, dada pela substituição de Leu156 e Phe448 por seus correspondentes em NahF, poderia favorecer a atividade da enzima sobre substratos aromáticos. Entretanto, ao contrário do que se esperava e de forma semelhante à enzima NahI, os três mutantes propostos não apresentaram atividade sobre o substrato natural de NahF, o 2-salicilaldeído, o que talvez possa ser atribuído à posição adjacente dos grupos substituintes. Assim como 2- e 3-carboxibenzaldeído são substratos fracos para XylG, dada a preferência ao 4-carboxibenzaldeído até mesmo sobre 2-HMS (Wang 2003), pode ser que o posicionamento *para* dos substituintes apresente melhor

alinhamento para a ligação do salicilaldeído quando comparado às posições *orto* e *meta*, que não possibilitariam um encaixe correto no sítio. Em NahF, contudo, observa-se que mudanças na posição da hidroxila em 3- e 4-salicilaldeído levam à diminuição dos valores de k_{cat}/K_M em cerca de 6 e 15 vezes, respectivamente, em relação ao substrato natural 2-salicilaldeído. Dessa forma, é válido testar a atividade de NahI e seus mutantes sobre 4-salicilaldeído, por exemplo. Entretanto, falta-lhe ainda o grupo carboxilato, identificado por Wang (2003) como essencial para a ligação do substrato. Mais ainda, para fins de comparação com a homóloga XylG, pode-se avaliar a atividade cinética de NahI e dos mutantes sobre benzaldeído e seus análogos.

O fato de NahI não apresentar atividade sobre o substrato natural de NahF sugere que estas enzimas tenham evoluído para cumprir funções específicas nas vias de degradação às quais pertencem, contribuindo talvez para a regulação do fluxo metabólico entre ambas. A observação de uma molécula de salicilaldeído localizada em um sítio atípico, entre as duas subunidades que compõem o dímero de NahF, suscita uma possível modulação alostérica para esta enzima (Coitinho 2013). Também segundo Coitinho (2013), há indícios de que salicilato, produto da reação catalisada por NahF, cause inibição não competitiva da própria enzima que o produz. Cabe, portanto, monitorar a atividade catalítica de NahI sobre 2-HMS na presença de salicilaldeído e salicilato, separadamente, a fim de avaliar uma possível comunicação do fluxo metabólico entre as vias superior e inferior.

A respeito da oxidação de 2-HMS, apenas o mutante F448S apresentou atividade. O K_M elevou-se nove vezes ($11,4 \pm 2,0 \mu\text{M}$) enquanto o valor de k_{cat} se manteve comparável ao de NahI ($0,9 \text{ s}^{-1}$), resultando em um decréscimo de aproximadamente nove vezes no valor de k_{cat}/K_M ($0,075 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$). Mais uma vez, a acessibilidade ao sítio ativo parece influenciar a ligação ao substrato, já que o sutil alargamento da entrada do sítio catalítico em F448S foi capaz de afetar ligeiramente a afinidade da enzima ao seu substrato natural, sem perturbar, contudo, o número de renovação (*turnover*). Esta última observação, somada ao fato de que os outros dois mutantes L156G e L156G-F448S não apresentaram qualquer atividade sobre 2-HMS, sugere que a manutenção de Leu156 parece ser crítica para a oxidação do substrato por NahI. Ainda que não tome parte diretamente na catálise, é provável que este aminoácido favoreça o melhor posicionamento de 2-HMS no sítio catalítico para que ocorra a reação.

De fato, em PnpE, uma γ -hidroximuconato semialdeído desidrogenase envolvida na via de degradação de *para*-nitrofenol em *Pseudomonas* sp. WBC-3, a mutação F150A, em

posição equivalente à Leu156 de NahI, comprometeu completamente a atividade da enzima sobre o substrato γ -hidroximuconato semialdeído (Su et al 2013). Segundo um alinhamento de sequências consenso das famílias de ALDHs (Perozich et al 1999), tal resíduo de leucina em NahI (Leu156) é representativo da família HMSALDH (hidroximuconato semialdeído desidrogenase), enquanto um resíduo de fenilalanina ou tirosina em posição correspondente é observado em todas as outras famílias, inclusive entre as ALDHs aromáticas. Exceção feita, contudo, para NahF. Uma vez que o resíduo de glicina possui a menor cadeia lateral dentre os aminoácidos, é provável que Gly150 de NahF favoreça a formação de um sítio catalítico mais amplo, capaz de acomodar aldeídos aromáticos estruturalmente diversos.

Quanto ao resíduo W163 de NahI, o equivalente W157 em PnpE parece contribuir para uma região hidrofóbica no sítio capaz de melhor acomodar uma pequena porção alifática da cadeia linear do substrato. Ainda que este resíduo não estabeleça contato direto com o γ -hidroximuconato semialdeído, a mutação W157A reduziu em 20% a atividade catalítica da enzima (Su et al 2013).

Tem-se observado que um determinado resíduo de asparagina, equivalente a Asn155 de NahI e conservado em 95% das ALDHs segundo (Perozich et al 1999), desempenha papel importante na catálise. Para a mutação N149D, PnpE perdeu sua atividade sobre γ -hidroximuconato semialdeído (Su et al 2013). De semelhante modo, perda de atividade também foi observada para o mutante N133A de Gox0499, uma ALDH da bactéria *Gluconobacter oxydans* com atividade sobre aldeídos aromáticos e alifáticos de cadeia longa (Yuan et al 2013). Tal resíduo de asparagina, que em NahI (Asn155) estabelece ligação de hidrogênio com o anel de nicotinamida do cofator (Figura 26, subseção 4.5), se posiciona na interface dos sítios de ligação ao substrato e ao cofator.

Assim, embora um mecanismo geral para ALDHs tenha sido proposto com base em análises de identidade de sequência e estudos cristalográficos (Wymore et al 2004), e os resíduos conservados de cisteína, glutamato e asparagina (respectivamente, C288, E254 e N155 em NahI) sejam considerados essenciais para a catálise, cabe descrever o possível mecanismo catalítico específico de NahI. A elucidação de mais estruturas da enzima complexada a outros ligantes é um auxílio na identificação de resíduos conservados ou não, necessários para a catálise e a especificidade ao substrato.

5. CONCLUSÕES E PERSPECTIVAS

Ainda que membros da superfamília ALDH apresentem um enovelamento estrutural comum, a elucidação cada vez mais ampla de estruturas tridimensionais destas enzimas tem permitido inferir distinções nos sítios catalíticos responsáveis por mudanças sistemáticas em suas propriedades cinéticas sobre uma variedade de substratos aldeídos. Assim, o acúmulo de conhecimento sobre os determinantes estruturais que tornam as ALDHs únicas em suas reações pode possibilitar o aprimoramento de características bioquímicas e cinéticas visando a possíveis aplicações biotecnológicas, como por exemplo, para fins de biorremediação.

Este trabalho, portanto, relata a caracterização estrutural e cinética, posto que esta em caráter parcial, de NahI, uma 2-hidroxiomuconato semialdeído desidrogenase pertencente à via inferior de degradação do naftaleno em *Pseudomonas putida* G7.

Conforme avaliado por espectroscopia de dicroísmo circular, o conteúdo de estrutura secundária de NahI mostra-se semelhante ao de 2-aminomuconato 6-semialdeído (2-AMS) desidrogenase de *Pseudomonas fluorescens* (PDB id 4I1W), um membro da superfamília ALDH que compartilha 61% de identidade de sequência com NahI. Embora alterações conformacionais significativas não sejam perceptíveis, dada a ligação ao NAD^+ , sua presença parece aumentar a estabilidade do complexo devido à manutenção de fitas- β do domínio de ligação ao cofator, mais provavelmente do motivo de *Rossmann* durante a desnaturação térmica. Além disso, sugere-se que NahI siga um processo de desnaturação em duas etapas. Quando complexada ao NAD^+ , a segunda temperatura de transição é maior, o que reforça a ideia de que o domínio de ligação ao cofator é menos suscetível à desnaturação térmica.

O modelo cristalográfico proposto para NahI apresenta o enovelamento α/β típico da superfamília das ALDHs, que se organiza em: domínio de oligomerização e domínios de ligação ao dinucleotídeo e ao substrato, este último também conhecido como catalítico. Além disso, operações de simetria cristalográficas revelam um arranjo tetramérico, consistente com o estado oligomérico inferido por SAXS. Rearranjos conformacionais induzidos pela ligação ao cofator não são perceptíveis quando superpostos os modelos de NahI com e sem NAD^+ , dados estes que apoiam aqueles observados por CD, em que os espectros obtidos a 20 °C, na presença ou na ausência do cofator, são muito semelhantes um do outro ao longo de toda a gama de comprimento de onda medido.

Ensaio de cinética confirmam a preferência de NahI por 2-hidroxiuconato semialdeído, todavia nenhuma atividade sobre salicilaldeído foi observada, ao contrário da homóloga XylG, uma 2-HMSD capaz de oxidar também substratos aromáticos. Sugere-se que a razão para tal comportamento esteja associada a um segmento de 21 resíduos de aminoácidos, notavelmente distinto entre ambas as enzimas e localizado em uma α -hélice adjacente ao sítio catalítico. Um alinhamento das sequências nucleotídicas de *nahI* e *xylG* revela um desvio de enquadramento da janela de leitura do segmento de 63 pares de bases correspondente à sequência de 21 resíduos de aminoácidos, o que pode ter conduzido a uma mudança dramática nas propriedades catalíticas de XylG.

Um estudo comparativo entre as estruturas tridimensionais de NahI e NahF, uma salicilaldeído desidrogenase com ampla atividade sobre aldeídos aromáticos e alifáticos de cadeia longa, atuante na via de degradação superior do naftaleno em *P. Putida* G7, permite indicar resíduos de aminoácidos que tenham participação na especificidade destas enzimas aos seus substratos. Sugere-se, para NahI, que a presença de resíduos com cadeia lateral volumosa no sítio catalítico inviabilize a acomodação de substratos aromáticos. Corrobora para isto a predição da expansão da cavidade catalítica, de 167 Å³ para 725 Å³, quando tão somente L156 e F448 são substituídos pelos resíduos equivalentes em NahF. Por esta razão, os três mutantes NahI-L156G / F448S / L156G-F448S foram propostos a fim de se compreender os determinantes estruturais de NahI para a seleção do substrato.

A estrutura tridimensional de tais mutantes revela um ligeiro afastamento da α -hélice justaposta à entrada do sítio catalítico com consequente expansão do volume da cavidade, superando os valores preditos anteriormente. Estas observações insinuam que L156 e F448, com suas cadeias laterais volumosas, não somente ocupam o sítio catalítico reduzindo o seu volume e bloqueando a passagem de substratos aromáticos, mas mantêm a integridade da abertura do bolsão ao aproximar a α -hélice circunvizinha que, de certo modo, acaba por limitar também o espaço da cavidade.

Ainda assim, o aumento do volume do sítio não proporciona incrementos da atividade catalítica sobre o aldeído aromático avaliado, 2-salicilaldeído. Ao contrário do que se esperava ao serem propostas as mutações em NahI, suas três variantes não apresentam atividade sobre o substrato natural de NahF, o que pode ser devido ao posicionamento dos grupos substituintes e dos resíduos que tomam parte na catálise.

Quanto à cinética dos mutantes sobre o substrato natural de NahI, 2-HMS, a atividade de NahI-F448S, a única observada, é reduzida em função do aumento de K_M , o que fomenta o papel crítico de L156 na especificidade da enzima ao seu substrato.

Os resultados desta investigação apresentam, pela primeira vez, características estruturais e cinéticas em relação à enzima NahI e abrem caminho para futuros estudos sobre a identidade e a função dos resíduos cruciais para a catálise e a descrição de como a reação se processa. A cristalização de NahI na presença de um inibidor competitivo torna favorável esclarecer estas questões. Cabe avaliar também a atividade cinética de NahI e mutantes sobre uma variedade de substratos aldeídos, bem como sobre seu substrato natural 2-HMS na presença de salicilaldeído e salicilato para verificar um possível papel regulatório destes últimos. Além do mais, elucidar a estrutura tridimensional da homóloga XylG possibilitará uma análise comparativa mais apurada quanto aos determinantes estruturais que viabilizam a oxidação de aldeídos aromáticos por esta ALDH.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213-21
- Agilent Technologies I. 2010. ArcticExpress Competent Cells and ArcticExpress (DE3) Competent Cells - Instruction Manual. ed. 230191-12
- Ahuja SK, Ferreira GM, Moreira AR. 2004. Utilization of enzymes for environmental applications. *Crit Rev Biotechnol* 24:125-54
- Ahvazi B, Coulombe R, Delarge M, Vedadi M, Zhang L, et al. 2000. Crystal structure of the NADP(+)-dependent aldehyde dehydrogenase from *Vibrio harveyi*: structural implications for cofactor specificity and affinity. *Biochemical Journal* 349:853-61
- Alcalde M, Ferrer M, Plou FJ, Ballesteros A. 2006. Environmental biocatalysis: from remediation with enzymes to novel green processes. *Trends Biotechnol* 24:281-7
- Andrade JA, Augusto F, Jardim ICSF. 2010. Bioremediação de solos contaminados por petróleo e seus derivados. *Eclética Química* 35:17 - 43
- Araujo SS, Neves CM, Guimaraes SL, Whitman CP, Johnson WH, Jr., et al. 2015. Structural and kinetic characterization of recombinant 2-hydroxymuconate semialdehyde dehydrogenase from *Pseudomonas putida* G7. *Arch Biochem Biophys* 579:8-17
- Atlas RM, Hazen TC. 2011. Oil biodegradation and bioremediation: a tale of the two worst spills in U.S. history. *Environ Sci Technol* 45:6709-15
- Baird WM, Hooven LA, Mahadevan B. 2005. Carcinogenic polycyclic aromatic hydrocarbon-DNA adducts and mechanism of action. *Environ Mol Mutagen* 45:106-14
- Bamforth SM, Singleton I. 2005. Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions. *J Chem Technol Biot* 80:723-36
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. 2000. The Protein Data Bank. *Nucleic Acids Res* 28:235-42
- Binkowski TA, Naghibzadeh S, Liang J. 2003. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res* 31:3352-5
- Bohm G, Muhr R, Jaenicke R. 1992. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng* 5:191-5
- Bordelon T, Montegudo SK, Pakhomova S, Oldham ML, Newcomer ME. 2004. A disorder to order transition accompanies catalysis in retinaldehyde dehydrogenase type II. *Journal of Biological Chemistry* 279:43085-91
- Bosch R, Garcia-Valdes E, Moore ER. 2000. Complete nucleotide sequence and evolutionary significance of a chromosomally encoded naphthalene-degradation lower pathway from *Pseudomonas stutzeri* AN10. *Gene* 245:65-74

- Brunger AT. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355:472-5
- Cao B, Geng A, Loh KC. 2008. Induction of ortho- and meta-cleavage pathways in *Pseudomonas* in biodegradation of high benzoate concentration: MS identification of catabolic enzymes. *Appl Microbiol Biotechnol* 81:99-107
- Cao B, Nagarajan K, Loh KC. 2009. Biodegradation of aromatic compounds: current status and opportunities for biomolecular approaches. *Appl Microbiol Biotechnol* 85:207-28
- Carneiro FR, Silva TC, Alves AC, Haline-Vaz T, Gozzo FC, Zanchin NI. 2006. Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1. *Biochem Biophys Res Commun* 343:260-8
- Cerniglia CE. 1984. Microbial metabolism of polycyclic aromatic hydrocarbons. *Adv Appl Microbiol* 30:31-71
- Cobessi D, Tete-Favier F, Marchal S, Azza S, Branlant G, Aubry A. 1999. Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. *Journal of Molecular Biology* 290:161-73
- Coitinho JB. 2013. Cristalização, estrutura e atividade de duas proteínas com apelo biotecnológico: Pb27 de *Paracoccidioides brasiliensis* e salicilaldeído desidrogenase (NahF) de *Pseudomonas putida*. Tese. Universidade Federal de Minas Gerais, Belo Horizonte. 156 pp.
- Coitinho JB, Costa DM, Guimaraes SL, de Goes AM, Nagem RA. 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* 68:93-7
- Cowtan K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62:1002-11
- Das N, Chandran P. 2011. Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol Res Int* 2011:941810
- Davies JJ, Evans WC. 1964. Oxidative metabolism of naphthalene by soil pseudomonads. The ring-fission mechanism. *Biochem J* 91:251-61
- Doniach S. 2001. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem Rev* 101:1763-78
- Dunn NW, Gunsalus IC. 1973. Transmissible plasmid coding early enzymes of naphthalene oxidation in *Pseudomonas putida*. *J Bacteriol* 114:974-9
- Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486-501
- Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica Section A - Foundations of Crystallography* 47:392-400

- Ferré-D'amaré AR, Burley SK. 1994. Use of Dynamic Light-Scattering to Assess Crystallizability of Macromolecules and Macromolecular Assemblies (Vol 2, Pg 357, 1994). *Structure* 2:567-
- Ferré-D'amaré AR, Burley SK. 1997. Dynamic light scattering in evaluating crystallizability of macromolecules. *Method Enzymol* 276:157-66
- Fischer H, Neto MD, Napolitano HB, Polikarpov I, Craievich AF. 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* 43:101-9
- Frishman D, Argos P. 1995. Knowledge-based protein secondary structure assignment. *Proteins* 23:566-79
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784-8
- González-Segura L, Rudiño-Piñera E, Muñoz-Clares RA, Horjales E. 2009. The crystal structure of a ternary complex of betaine aldehyde dehydrogenase from *Pseudomonas aeruginosa* provides new insight into the reaction mechanism and shows a novel binding mode of the 2'-phosphate of NADP⁺ and a novel cation binding site. *J. Mol. Biol.* 385:542–57
- Grimm AC, Harwood CS. 1999. NahY, a catabolic plasmid-encoded receptor required for chemotaxis of *Pseudomonas putida* to the aromatic hydrocarbon naphthalene. *J Bacteriol* 181:3310-6
- Habe H, Omori T. 2003. Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. *Biosci Biotechnol Biochem* 67:225-43
- Hammersley AP, Svensson SO, Hanfland M, Fitch AN, Hausermann D. 1996. Two-dimensional detector software: From real detector to idealised image or two-theta scan. *High Pressure Res* 14:235-48
- Harayama S, Mermod N, Rekik M, Lehrbach PR, Timmis KN. 1987a. Roles of the divergent branches of the *meta*-cleavage pathway in the degradation of benzoate and substituted benzoates. *J Bacteriol* 169:558-64
- Harayama S, Rekik M. 1989. Bacterial aromatic ring-cleavage enzymes are classified into two different gene families. *J Biol Chem* 264:15328-33
- Harayama S, Rekik M. 1993. Comparison of the nucleotide sequences of the *meta*-cleavage pathway genes of TOL plasmid pWWO from *Pseudomonas putida* with other *meta*-cleavage genes suggests that both single and multiple nucleotide substitutions contribute to enzyme evolution. *Molecular and General Genetics* 239:81-9
- Harayama S, Rekik M, Wasserfallen A, Bairoch A. 1987b. Evolutionary relationships between catabolic pathways for aromatics: Conservation of gene order and nucleotide sequences of catechol oxidation genes of pWWO and NAH7 plasmids. *Mol Gen Genet* 210:241-7
- Harding SE. 1994. Determination of diffusion coefficients of biological macromolecules by dynamic light scattering. *Methods Mol Biol* 22:97-108

- Haritash AK, Kaushik CP. 2009. Biodegradation aspects of polycyclic aromatic hydrocarbons (PAHs): a review. *J Hazard Mater* 169:1-15
- He Z, Davis JK, Spain JC. 1998. Purification, characterization, and sequence analysis of 2-aminomuconic 6-semialdehyde dehydrogenase from *Pseudomonas pseudoalcaligenes* JS45. *J Bacteriol* 180:4591-5
- Heinig M, Frishman D. 2004. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* 32:W500-2
- Hempel J, Nicholas H, Lindahl R. 1993. Aldehyde dehydrogenases: widespread structural and functional diversity within a shared framework. *Protein Sci* 2:1890-900
- Ho KK, Hurley TD, Weiner H. 2006. Selective alteration of the rate-limiting step in cytosolic aldehyde dehydrogenase through random mutagenesis. *Biochemistry* 45:9445-53
- Hugo N, Armengaud J, Gaillard J, Timmis KN, Jouanneau Y. 1998. A novel [2Fe-2S] ferredoxin from *Pseudomonas putida* mt2 promotes the reductive reactivation of catechol 2,3-dioxygenase. *Journal of Biological Chemistry* 273:9622-9
- Hugo N, Meyer C, Armengaud J, Gaillard J, Timmis KN, Jouanneau Y. 2000. Characterization of three XylT-like [2Fe-2S] ferredoxins associated with catabolism of cresols or naphthalene: Evidence for their involvement in catechol dioxygenase reactivation. *Journal of Bacteriology* 182:5580-5
- Huo L, Davis I, Liu F, Andi B, Esaki S, et al. 2015. Crystallographic and spectroscopic snapshots reveal a dehydrogenase in action. *Nat Commun* 6:5935
- Inoue J, Shaw JP, Rekik M, Harayama S. 1995. Overlapping substrate specificities of benzaldehyde dehydrogenase (the xylC gene product) and 2-hydroxymuconic semialdehyde dehydrogenase (the xylG gene product) encoded by TOL plasmid pWW0 of *Pseudomonas putida*. *J Bacteriol* 177:1196-201
- Jacques DA, Guss JM, Svergun DI, Trehwella J. 2012. Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta Crystallogr D* 68:620-6
- Jacques RJS, Bento FM, Camargo FAO. 2007. Biodegradação de hidrocarbonetos aromáticos policíclicos. *Ciência e Natura* 29:7 - 24
- Kanally RA, Harayama S. 2000. Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *J Bacteriol* 182:2059-67
- Lamb AL, Newcomer ME. 1999. The structure of retinal dehydrogenase type II at 2.7 Å resolution: implications for retinal specificity. *Biochemistry* 38:6003-11
- Lindahl R. 1992. Aldehyde dehydrogenases and their role in carcinogenesis. *Crit Rev Biochem Mol Biol* 27:283-335
- Liu ZJ, Sun YJ, Rose J, Chung YJ, Hsiao CD, et al. 1997. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat Struct Biol* 4:317-26

- Lorber B, Fischer F, Bailly M, Roy H, Kern D. 2012. Protein analysis by dynamic light scattering: Methods and techniques for students. *Biochem Mol Biol Edu* 40:372-82
- Luo M, Singh RK, Tanner JJ. 2013. Structural Determinants of Oligomerization of Delta(1)-Pyrroline-5-Carboxylate Dehydrogenase: Identification of a Hexamerization Hot Spot. *Journal of Molecular Biology* 425:3106-20
- Maddocks SE, Oyston PC. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* 154:3609-23
- Matte A, Cygler M. 2007. Using dynamic light scattering to improve protein solution behavior for crystallization. *Am Lab* 39:35-7
- Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol* 33:491-7
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* 40:658-74
- Murray K, Duggleby CJ, Sala-Trepat JM, Williams PA. 1972. The metabolism of benzoate and methylbenzoates via the meta-cleavage pathway by *Pseudomonas arvilla* mt-2. *Eur J Biochem* 28:301-10
- Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, et al. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355-67
- Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-55
- Netto APD, Moreira JC, Dias AEXO, Arbilla G, Ferreira LFV, et al. 2000. Avaliação da contaminação humana por hidrocarbonetos policíclicos aromáticos (HPAs) e seus derivados nitrados (NHPAs): uma revisão metodológica *Quim Nova* 23:765 - 73
- Neves CML. 2013. Produção de enzimas recombinantes da via de degradação do naftaleno para caracterização estrutural. Dissertação. Universidade Federal de Minas Gerais, Belo Horizonte. 70 pp.
- Ni L, Sheikh S, Weiner H. 1997. Involvement of glutamate 399 and lysine 192 in the mechanism of human liver mitochondrial aldehyde dehydrogenase. *J Biol Chem* 272:18823-6
- Niesen FH, Berglund H, Vedadi M. 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2:2212 - 21
- Otwinowski Z, Minor W. 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology*, ed. CW Carter, RM Sweet, pp. 307 - 26. New York: Academic Press
- Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. 1995. How to measure and predict the molar absorption coefficient of a protein *Protein Science* 4:2411-23
- Paul D, Pandey G, Pandey J, Jain RK. 2005. Accessing microbial diversity for bioremediation and environmental restoration. *Trends Biotechnol* 23:135-42

- Peixoto RS, Vermelho AB, Rosado AS. 2011. Petroleum-degrading enzymes: bioremediation and new prospects. *Enzyme Res* 2011:475193
- Pemberton TA, Srivastava D, Sanyal N, Henzl MT, Becker DF, Tanner JJ. 2014. Structural Studies of Yeast Delta(1)-Pyrroline-5-carboxylate Dehydrogenase (ALDH4A1): Active Site Flexibility and Oligomeric State. *Biochemistry* 53:1350-9
- Peng RH, Xiong AS, Xue Y, Fu XY, Gao F, et al. 2008. Microbial biodegradation of polyaromatic hydrocarbons. *FEMS Microbiol Rev* 32:927-55
- Perez-Miller SJ, Hurley TD. 2003. Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. *Biochemistry* 42:7100-9
- Perozich J, Kuo I, Lindahl R, Hempel J. 2001. Coenzyme specificity in aldehyde dehydrogenase. *Chem Biol Interact* 130-132:115-24
- Perozich J, Kuo I, Wang BC, Boesch JS, Lindahl R, Hempel J. 2000. Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase. *Eur J Biochem* 267:6197-203
- Perozich J, Nicholas H, Wang BC, Lindahl R, Hempel J. 1999. Relationships within the aldehyde dehydrogenase extended family. *Protein Sci* 8:137-46
- Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, et al. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography* 45:342-50
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-12
- Porod G. 1982. General theory. In Small angle X-ray scattering, edited by O. Glatter & O. Kratky. *Academic Press – London*:17-51
- Purcaro G, Moret S, Conte LS. 2013. Overview on polycyclic aromatic hydrocarbons: occurrence, legislation and innovative determination in foods. *Talanta* 105:292-305
- Rambo RP, Tainer JA. 2011. Characterizing Flexible and Intrinsically Unstructured Biological Macromolecules by SAS Using the Porod-Debye Law. *Biopolymers* 95:559-71
- Rhodes G. 2000. Crystallography Made Crystal Clear - A Guide for Users of Macromolecular Models. ed. A Press
- Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320-4
- Rossmann MG, Moras D, Olsen KW. 1974. Chemical and biological evolution of nucleotide-binding protein. *Nature* 250:194-9
- Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725-38

- Rubin H. 2001. Synergistic mechanisms in carcinogenesis by polycyclic aromatic hydrocarbons and by tobacco smoke: a bio-historical perspective with updates. *Carcinogenesis* 22:1903-30
- Saeed M, Higginbotham S, Gaikwad N, Chakravarti D, Rogan E, Cavalieri E. 2009. Depurinating naphthalene-DNA adducts in mouse skin related to cancer initiation. *Free Radic Biol Med* 47:1075-81
- Sala-Trepat JM, Murray K, Williams PA. 1972. The metabolic divergence in the meta cleavage of catechols by *Pseudomonas putida* NCIB 10015. Physiological significance and evolutionary implications. *Eur J Biochem* 28:347-56
- Samanta SK, Singh OV, Jain RK. 2002. Polycyclic aromatic hydrocarbons: environmental pollution and bioremediation. *Trends Biotechnol* 20:243-8
- Schneidman-Duhovny D, Hammel M, Sali A. 2010. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* 38:W540-4
- Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. 2013. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J* 105:962-74
- Semisotnov GV, Kihara H, Kotova NV, Kimura K, Amemiya Y, et al. 1996. Protein globularization during folding. A study by synchrotron small-angle X-ray scattering. *Journal of Molecular Biology* 262:559-74
- Seo JS, Keum YS, Li QX. 2009. Bacterial degradation of aromatic compounds. *Int J Environ Res Public Health* 6:278-309
- Sophos NA, Vasiliou V. 2003. Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem Biol Interact* 143-144:5-22
- Sota M, Yano H, Ono A, Miyazaki R, Ishii H, 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* 188:4057-67
- Stansbury KH, Flesher JW, Gupta RC. 1994. Mechanism of aralkyl-DNA adduct formation from benzo[a]pyrene in vivo. *Chem Res Toxicol* 7:254-9
- Steinmetz CG, Xie P, Weiner H, Hurley TD. 1997. Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure* 5:701-11
- Su J, Zhang C, Zhang JJ, Wei T, Zhu D, et al. 2013. Crystal structure of the gamma-hydroxymuconic semialdehyde dehydrogenase from *Pseudomonas* sp. strain WBC-3, a key enzyme involved in *para*-Nitrophenol degradation. *BMC Struct Biol* 13:30
- Sutherland TD, Horne I, Weir KM, Coppin CW, Williams MR, et al. 2004. Enzymatic bioremediation: from enzyme discovery to applications. *Clin Exp Pharmacol Physiol* 31:817-21
- Svergun D, Barberato C, Koch MHJ. 1995. CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* 28:768-73

- Svergun DI. 1992. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography* 25:495-503
- Taylor RG, Walker DC, McInnes RR. 1993. E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res* 21:1677-8
- Tiburtius ERL, Peralta-Zamora P, Leal ES. 2004. Contamination of waters by BTXs and processes used in the remediation of contaminated sites. *Quim Nova* 27:441-6
- Wallace AC, Laskowski RA, Thornton JM. 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8:127-34
- Wang SC. 2003. Studies of bacterial catabolic enzymes: implications for the evolution of enzymes and metabolic pathways. Tese. University of Texas, Austin. 258 pp.
- Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, et al. 1999. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531-52
- Williams PA, Sayers JR. 1994. The evolution of pathways for aromatic hydrocarbon oxidation in *Pseudomonas*. *Biodegradation* 5:195-217
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235-42
- Wymore T, Hempel J, Cho SS, Mackerell AD, Jr., Nicholas HB, Jr., Deerfield DW, 2nd. 2004. Molecular recognition of aldehydes by aldehyde dehydrogenase and mechanism of nucleophile activation. *Proteins* 57:758-71
- Yen KM, Gunsalus IC. 1982. Plasmid gene organization: naphthalene/salicylate oxidation. *Proc Natl Acad Sci U S A* 79:874-8
- Yen KM, Gunsalus IC. 1985. Regulation of naphthalene catabolic genes of plasmid NAH7. *J Bacteriol* 162:1008-13
- Yuan ZN, Yin B, Wei DZ, Yuan YRA. 2013. Structural basis for cofactor and substrate selection by cyanobacterium succinic semialdehyde dehydrogenase. *J Struct Biol* 182:125-35
- Zhang L, Ahvazi B, Szittner R, Vrielink A, Meighen E. 1999. Change of nucleotide specificity and enhancement of catalytic efficiency in single point mutants of *Vibrio harveyi* aldehyde dehydrogenase. *Biochemistry* 38:11440-7



Structural and kinetic characterization of recombinant 2-hydroxymuconate semialdehyde dehydrogenase from *Pseudomonas putida* G7

Simara Semíramis de Araújo^{a,1}, Cíntia Mara Leal Neves^{a,1}, Samuel Leite Guimarães^a, Christian P. Whitman^b, William H. Johnson Jr.^b, Ricardo Aparicio^c, Ronaldo Alves Pinto Nagem^{a,*}

^a Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, MG 31270-901, Brazil

^b Division of Medicinal Chemistry, College of Pharmacy, The University of Texas, Austin, TX 78712-1071, USA

^c Laboratório de Biologia Estrutural e Cristalografia, Instituto de Química, Universidade Estadual de Campinas, CP 6154, Campinas, SP 13083-970, Brazil

ARTICLE INFO

Article history:

Received 9 April 2015

and in revised form 21 May 2015

Available online 29 May 2015

Keywords:

Pseudomonas putida G7

Naphthalene degradation

2-Hydroxymuconate semialdehyde dehydrogenase

Structure

Kinetics

ABSTRACT

The first enzyme in the oxalocrotonate branch of the naphthalene-degradation lower pathway in *Pseudomonas putida* G7 is Nahl, a 2-hydroxymuconate semialdehyde dehydrogenase which converts 2-hydroxymuconate semialdehyde to 2-hydroxymuconate in the presence of NAD⁺. Nahl is in family 8 (ALDH8) of the NAD(P)⁺-dependent aldehyde dehydrogenase superfamily. In this work, we report the cloning, expression, purification and preliminary structural and kinetic characterization of the recombinant Nahl. The *nahl* gene was subcloned into a T7 expression vector and the enzyme was overexpressed in *Escherichia coli* ArcticExpress as a hexa-histidine-tagged fusion protein. After purification by affinity and size-exclusion chromatography, dynamic light scattering and small-angle X-ray scattering experiments were conducted to analyze the oligomeric state and the overall shape of the enzyme in solution. The protein is a tetramer in solution and has nearly perfect 222 point group symmetry. Protein stability and secondary structure content were evaluated by a circular dichroism spectroscopy assay under different thermal conditions. Furthermore, kinetic assays were conducted and, for the first time, K_M ($1.3 \pm 0.3 \mu\text{M}$) and k_{cat} (0.9 s^{-1}) values were determined at presumed NAD⁺ saturation. Nahl is highly specific for its biological substrate and has no activity with salicylaldehyde, another intermediate in the naphthalene-degradation pathway.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Polycyclic aromatic hydrocarbons (PAHs²) are a group of hydrophobic organic compounds consisting of two or more fused aromatic rings in linear or angular arrangements. They are widely distributed in the environment due to natural as well as anthropogenic sources [1]. Members of this group include naphthalene, anthracene, phenanthrene, pyrene, and benzo[a]pyrene, among others.

* Corresponding author. Fax: +55 31 3409 2614.

E-mail address: nagem@icb.ufmg.br (R.A.P. Nagem).

¹ SSA and CMLN equally contributed to this work.

² Abbreviations used: PAHs, polycyclic aromatic hydrocarbons; ALDH, aldehyde dehydrogenase; CDD, Conserved Domain Database; SF, superfamily; CD, circular dichroism; SEC, size-exclusion chromatography; DLS, dynamic light scattering; SAXS, small-angle X-ray scattering; PCR, polymerase chain reaction; TEV, tobacco etch virus; IPTG, isopropyl beta-D-thiogalactopyranoside; PDB, Protein Data Bank; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis.

Naphthalene is the simplest and most abundant PAH, and serves as a model compound for PAH studies. The toxicity of naphthalene has been well documented. It has been suggested that naphthalene undergoes metabolic activation to 1,2-naphthoquinone, which reacts with DNA and leads to the formation of depurinating adducts. The abasic sites in DNA formed by loss of the depurinating adducts can generate mutations that might initiate cancer [2].

Due to their ubiquitous occurrence, potential to bio-accumulate and carcinogenic activity, PAHs are significant environmental concerns [3]. The US Environmental Protection Agency has listed 16 PAHs as priority pollutants for remediation [4] highlighting the importance of their removal from the environment.

Owing to the relatively high cost and the non-specificity of conventional physical and chemical environmental decontamination methods, bioremediation is a promising strategy for environmental cleanup [5]. It is a safe, less disruptive and cost-effective treatment to remove toxic compounds from the environment [6]. In the past

few years, enzymatic bioremediation has resurfaced as an attractive alternative to augment currently available bio-treatment techniques. The process starts with a search for microorganisms capable of feeding on a particular pollutant, and then focuses on the identification of the enzymes responsible for degradation [7].

A wide phylogenetic diversity of bacteria such as species of the genus *Alcaligenes*, *Acinetobacter*, *Pseudomonas*, *Rhodococcus*, and *Nocardia* is capable of aerobic degradation of aromatic compounds into less toxic or non-toxic products [8]. *Pseudomonas* species and closely related organisms have been the most extensively studied due to their superior performance in degrading PAHs as sole carbon and energy sources [8,9].

The gram-negative bacterium *Pseudomonas putida* strain G7 harbors an 83 kilobase plasmid associated with naphthalene metabolism, NAH7 [10]. The naphthalene catabolic genes (*nah*) of NAH7 are organized into two operons. The *nah* operon (*nahAaAbAcAdBFCED*) encodes the upper pathway enzymes involved in the conversion of naphthalene to salicylate. The latter results from the oxidation of salicylaldehyde by the action of salicylaldehyde dehydrogenase (NahF), the last enzyme of the upper pathway [11]. On the other hand, the *sal* operon (*nahGTHINLOMKJ*) codes for the lower pathway enzymes involved in the conversion of salicylate to pyruvate and acetaldehyde [10,12]. The lower pathway starts with the oxidation of salicylate to catechol by salicylate hydroxylase (NahG), which is extradiol-cleaved by catechol-2,3-dioxygenase (NahH) and further transformed to pyruvate and acetyl-CoA by the remaining *meta*-cleavage pathway gene products (NahI, NahJ, NahK, NahN, NahL, NahM, and NahO) [13] (Fig. 1).

NahI from *P. putida* G7 (2-hydroxymuconate semialdehyde dehydrogenase – NCBI entry YP_534834.1; 486 amino acid residues and 51,517 Da), the first enzyme in the oxalocrotonate branch of the naphthalene-degradation lower pathway, belongs to the NAD(P)⁺-dependent aldehyde dehydrogenase (ALDH) superfamily, according to the Conserved Domain Database classification (CDD; [14]). The members of this superfamily (SF) catalyze the oxidation of a wide variety of endogenous and exogenous aliphatic and aromatic aldehydes to their corresponding carboxylic acids using NAD⁺ or NADP⁺ as a cofactor. They share the same scaffold and many conserved residues with important catalytic roles despite their overall low sequence identity (below 15%), modes of oligomerization and substrate specificity.

Within the ALDH-SF, NahI is further classified into the ALDH8 family together with different 2-hydroxymuconate semialdehyde dehydrogenases (2-HMSDs; e.g., XylG from *P. putida* mt-2 – NCBI entry P23105.1, AphC from *Comamonas testosteroni* TA441 – NCBI entry BAA88501.1, TomC from *Burkholderia cepacia* G4 – NCBI entry AAP32788.1) and other enzymes (e.g., human aldehyde dehydrogenase family 8 member A1 protein or ALDH8A1 – NCBI entry Q9H2A2.1, 5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase from *Escherichia coli* C – NCBI entry CAA86041.1). ALDH8 members share an overall sequence identity above 30% and, as reported in the literature, they catalyze reactions with different substrates. While NahI is thought to participate only in the conversion of 2-hydroxymuconate semialdehyde to 2-hydroxymuconate using NAD⁺ as a cofactor (EC 1.2.1.85; [15,16], XylG, another 2-HMSD involved in toluene and xylene degradation, oxidizes not only 2-hydroxymuconate semialdehyde, but also aromatic substrates such as benzaldehyde and its analogs [17]. Furthermore, human ALDH8A1, another ALDH8 member, converts 9-*cis*-retinal to 9-*cis*-retinoic acid [18]. Therefore, it is likely that subtle differences in the substrate pocket of these enzymes are responsible for their diverse kinetic properties using a variety of aldehyde substrates [19].

In this work, we report the cloning and expression of recombinant NahI from *P. putida* G7 (hexa histidine-tag NahI; 6×His-NahI). In order to characterize the secondary and quaternary structures of this enzyme in solution, we purified the recombinant protein and analyzed it by circular dichroism (CD), size-exclusion chromatography (SEC), dynamic light scattering (DLS) and small-angle X-ray scattering (SAXS). Furthermore, as a preliminary step to better understand the specificity of this enzyme, we performed kinetic assays with its biological substrate (2-hydroxymuconate semialdehyde) and salicylaldehyde, another intermediate in the pathway. Taken together, these findings validate the NahI biological unit and substrate specificity, reinforcing its precise role in the naphthalene-degradation pathway.

Materials and methods

Cloning

The coding region for 2-hydroxymuconate semialdehyde dehydrogenase (NahI, NCBI Gene ID: 3974218) was amplified from the

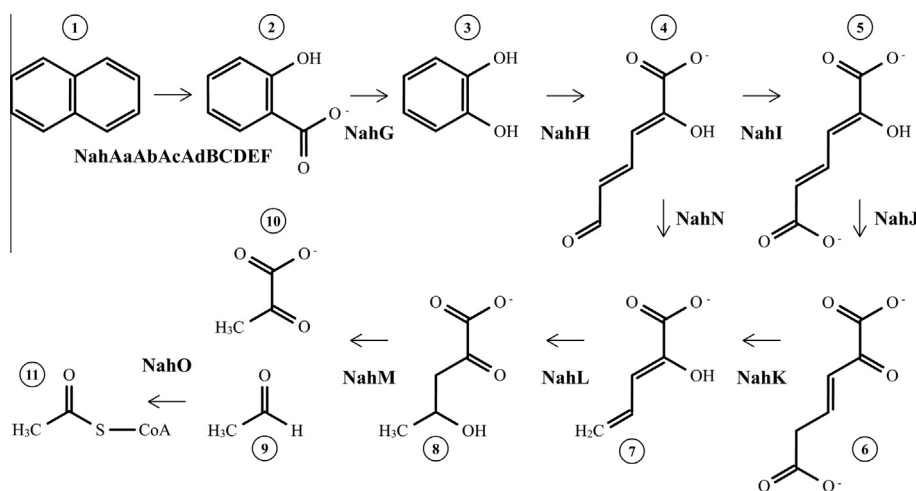


Fig. 1. Naphthalene-degradation lower pathway (from salicylate to pyruvate and acetyl-CoA). This pathway is composed of the enzymes salicylate hydroxylase (NahG), catechol 2,3-dioxygenase (NahH), 2-hydroxymuconate semialdehyde dehydrogenase (NahI), 4-oxalocrotonate tautomerase (NahJ), 4-oxalocrotonate decarboxylase (NahK), 2-hydroxymuconate semialdehyde hydrolase (NahN), vinylpyruvate hydratase (NahL), 4-hydroxy-2-oxopentanoate aldolase (NahM), and acetaldehyde dehydrogenase (NahO). The main substrates and products for these enzymes are also depicted. The metabolites are: naphthalene (1), salicylate (2), catechol (3), 2-hydroxymuconate semialdehyde (4), 2-oxo-3-hexenedioate (6), 2-hydroxy-2,4-pentadienoate (7), 4-hydroxy-2-oxopentanoate (8), acetaldehyde (9), pyruvate (10), and acetyl-CoA (11).

NAH7 plasmid by the polymerase chain reaction (PCR) using sense 5'-**CATATG**AAAGAGATCAAGCATTTTC-3' and antisense 5'-**AAGCTT**CAGGGATGCAGTCATG-3' primers. The *NdeI* and *HindIII* restriction enzyme cloning sites are shown in bold, respectively. The PCR was carried out in a final volume of 60 μ L containing 225 ng of template plasmid, 0.20 mM dNTPs, 1 $\times$ Platinum Taq buffer, 2 mM $MgSO_4$, 1 Unit Platinum Taq DNA polymerase (Invitrogen) and 0.6 μ M of the forward and reverse primers. DNA was amplified using a thermal cycler with the following protocol: 95 °C for 120 s in the first cycle; 95 °C for 60 s, 47 °C for 60 s and 68 °C for 90 s; for 25 cycles. A final extension step was carried out at 68 °C for 300 s.

The amplified product was cloned into the pGEM-T vector (Promega Corporation) and the recombinant plasmid (pGEM-*nahl*) was inserted into *E. coli* DH5 α competent cells, which were screened for positive clones by colony PCR using the sense and antisense primers for *nahl* amplification. DNA fragments obtained from the digestion of pGEM-*nahl* with *NdeI* and *HindIII* were ligated into pET28a(TEV) vector [20], previously digested with the same restriction enzymes, resulting in the recombinant expression vector pET28a(TEV)-*nahl*. pET28a(TEV) is a modified pET28a vector (Novagen) that adds a 6 $\times$ His-tag followed by a tobacco etch virus (TEV) protease-cleavage site at the N-terminus of the target protein (MGHHHHHHENLYFQGH-target protein).

Electrocompetent *E. coli* DH5 α cells were transformed with pET28a(TEV)-*nahl* for storage while *E. coli* BL21 (DE3) and *E. coli* BL21 ArcticExpress (DE3) (Agilent Technologies) cells transformed with the same vector were used for protein expression. Plasmid DNA was extracted from cells using the AxyPrep Plasmid Miniprep kit according to the manufacturer's instructions (Axygen Biosciences). Purification of PCR products or DNA fragments from agarose gel slices was carried out by using Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare Life Sciences). Gene insertion was verified by colony PCR and the fidelity of the *nahl* nucleotide sequence was confirmed by DNA sequencing using T7 primers in a MegaBACE Sequencing System instrument (GE Healthcare Life Sciences).

Protein expression

Recombinant Nahl was first overexpressed in *E. coli* BL21 (DE3) cells harboring the pET28a(TEV)-*nahl* plasmid. A single colony from an agar plate was used to inoculate 20 mL of 2 $\times$ YT medium (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl) supplemented with 50 μ g/mL kanamycin. After overnight growth at 37 °C with agitation at 200 rpm, this primary culture was used to inoculate 1 L 2 $\times$ YT medium supplemented with the same antibiotic and allowed to grow at 37 °C until the OD₆₀₀ (optical density at 600 nm) reached 0.5–0.7. Protein expression was induced by the addition of isopropyl beta-D-thiogalactopyranoside (IPTG) to a final concentration of 1.0 mM. The cell culture was allowed to grow at 37 °C for 3 h at 220 rpm. Cells were harvested by centrifugation at 10,000g for 10 min at 4 °C.

In a second attempt to overexpress recombinant Nahl, transformed *E. coli* BL21 ArcticExpress (DE3) cells were inoculated into a 20 mL 2 $\times$ YT medium containing 50 μ g/mL kanamycin and 20 μ g/mL gentamicin and grown overnight (200 rpm at 37 °C). This culture was then transferred to a 1 L 2 $\times$ YT medium without antibiotics and grown at 30 °C for 3 h. Subsequently, the temperature was reduced to 12 °C and heterologous protein expression (6 $\times$ His tagged protein; 6 $\times$ His-Nahl) was induced by the addition of IPTG to a final concentration of 1.0 mM. Incubation proceeded for 24 h at 12 °C before harvesting the cells by centrifugation (10,000g for 10 min at 4 °C).

The protein expression level was checked by Coomassie blue stained 12% or 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE).

Protein purification

Harvested cells were resuspended in 20 mL of buffer A (20 mM sodium phosphate pH 7.4, 500 mM NaCl and 30 mM imidazole) and disrupted using an EmulsiFlex-C3 homogenizer (Avestin). The lysate was clarified by centrifugation at 15,000g for 20 min at 4 °C. The supernatant was filtered through a 0.45 μ m pore-size membrane filter and loaded onto a 5-mL Ni²⁺ affinity column HisTrap HP (GE Healthcare Life Sciences) connected to an ÄKTAprime chromatography system (GE Healthcare Life Sciences). Both the column and system were previously equilibrated with buffer A. In order to remove nonspecific binding proteins, the column was washed with five column volumes of buffer A. Subsequently, the recombinant protein (6 $\times$ His-Nahl) was eluted with a linear imidazole gradient using a combination of buffers A and B (20 mM sodium phosphate pH 7.4, 500 mM NaCl and 500 mM imidazole). Fractions containing 6 $\times$ His-Nahl were pooled and buffer-exchanged into buffer C (50 mM Tris–HCl pH 7.4 and 50 mM NaCl) using the same ÄKTAprime system and a HiPrep 26/10 desalting column (GE Healthcare Life Sciences).

The 6 $\times$ His-Nahl imidazole-free sample was then loaded onto a HiLoad 16/60 Superdex 200 (GE Healthcare Life Sciences) size-exclusion chromatography column equilibrated with buffer D (50 mM sodium phosphate buffer pH 7.4 and 50 mM NaCl) for a final purification step. Alternatively, samples were incubated with recombinant histidine-tagged TEV protease (5% w/w) at 4 °C for 16 h for 6 $\times$ His-tag removal. A second Ni²⁺ affinity chromatography step was then used to further separate untagged Nahl (GH-Nahl) from contaminants – including histidine-tagged TEV protease – prior to a final purification step.

Whenever required, tagged and untagged Nahl samples obtained from SEC were concentrated by ultrafiltration using a Vivaspinn sample concentrator (10 kDa MWC, GE Healthcare Life Sciences).

Protein purity was checked by Coomassie blue stained 12% or 15% SDS–PAGE at all steps. The protein concentration at each purification step was determined from its absorbance at 280 nm using the appropriate extinction coefficient. The 6 $\times$ His-Nahl and GH-Nahl molar extinction coefficients (57,410 and 55,920 M^{−1} cm^{−1}, respectively) were estimated from their amino acid composition by ProtParam tool [21] using the ExPASy Server [22].

Circular dichroism spectroscopy

The secondary structure content of GH-Nahl in the presence or absence of NAD⁺ was assessed by circular dichroism (CD) spectroscopy using a JASCO J-815 spectropolarimeter (Jasco Corporation). Far UV CD spectra (190–260 nm) were recorded at a protein concentration of 0.3 mg/mL in the temperature range of 20–95 °C, with a scan rate of 100 nm/min, a step size of 1.0 nm and 5 accumulations per scan. The same 0.5 mm path length quartz cuvette was used throughout the experiments. Spectra were baseline-corrected by subtracting the spectra of buffer solution (50 mM sodium phosphate buffer pH 7.4 and 50 mM NaCl) from all measurements. Samples were clarified by centrifugation (13,000g for 10 min at 4 °C) before assays. The content of GH-Nahl secondary structure elements was calculated with the program CDNN v.2.1 [23]. Assignments of the secondary structure elements from 3D structures of a few proteins deposited in the Protein Data Bank (PDB) [24] were obtained using the “STRIDE:

Protein secondary structure assignment from atomic coordinates" server [25].

Enzymatic assays and kinetics

2-Hydroxymuconate semialdehyde was generated enzymatically using partially purified catechol-2,3-dioxygenase (C230) [26,27], as follows. Catechol (200 mg, 1.8 mmol) was dissolved in 20 mM sodium phosphate buffer (30 mL, pH 7.5). To this mixture was added a 100 μ L aliquot of C230 (11.5 mg/mL). Pure oxygen gas was bubbled directly into the reaction as it stirred in order to enhance activity. Subsequently, aliquots (50 μ L) of C230 were added every 15 min (4 $\times$) and the pH adjusted to 7.3–7.6 with aliquots of 1 M NaOH. The reaction was complete when the pH no longer required adjustment, indicating that the acid was no longer being generated. As the reaction progressed, a yellow color appeared, which turned to deep red, indicative of the formation of 2-hydroxymuconate semialdehyde. The solution was acidified to pH $\sim$ 2 with 8.5% phosphoric acid and extracted with ethyl acetate (3 $\times$ 30 mL). The organic layers were pooled, dried over anhydrous Na₂SO₄, filtered, and evaporated to dryness at room temperature. The resulting solid was dissolved in ethyl acetate and recrystallized. The compound is stable at room temperature as a solid. A ¹H NMR spectrum in deuterated methanol indicates that the cyclic acetal of 2-hydroxymuconate semialdehyde is isolated. However, in aqueous buffer at pH 8.0, only the ring-opened form of 2-hydroxymuconate semialdehyde is present. The yield of 2-hydroxymuconate semialdehyde was 48 mg (0.3 mmol, 17%). Cyclic Acetal: ¹H NMR (CD₃OD, 500 MHz) δ 5.61 (1H, d, H6, J = 3.9 Hz), 5.82 (1H, ddd, H5, J = 1.1, 3.9, 9.5 Hz), 6.35 (1H, dd, H4, J = 6.1, 9.5 Hz), 6.54 (1H, dd, H3, J = 1.1, 6.1 Hz) ppm; ¹³C NMR (CD₃OD, 500 MHz) δ 96.76 (C6), 109.16 (C3), 120.15, 123.39 (C4, C5), 141.66 (C2), 164.54 (C1) ppm. Ring-opened compound 4 (in Fig. 1): δ 5.65 (2H, m, H3 and H5), 7.52 (1H, t, H4, J = 13.37 Hz), 8.65 (1H, d, H6, J = 8.8 Hz) ppm.

2-Hydroxymuconate semialdehyde dehydrogenase (Nahl) activity was monitored by following the rate of disappearance of 2-hydroxymuconate semialdehyde at 375 nm (λ_{max} = 22,966 M⁻¹ cm⁻¹). The production of NADH at 340 nm could not be observed due to substrate absorbance. The reaction mixture consisted of 50 mM potassium phosphate buffer pH 8.5, 50 mM NaCl, containing several concentrations of 2-hydroxymuconate semialdehyde, 200 μ M NAD⁺, and 0.003 mg/mL of enzyme. Both substrate and cofactor were dissolved in 50 mM potassium phosphate buffer pH 8.5 before adding to the reaction mixture. The assay was initiated by addition of 2–40 μ L of 2-hydroxymuconate semialdehyde to a final reaction volume of 1.0 mL. All the spectrophotometric assays were performed in quartz cuvettes of 10 mm path length at 25 °C, and data were obtained using an Agilent 8453 diode-array spectrophotometer. The Michaelis–Menten kinetic parameters were fitted by non-linear regression of the initial velocities (v_0) versus concentrations of the substrate 2-hydroxymuconate semialdehyde using GraFit program (Erithacus Software Ltd., Staines, U.K.).

The kinetic activity of Nahl with salicylaldehyde, another intermediate in the naphthalene-degradation pathway, was also evaluated under similar conditions. However, activity was monitored by following the rate of formation of NADH at 340 nm.

Size-exclusion chromatography and dynamic light scattering analysis

A single 9 mL pool of GH-Nahl (at 2.3 mg/mL), obtained after SEC, was used to prepare diluted and concentrated protein aliquots (0.5, 1.5 and 6 mg/mL). Each aliquot was separated into two fractions and NAD⁺ was added to one of them to a final molar concentration 10 times greater than that of protein, and incubated for an

hour. Each fraction was then used in dynamic light scattering (DLS) and SEC experiments.

The DLS assays were performed at 25 °C on a Zetasizer Nano ZS-90 (Malvern) instrument. Each sample was centrifuged at 13,000g for 10 min at 4 °C before the assay, in order to remove impurities or insoluble particles. Each sample was measured from 10 to 100 times (depending on the protein concentration) for 10 s in a 10 mm path length quartz cuvette. Data were collected and analyzed using Malvern Zetasizer software v. 6.01, where the "Protein Utility" routine was used to estimate particle mass based on hydrodynamic radius (R_h).

The same samples used in the DLS assays were evaluated by SEC. The samples were individually loaded onto a HiLoad 16/60 Superdex 200 column equilibrated with buffer D using a flow rate of 1.0 mL/min. The chromatography profiles of GH-Nahl in the presence or absence of NAD⁺ were compared with those of the following standards: cytochrome C (12.4 kDa), albumin (66 kDa), alcohol dehydrogenase (150 kDa) and β -amylase (200 kDa).

Small-angle X-ray scattering

Data reduction and overall parameters

Small-angle X-ray scattering (SAXS) curves for GH-Nahl samples in buffer D were recorded at the Laboratório Nacional de Luz Síncrotron (Brazilian Synchrotron Light Laboratory, LNLS) at the SAXS-1 beam line, equipped with a Dectris Pilatus 300 K detector (84 mm $\times$ 107 mm). Samples were centrifuged (13,000g for 10 min at 4 °C) prior to data collection. The sample-to-detector distance was set to 1617.34 mm and the radiation wavelength used was 1.55 Å, with q ranging from 0.0069 to 0.2692 Å⁻¹ ($q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle). Potential radiation damage was monitored by collecting, for each sample, a series of consecutive curves with increasing exposure times (from 5 to 60 s). Data integration, normalization to the intensity of the transmitted beam and sample attenuation, and buffer scattering subtraction were carried out with Fit2D [28]. Absolute calibration of scattering data was performed using water as a secondary standard [29,30]. Data analysis and plotting were performed with ATSAS 2.5.1.1 [31] and GNUPLOT (<http://www.gnuplot.info>) under Slackware Linux (<http://www.slackware.com>). Values for the radius of gyration (R_g) and the zero-angle scattering intensity ($I(0)$) were obtained both from the Guinier approximation [32] $I(q) = I(0)\exp(-q^2R_g^2/3)$, valid for $qR_g \lesssim 1.3$, and by the indirect Fourier transform method as implemented in the program GNOM [33]. The latter was used to calculate the pair-distance distribution function, $P(r)$, from the scattering curve. The conformational state of the protein in solution was evaluated by a Kratky plot ($q^2I(q) \times q$) obtained directly from the scattering curve [34–36]. Estimates for the molecular mass of the scattering particles in solution were obtained by a method which is independent of protein concentration measurements, based on the Porod invariant, $\int_0^\infty q^2I(q)dq$ (integration area under the Kratky plot) [37].

Protein structure prediction

CRY SOL [38] was used to fit the theoretical scattering curves computed for structures available in the Protein Data Bank (PDB) [24] to the experimental data measured for GH-Nahl. In each case, the maximum distance observed in each structure was calculated with the program NCONT [39]. Once the most probable tetrameric arrangement of GH-Nahl was determined (as being analogous to the PDB entry 4I1W), a three-dimensional model for GH-Nahl monomer was predicted by threading using its amino acid sequence, as implemented in the I-TASSER platform [40]. The resulting monomer structure was then individually superposed onto each one of the four chains composing the biological unit,

resulting in the final GH-Nahl putative tetramer. The FoXS program [41,42] was used to calculate the theoretical scattering curves from the dimer (similar to chains A and B of PDB entry 4I1W) and tetramer (similar to PDB entry 4I1W) models of Nahl and its Minimal Ensemble calculation routine was used to estimate the percentage of each oligomer in solution.

Ab initio shape recovery from SAXS curves

Forty dummy atom models for the GH-Nahl tetramer were generated with DAMMIF [43] in mode slow, enforcing particle symmetry 222 ("P222" option in DAMMIF). Post-processing within the entire set of models was carried out both with DAMAVER [44] and DAMCLUST [31] for evaluation of model similarity and clustering of models, respectively. Superposition of models was done with SUPCOMB [45]. Visualization and graphical analysis of three-dimensional models were performed with RASMOL [46] and PyMOL (The PyMOL Molecular Graphics System, Version 1.7.1.1, Schrödinger, LCC). Computational jobs were automated with C shell scripts.

Results and discussion

Expression and purification

The recombinant form of 2-hydroxymuconate semialdehyde dehydrogenase from *P. putida* G7 (6×His-Nahl) was first expressed at 37 °C for 3 h in *E. coli* BL21 (DE3) cells. This first protein expression attempt resulted in an insoluble 6×His-Nahl (data not shown). To overcome this, *E. coli* BL21 ArcticExpress (DE3) was chosen as an alternative bacterial strain for 6×His-Nahl expression at 12 °C for 24 h after induction with IPTG. These cells co-express the cold-adapted chaperonins Cpn10 and Cpn60 from *Oleispira antarctica*, which share 74% and 54% amino acid sequence identity to the *E. coli* GroEL and GroES chaperonins, respectively. They confer improved protein processing at lower temperatures (from 12 to 4 °C), potentially increasing the yield of active and soluble recombinant protein [47]. As observed in SDS–PAGE gels (Fig. 2), this expression protocol resulted in a pronounced fraction of soluble 6×His-Nahl, as shown by the large expression band around 55 kDa corresponding to the molecular mass of the 6×His-Nahl monomer (53,518 Da). Only a small portion of the recombinant enzyme was found in the insoluble fraction (Fig. 2). Metal affinity chromatography (MAC; Ni²⁺-affinity) followed by size-exclusion chromatography (SEC) were used for protein purification which allowed the recovery of approximately 30 mg of pure 6×His-Nahl per liter of cell culture. After His-tag removal with TEV protease, the cleaved Nahl enzyme (GH-Nahl; 51,712 Da) contained two extra amino acid residues (glycine and histidine) at its N-terminus compared to the native Nahl. GH-Nahl remained soluble with no signs of protein aggregation or precipitation.

Circular dichroism spectroscopy

Circular dichroism spectroscopy was used to evaluate the proper folding of recombinant GH-Nahl and the structural effects of temperature and NAD⁺ binding. For this purpose, GH-Nahl obtained after SEC purification, in the presence or absence of NAD⁺, was independently evaluated at different temperatures (from 20 to 95 °C, in 5 °C increments) in a spectropolarimeter. The results are shown in Fig. 3.

The CD spectra of GH-Nahl at 20 °C, in the presence and absence of NAD⁺, are very similar to each other over the entire measured wavelength range (Fig. 3A), with only a slight difference below 195 nm. This difference could be explained either by the higher level of noise in both sample and buffer spectra at this region or

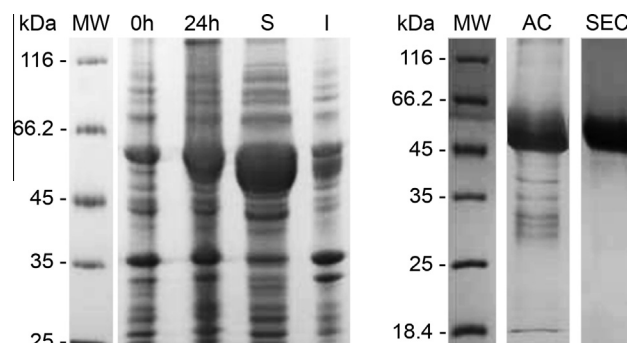


Fig. 2. Coomassie blue-stained 15% SDS–PAGE analysis of 6×His-Nahl expression in *E. coli* ArcticExpress (DE3) and purification. Lane MW: protein molecular weight marker; Lane 0 h: crude bacterial extract before IPTG induction; Lane 24 h: crude bacterial extract after 24 h of induction with 1.0 mM IPTG; Lanes S and I: soluble and insoluble fractions, respectively, after cell lysis. Lanes AC and SEC: purified fractions of 6×His-Nahl after Ni²⁺-affinity and size-exclusion chromatography.

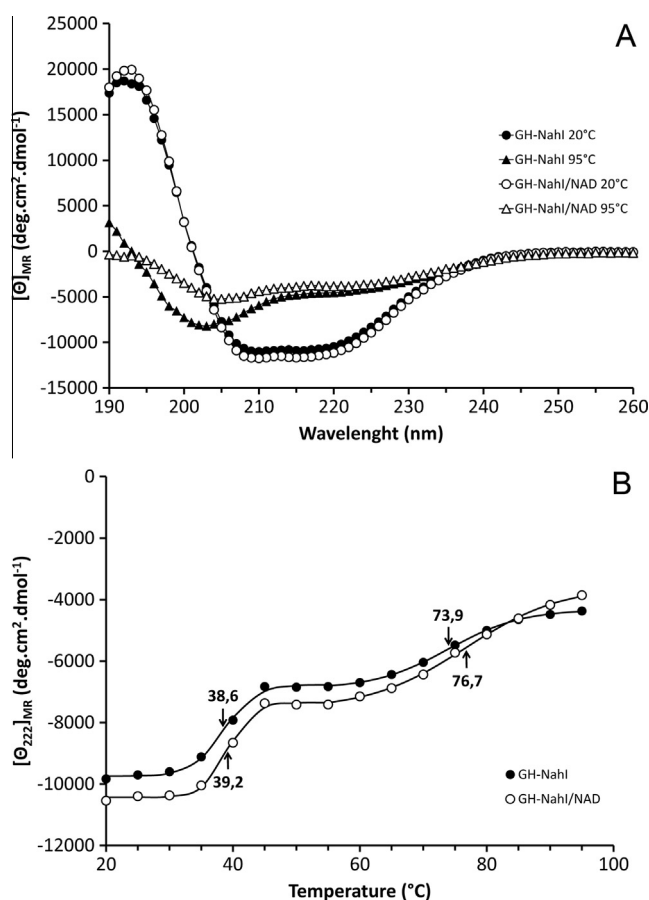


Fig. 3. A. CD spectra of GH-Nahl at 20 °C and 95 °C, in the absence and presence of NAD⁺, in 50 mM sodium phosphate buffer pH 7.4 and 50 mM NaCl. The protein concentration in both solutions is 0.3 mg/mL and the cell path lengths are 0.5 mm. B. CD melting curves recorded at 222 nm for GH-Nahl and GH-Nahl/NAD⁺ between 20 °C and 95 °C.

by a very small difference in the beta-sheet content between samples. The deconvolution of any of these two spectra using CDNN 2.0 software is insensitive to this small difference and both suggest GH-Nahl is composed of 35% α -helices, 17% β -sheets and 46% other secondary structure elements (17% β -turns and 29% random coils). These results suggest that NAD⁺ binding to GH-Nahl does not promote any significant conformational changes in the secondary

structure of the enzyme. Moreover, they are in good agreement with the secondary structure content of 2-aminomuconate 6-semialdehyde (2-AMS) dehydrogenase from *Pseudomonas fluorescens* (PDB ID 4I1W) which shares 61% sequence identity to GH-Nahl and shows 43% α -helices and 22% β -sheets, as assigned by the STRIDE server [25]. Interestingly, Nahl shows a similar fold to that of NahF (PDB ID 4JZ6) [48], a salicylaldehyde dehydrogenase from *P. putida* G7 belonging to the same naphthalene-degradation pathway. They share only 33% sequence identity, but NahF is composed of 43% α -helices and 20% β -sheets.

As the temperature is increased up to 95 °C, CD spectra are modified towards a partially unfolded protein profile resulting in the loss of GH-Nahl α -helices content. Interestingly, the apo and holo (NAD⁺ bound) forms of the enzyme behave differently under thermal denaturation. This difference is more evident from 40 °C in the 195–210 nm range, with the greatest dissimilarities at 95 °C, as shown in Fig. 3A. This difference suggests that NAD⁺ binding to Nahl increases the stability of the complex due to the maintenance of a few β -strands in the NAD⁺ binding domain (most probable a Rossmann fold domain) during thermal denaturation. As β -strands have a positive contribution over the mean residue ellipticity ($[\theta]_{MR}$) in the 195–210 nm range, the GH-Nahl/NAD⁺ CD curve is less negative than the curve observed when monitoring the enzyme alone.

In addition, when monitoring the mean residue ellipticity at a fixed 222 nm wavelength ($[\theta]_{222nm}$) as a function of temperature (Fig. 3B) it can be seen that GH-Nahl follows a two-step denaturation process with the first transition temperature at 38.6 °C and a second transition temperature at 73.9 °C. The GH-Nahl/NAD⁺ complex also shows a two-step denaturation process but the binding of NAD⁺ to Nahl has a greater impact only over the second transition temperature (76.7 °C). This behavior supports the idea that, in these both situations, the co-factor binding domain is the less susceptible to thermal denaturation.

Size-exclusion chromatography and dynamic light scattering analysis

Size-exclusion chromatography and dynamic light scattering assays were conducted to indirectly evaluate the oligomeric state of GH-Nahl in solution. SEC profiles obtained throughout this work

using protein sample with and without NAD⁺ showed a single well-resolved peak corresponding to the enzyme elution volume (V_e). More than three chromatographic runs were conducted per sample (with and without NAD⁺). The elution volume in each run did not remain the same, even when using the same sample. GH-Nahl and GH-Nahl/NAD⁺ samples eluted between 76.7 and 79.3 mL, consistent with an apparent molecular mass of 184.3–135 kDa (Fig. 4). These results suggest that NAD⁺ binding does not interfere with the biological assembly of GH-Nahl, which might be formed by three or four Nahl monomers.

The results obtained using dynamic light scattering also suggest that GH-Nahl, either in the presence or absence of NAD⁺, might assemble as a trimer. GH-Nahl and GH-Nahl/NAD⁺ behave as a particle with a hydrodynamic diameter (D_h) of 9.8 and 9.9 nm in solution, which for globular proteins might correspond to molecular masses of 139 and 142 kDa, respectively.

Even though these results suggest that GH-Nahl might assemble as a trimer in solution, there is little evidence in the literature that ALDHs monomers associate like this. The known ALDHs assemble as dimers with interlaced domains, tetramers with two copies of dimers, or as a trimer-of-dimers hexamers [49,50]. The theoretical molecular mass of GH-Nahl monomer is 51.7 kDa, while the expected molecular masses for the dimer and the tetramer are 103.4 and 206.8 kDa, respectively. A similar situation was reported by He and co-workers [51], where gel filtration results suggested that 2-AMS dehydrogenase from *Pseudomonas pseudoalcaligenes* might assemble as a trimer. However, the authors concluded that “because a homotrimeric structure is uncommon, more research is required to rigorously determine the native structure of the 2-AMS dehydrogenase” [51].

Therefore, as a final attempt to correctly identify the oligomeric state of GH-Nahl and its overall shape in solution, a small-angle X-ray scattering experiment was conducted.

Small-angle X-ray scattering

Recombinant Nahl is a tetramer in solution

To check for a possible dependence of the oligomeric state on protein concentration, three samples of GH-Nahl were made up at different protein concentrations (1.5, 5.8 and 11.3 mg/mL) and

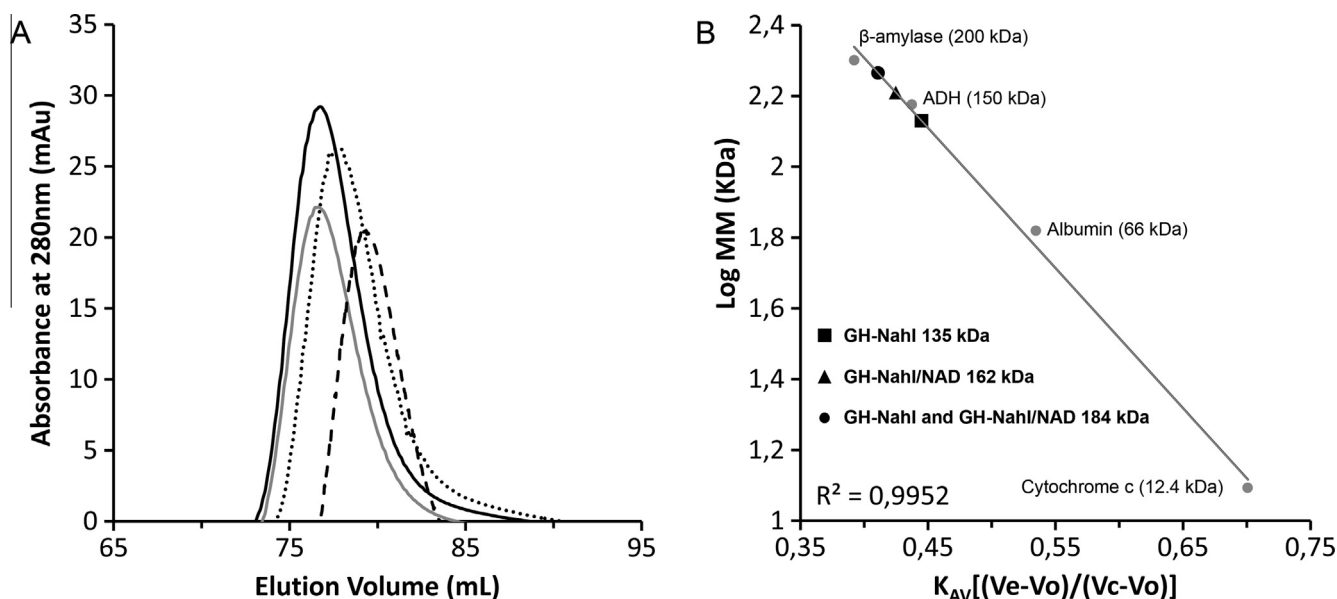


Fig. 4. A. Size-exclusion chromatography curves for GH-Nahl in the absence (dashed line, V_e = 79.3 mL/gray solid line, V_e = 76.7 mL) and presence of NAD⁺ (dotted line, V_e = 77.8 mL, black solid line/ V_e = 76.7 mL). B. Calibration curve generated using molecular weight standards.

analyzed. The results showed mass estimates equal to 185, 190, and 177 kDa, respectively. No evidence for higher order oligomers at increasing protein concentration was found. Within the inherent error of the technique [37], these results unequivocally indicate that GH-Nahl is a tetramer in solution (predicted molecular weight equal to 206.8 kDa). This result agrees with the size-exclusion chromatography data (184 kDa) reported in section “Size-exclusion chromatography and dynamic light scattering analysis”.

In all cases, radiation damage was only visible for longer (and cumulative) exposure times. Therefore, the first scattering curve collected with 20 s exposure time, from the sample at 5.8 mg/mL, shown in Fig. 5A, was kept for further analysis due to its higher signal-to-noise ratio and reliability observed for mass determination procedure. The clear linear behavior of the Guinier plot (inset of Fig. 5A) strongly suggests a monodisperse protein sample, allowing a reliable estimate of the radius of gyration as $R_g = 38.2$ Å. This value is in excellent agreement with that obtained with GNOM [33], $R_g = 37.4$ Å, during the calculation of the distance distribution function, $P(r)$, shown in Fig. 5B. This $P(r)$ function suggests a globular assembly in solution, with a maximum dimension equal to 106 Å. To complete this preliminary analysis, the Kratky plot presented in Fig. 5C, with a clearly defined maximum and decay at higher angles, typical for well-folded domains, indicates that the protein has a compact structure in solution.

The most probable oligomeric assembly

To restore the molecular envelope of GH-Nahl from the scattering curve, it is important to know whether or not symmetry is present in its oligomeric assembly in solution. Thus, possible tetrameric arrangements were first investigated by using CRY SOL [38]. Theoretical scattering curves for the 12 closest homologs, whose structures were available in the PDB, were fitted to the experimental SAXS data measured for GH-Nahl. Although these structures were determined by X-ray crystallography, it is worth noting that the content of the asymmetric unit (ASU) of the crystal does not necessarily correspond to the oligomeric state of the protein in solution. In the case of GH-Nahl, however, it was not necessary to generate crystallographic symmetry mates to take into account other possible tetrameric arrangements because the tetramer contained in the ASU of the closest homologs already exhibited a good fit to the experimental data. In addition, the R_g calculated by CRY SOL and the maximum dimension obtained with NCONT [39] for the closest homolog structures were in excellent agreement with the corresponding experimental values for GH-Nahl (data not shown). Further confirmation of the oligomeric state and monodispersity of the sample was obtained from a calculation of the minimal ensemble of molecular models required to best fit the experimental data (Fig. 5A). A minimal ensemble search was performed using FoXS software [41,42] assuming the existence of a combination of dimers and tetramers of Nahl in solution. The percentage of tetramers in solution was estimated to be 99%. Altogether, this analysis indicated that GH-Nahl is present as a tetramer in solution, assuming an arrangement compatible with the tetramer corresponding to the ASU content of the PDB entry 4I1W (2-AMS dehydrogenase from *P. fluorescens*, with an amino acid sequence identity of 61%).

The 3D model of the PDB entry 4I1W possesses a nearly perfect 222 symmetry, with three mutually almost-perpendicular, nearly twofold, rotation axes. It is known that the solution resulting from the shape determination of the SAXS curves is not unique, and, in some cases, the reliability of this procedure may be increased by the addition of information which includes particle symmetry and/or anisotropy. Reconstruction of the Nahl tetramer using DAMMIF [43] was done without restricting the particle symmetry (“P1” option), by imposing a twofold (“P2” option) axis or three

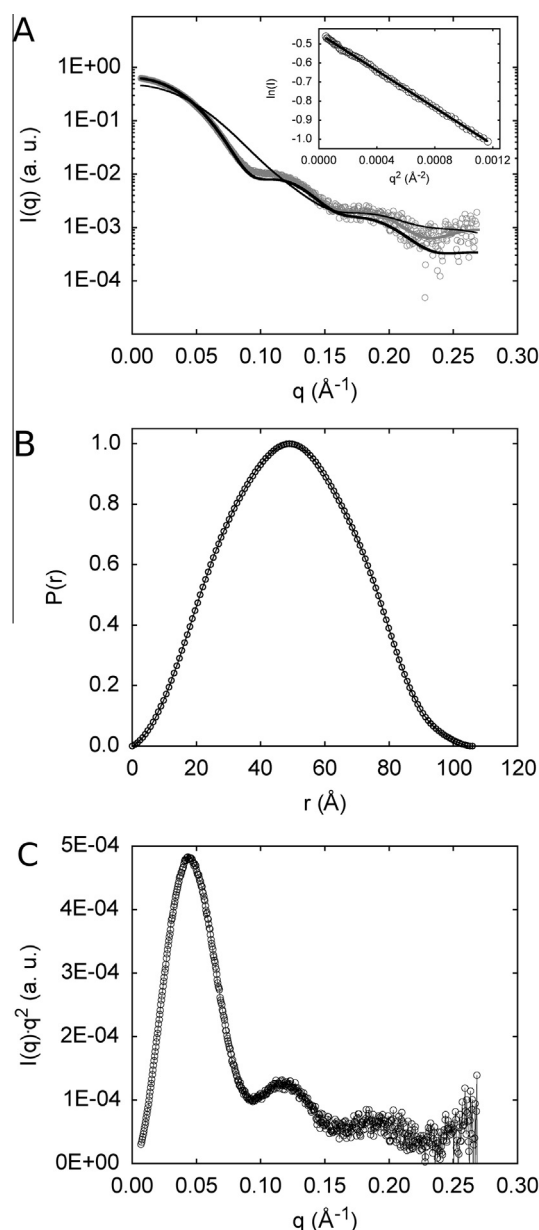


Fig. 5. SAXS data and overall parameters. A. Experimental Small-Angle X-ray Scattering curve (open circles) of GH-Nahl in solution, with the GNOM [33] fitting (thicker line in gray) obtained during the computation of the pair-distance distribution function, $P(r)$. The Guinier region and the corresponding linear fitting are shown in the inset. Theoretical SAXS curves for modeled Nahl dimer (thin line in black) and tetramer (thicker line in black) were calculated with FoXS program [41,42] and fitted to SAXS experimental data. B. Pair-distance distribution function derived from the experimental curve. The bell-shaped profile, with a centered maximum, is typical of a globular scattering particle in solution. C. Kratky plot calculated from the experimental data; typical of compact proteins, with well-folded domains. A few negative intensity points resulting from buffer subtraction in the noisy region, at higher scattering angles, were deleted previously to FoXS calculations and for figures composition. Plots were done with GNU PLOT (<http://www.gnuplot.info>) and edited with GIMP (<http://www.gimp.org>) under Slackware Linux (<http://www.slackware.com>).

perpendicular twofold rotation axes (“P222” option). The latter option significantly improved the stability and reproducibility of the solution.

The similarity among the reconstructed models was initially assessed by calculating the normalized spatial discrepancy (NSD) parameter as performed by the program suite DAMAVER [44] (once more, enforcing 222 symmetry). For an ideal superposition,

a NSD value of 0 implies very similar objects while values greater than 1 indicate that systematic differences occur [45]. In the present case, a $\text{NSD} = 0.92 \pm 0.07$ indicated that the ensemble exhibits acceptable internal agreement, but with a visible variability emerging from different portions of the dummy atom (reconstructed) models. This characteristic is a consequence of the expected diversity resulting from the enforced particle symmetry. Therefore, to evaluate more efficiently the variability and uniqueness of the solutions, all models were clustered into groups by using the NSD criteria within the program DAMCLUST [31]. The total set of 40 dummy atom models could be optimally grouped into 9 clusters. For the 7 non-isolated clusters obtained, the average NSD value calculated for each member cluster against the respective representative model was 0.55 ± 0.09 , thus indicating the procedure resulted in clusters containing highly similar models.

To interpret the resulting experimental low resolution models derived from the SAXS curve and to gain further insight into the biological problem under investigation, the 3D model in PDB entry 4I1W was used as a template to assemble a putative Nahl tetramer based on monomer chains predicted by the I-TASSER server [40], which was subsequently used as a guide to visually scrutinize the dummy atom clusters. Even though virtually all clusters individually resembled in some degree the expected Nahl tetramer, a striking result was obtained for a single cluster, easily recognizable by its well-marked 222 symmetry, which was exceptionally similar to that exhibited by the putative Nahl model, as can be seen in Fig. 6.

Enzymatic assays and kinetics

The kinetic constants for the recombinant GH-Nahl were estimated using 2-hydroxymuconate semialdehyde as the variable substrate and NAD^+ fixed at a presumed saturating concentration of 0.2 mM. The 2-hydroxymuconate semialdehyde dehydrogenase activity was measured at 25 °C by monitoring the decrease in absorbance at 375 nm corresponding to the decrease in the concentration of substrate. A plot of the initial velocities [v_0] of the enzyme-catalyzed reaction versus the substrate concentrations [2-hydroxymuconate semialdehyde] (Fig. 7A) shows that the enzyme follows Michaelis–Menten kinetics. Analysis of the experimental data gave K_M and k_{cat} values of $1.3 \pm 0.3 \mu\text{M}$ and 0.9 s^{-1} , respectively. The resulting k_{cat}/K_M value of $0.66 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ is marginally lower than that observed for XylG from *P. putida* mt-2 ($k_{\text{cat}}/K_M = 1.6 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$) [17].

Surprisingly, GH-Nahl does not show any activity (under the experimental conditions) with salicylaldehyde whereas the corresponding enzyme in *P. putida* mt-2 is able to oxidize aromatic aldehydes, although with lower specificity ($0.8\text{--}30 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$) [17].

A comparison between Nahl (*P. putida* G7) and XylG (*P. putida* mt-2) amino acid sequences and the use of the predicted Nahl monomer structure obtained in this work provide one possible explanation for this different type of behavior. These enzymes exhibit primary structure identity of 87.6% over the entire polypeptide chain. They show a remarkable difference extending over 21 consecutive amino acid residues (between Ala-87 and Ala-109; Nahl numbering). Nahl (*P. putida* G7) shows the following amino acid sequence 87-AECLDTGKPKSLASHIDIPRGAA-109 and XylG (*P. putida* mt-2) contains 87-ARMPGHRQAEVAGQPHRHSARRA-109, without any similarities (Fig. 7 B). This region (Ala-87 to Ala-109) in the modeled structure is located around the predicted catalytic site (Fig. 7 D), suggesting a possible role in substrate specificities. To date, there are no three-dimensional structures for Nahl or XylG available in the PDB.

Harayama and Reikik [52] proposed that the meta-cleavage pathway of the TOL plasmid pWWO from *P. putida* mt-2 is

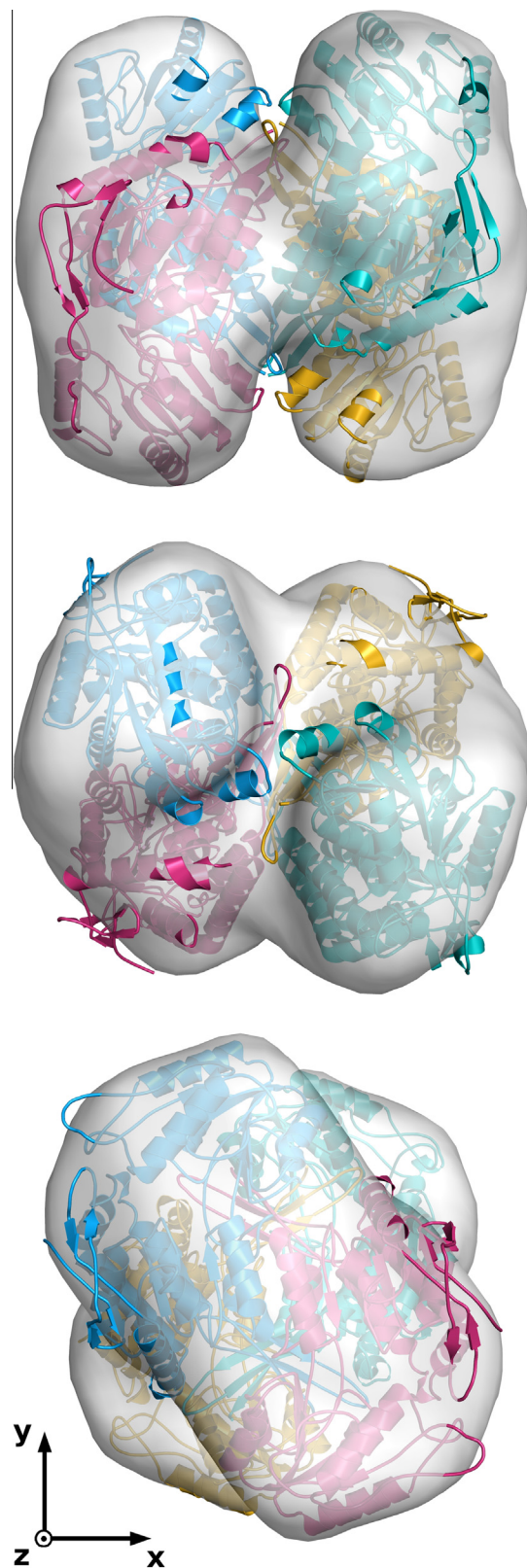


Fig. 6. Superposition of the experimental envelope (shown as a surface) onto the predicted Nahl tetramer, represented as a cartoon with monomer chains shown in different colors. The middle and bottom views are rotated clockwise by 90° around the *x*- and *y*-axes, respectively. The axes of the coordinate system of the page were chosen to make evident the striking coincidence of the envelope symmetry, with the nearly perfect 222 point-group symmetry of the tetramer. Drawings were prepared with PyMOL and edited using GIMP (<http://www.gimp.org>) under Slackware Linux (<http://www.slackware.com>). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

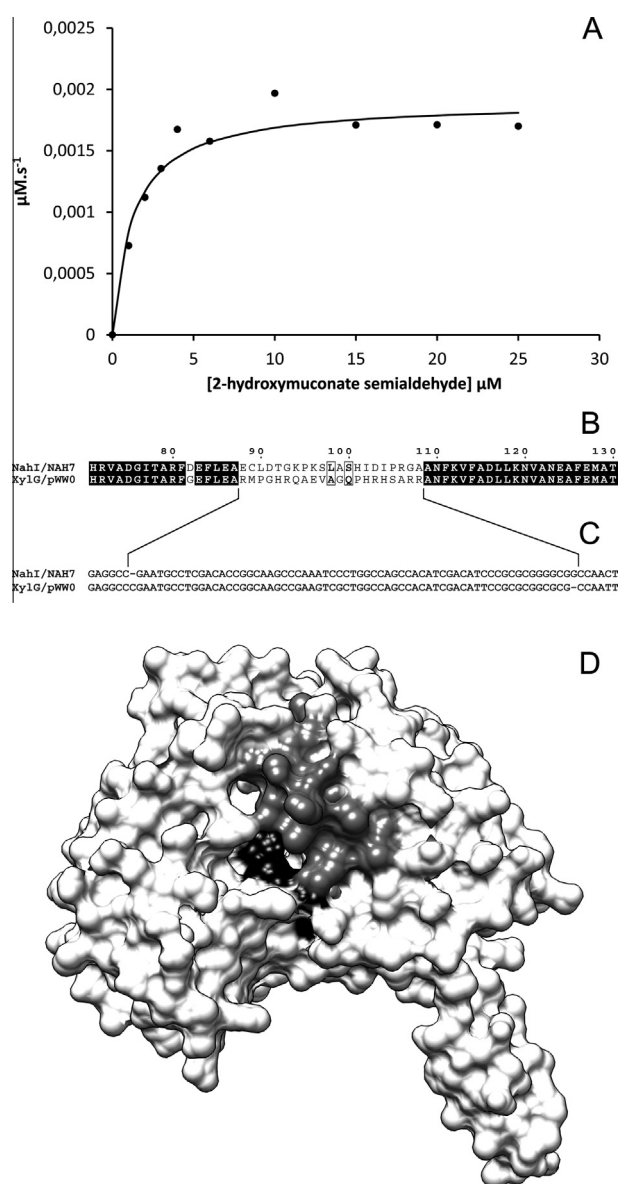


Fig. 7. A. Kinetic activity of the recombinant GH-Nahl/NAD⁺ complex with 2-hydroxymuconate semialdehyde. A plot of the initial velocities [v_0] of the enzyme-catalyzed reaction versus the concentration of 2-hydroxymuconate semialdehyde shows that the enzyme follows Michaelis-Menten kinetics. B and C. Partial amino acid and nucleotide sequence alignment between Nahl and XylG. A notable difference extending over 21 consecutive amino acid residues (Ala-87 to Ala-109) may be responsible for the different kinetic behavior observed for XylG (B). An insertion and a deletion mutation flanking this different fragment might have produced a reading frameshift (C). D. Predicted Nahl monomer showing that the fragment (Ala-87 to Ala-109 colored in gray) stands in the vicinity of the predicted catalytic site (black).

composed of hybrid genes evolved from a recombination product of the ancestral homologs *nah* (NAH7 from *P. putida* G7) and *dmp* (pVI150 from *P. putida* CF600) sequences. Since naturally occurring TOL and NAH plasmids often contain multiple meta-cleavage pathway operons whose nucleotide sequences are similar but not identical, in organisms containing these TOL and NAH plasmids, the chance of recombination between non-identical meta-cleavage pathway genes may be significant [52]. Thus, it is tempting to suggest that this fragment might have been incorporated into the *xylG* gene leading to a dramatic change in the specificity of this enzyme. However, a comparison of *nahl* and *xylG* nucleotide sequences revealed that an insertion and a deletion mutation flanking this

63 bp fragment might have produced a reading frameshift (Fig. 7C). Therefore, it is possible that the *xylG* sequence has been positively selected, since its product possesses some advantageous catalytic properties.

Conclusions

In this work, we report the successful cloning, expression, purification of 2-hydroxymuconate semialdehyde dehydrogenase (Nahl) from *P. putida* G7. The enzyme was overexpressed in *E. coli* Arctic Express (DE3) at 12 °C as an N-terminal hexa-histidine-tagged fusion protein, which was purified by affinity and gel filtration chromatography. His-tag removal with recombinant TEV protease was successfully employed and an almost native enzyme (GH-Nahl) was obtained. The recombinant enzyme was structurally and kinetically characterized, and displays results consistent with the native enzyme.

The GH-Nahl secondary structure content, as assessed by circular dichroism (CD) spectroscopy, was similar to 2-AMS dehydrogenase from *P. fluorescens* (PDB ID 4I1W), another ALDH which shares 61% sequence identity to the recombinant enzyme. Despite the fact that NAD⁺ binding to GH-Nahl did not cause significant conformational changes in the secondary structure of the enzyme, it seems to increase the stability of the complex probably by maintaining a few β -strands in the NAD⁺ binding domain during thermal denaturation. In addition, GH-Nahl followed a two-step denaturation process. The GH-Nahl/NAD⁺ complex had a higher second transition temperature, which reinforces the idea that the co-factor binding domain is less susceptible to thermal denaturation.

Even though SEC and DLS results suggest that GH-Nahl might assemble as a trimer, the low-resolution Nahl structure in solution confirms that this enzyme has a tetrameric arrangement similar to the 4I1W tetramer (i.e., 2-AMS dehydrogenase). The putative Nahl tetramer is formed by two dimers of dimers that are oriented anti-parallel to each other to form a compact structure. The two dimers are easily identified and have the same structure topology as the ALDHs members that have been previously characterized experimentally.

The kinetic assay confirmed the preference of Nahl for 2-hydroxymuconate semialdehyde, with a k_{cat}/K_M value similar to that of another dehydrogenase (XylG) from the same subfamily. Surprisingly, Nahl showed no activity with salicylaldehyde, in contrast to XylG, which has activity with different aromatic substrates. The reason for this difference might be associated with a specific 21-amino acid fragment, which is remarkably different for the two enzymes and located near the predicted catalytic site. Comparison of *nahl* and *xylG* nucleotide sequences revealed that an insertion and a deletion mutation might have produced a reading frameshift, leading to a dramatic change in the catalytic properties of a 2-hydroxymuconate semialdehyde dehydrogenase.

Acknowledgments

We gratefully thank Dr Masataka Tsuda and Dr Masahiro Sota for kindly providing us with the *P. putida* G7 strain and the Laboratório Nacional de Luz Síncrotron (CNPEM, Campinas, Brazil) for provision of synchrotron-radiation facilities (SAXS-1 beam line). This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Project 482173/2010-6), Fundação de Desenvolvimento à Pesquisa do Estado de Minas Gerais (FAPEMIG; Projects EDT-0185-0.00-07, Rede-170/08 and RDP-00174-10), VALE S.A. (RDP-00174-10), the National Institutes of Health Grant (GM-41239, CPW), and the Robert A. Welch Foundation (F-1334, CPW). SSA, CMLN and RA are recipients of scholarships and research fellowship from CNPq.

References

- [1] R.H. Peng, A.S. Xiong, Y. Xue, X.Y. Fu, F. Gao, W. Zhao, Y.S. Tian, Q.H. Yao, *FEMS Microbiol. Rev.* 32 (2008) 927–955.
- [2] M. Saeed, S. Higginbotham, N. Gaikwad, D. Chakravarti, E. Rogan, E. Cavalieri, *Free Radic. Biol. Med.* 47 (2009) 1075–1081.
- [3] A.K. Haritash, C.P. Kaushik, *J. Hazard. Mater.* 169 (2009) 1–15.
- [4] EPA, *Toxicological review of naphthalene* (2004).
- [5] S. Singh, S.H. Kang, A. Mulchandani, W. Chen, *Curr. Opin. Biotechnol.* 19 (2008) 437–444.
- [6] T.D. Sutherland, I. Horne, K.M. Weir, C.W. Coppin, M.R. Williams, M. Selleck, R.J. Russell, J.G. Oakeshott, *Clin. Exp. Pharmacol. Physiol.* 31 (2004) 817–821.
- [7] M. Alcalde, M. Ferrer, F.J. Plou, A. Ballesteros, *Trends Biotechnol.* 24 (2006) 281–287.
- [8] B. Cao, K. Nagarajan, K.C. Loh, *Appl. Microbiol. Biotechnol.* 85 (2009) 207–228.
- [9] R.A. Rossello-Mora, J. Lalucat, E. Garcia-Valdes, *Appl. Environ. Microbiol.* 60 (1994) 966–972.
- [10] K.M. Yen, I.C. Gunsalus, *Proc. Natl. Acad. Sci. U. S. A.* 79 (1982) 874–878.
- [11] J.I. Davies, W.C. Evans, *Biochem. J.* 91 (1964) 251–261.
- [12] K.M. Yen, I.C. Gunsalus, *J. Bacteriol.* 162 (1985) 1008–1013.
- [13] R. Bosch, E. Garcia-Valdes, E.R. Moore, *Gene* 245 (2000) 65–74.
- [14] A. Marchler-Bauer, C. Zheng, F. Chitsaz, M.K. Derbyshire, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, S. Lu, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, S.H. Bryant, *Nucleic Acids Res.* 41 (2013) D348–352.
- [15] J.M. Sala-Trepat, W.C. Evans, *Eur. J. Biochem.* 20 (1971) 400–413.
- [16] J.M. Sala-Trepat, K. Murray, P.A. Williams, *Eur. J. Biochem.* 28 (1972) 347–356.
- [17] J. Inoue, J.P. Shaw, M. Reikik, S. Harayama, *J. Bacteriol.* 177 (1995) 1196–1201.
- [18] M. Lin, J.L. Napoli, *J. Biol. Chem.* 275 (2000) 40106–40112.
- [19] M.F. Wang, C.L. Han, S.J. Yin, *Chem. Biol. Interact.* 178 (2009) 36–39.
- [20] F.R. Carneiro, T.C. Silva, A.C. Alves, T. Haline-Vaz, F.C. Gozzo, N.I. Zanchin, *Biochem. Biophys. Res. Commun.* 343 (2006) 260–268.
- [21] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, in: J.M. Walker (Ed.), *The Proteomics Protocols Handbook*, Humana Press, 2005, pp. 571–607.
- [22] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, *Nucleic Acids Res.* 31 (2003) 3784–3788.
- [23] G. Bohm, R. Muhr, R. Jaenicke, *Protein Eng.* 5 (1992) 191–195.
- [24] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Res.* 28 (2000) 235–242.
- [25] D. Frishman, P. Argos, *Proteins* 23 (1995) 566–579.
- [26] P. Cerdan, A. Wasserfallen, M. Reikik, K.N. Timmis, S. Harayama, *J. Bacteriol.* 176 (1994) 6074–6081.
- [27] T. Kobayashi, T. Ishida, K. Horiike, Y. Takahara, N. Numao, A. Nakazawa, T. Nakazawa, M. Nozaki, *J. Biochem.* 117 (1995) 614–622.
- [28] A.P. Hammersley, S.O. Svensson, M. Hanfland, A.N. Fitch, D. Hausermann, *High Pressure Res.* 14 (1996) 235–248.
- [29] E. Mylonas, D.I. Svergun, *J. Appl. Crystallogr.* 40 (2007) S245–S249.
- [30] D. Orthaber, A. Bergmann, O. Glatter, *J. Appl. Crystallogr.* 33 (2000) 218–225.
- [31] M.V. Petoukhov, D. Franke, A.V. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P.V. Konarev, D.I. Svergun, *J. Appl. Crystallogr.* 45 (2012) 342–350.
- [32] G. Porod, in: O. Glatter, O. Kratky (Eds.), *Small Angle X-ray Scattering*, Academic Press, London, 1982, pp. 17–51.
- [33] D.I. Svergun, *J. Appl. Crystallogr.* 25 (1992) 495–503.
- [34] S. Doniach, *Chem. Rev.* 101 (2001) 1763–1778.
- [35] R.P. Rambo, J.A. Tainer, *Biopolymers* 95 (2011) 559–571.
- [36] G.V. Semisotnov, H. Kihara, N.V. Kotova, K. Kimura, Y. Amemiya, K. Wakabayashi, I.N. Serdyuk, A.A. Timchenko, K. Chiba, K. Nikaido, T. Ikura, K. Kuwajima, *J. Mol. Biol.* 262 (1996) 559–574.
- [37] H. Fischer, M.D. Neto, H.B. Napolitano, I. Polikarpov, A.F. Craievich, *J. Appl. Crystallogr.* 43 (2010) 101–109.
- [38] D. Svergun, C. Barberato, M.H.J. Koch, *J. Appl. Crystallogr.* 28 (1995) 768–773.
- [39] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.* 67 (2011) 235–242.
- [40] A. Roy, A. Kucukural, Y. Zhang, *Nat. Protoc.* 5 (2010) 725–738.
- [41] D. Schneidman-Duhovny, M. Hammel, J.A. Tainer, A. Sali, *Biophys. J.* 105 (2013) 962–974.
- [42] D. Schneidman-Duhovny, M. Hammel, A. Sali, *Nucleic Acids Res.* 38 (2010) W540–544.
- [43] D. Franke, D.I. Svergun, *J. Appl. Crystallogr.* 42 (2009) 342–346.
- [44] V.V. Volkov, D.I. Svergun, *J. Appl. Crystallogr.* 36 (2003) 860–864.
- [45] M.B. Kozin, D.I. Svergun, *J. Appl. Crystallogr.* 34 (2001) 33–41.
- [46] R.A. Sayle, E.J. Milnerwhite, *Trends Biochem. Sci.* 20 (1995) 374–376.
- [47] I. Agilent Technologies, *ArcticExpress Competent Cells and ArcticExpress (DE3) Competent Cells – Instruction Manual*, (230191–12, Ed.), 2010.
- [48] J.B. Coitinho, D.M. Costa, S.L. Guimaraes, A.M. de Goes, R.A. Nagem, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 68 (2012) 93–97.
- [49] M. Luo, R.K. Singh, J.J. Tanner, *J. Mol. Biol.* 425 (2013) 3106–3120.
- [50] T.A. Pemberton, D. Srivastava, N. Sanyal, M.T. Henzl, D.F. Becker, J.J. Tanner, *Biochemistry* 53 (2014) 1350–1359.
- [51] Z. He, J.K. Davis, J.C. Spain, *J. Bacteriol.* 180 (1998) 4591–4595.
- [52] S. Harayama, M. Reikik, *Mol. Gen. Genet.* 239 (1993) 81–89.