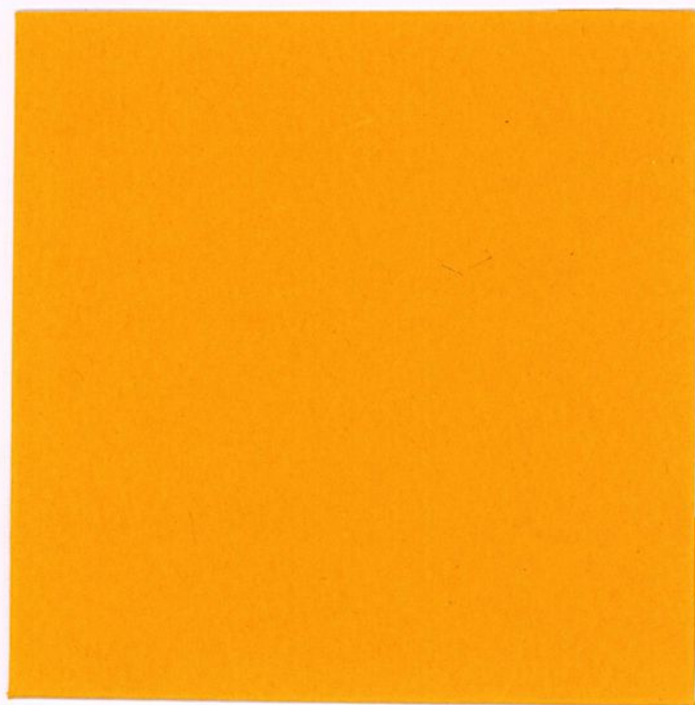


Helen Andrade Arcuri

**Bioinformática Estrutural Aplicada ao Estudo da Via
Metabólica do Ácido Chiquímico de *Xylella fastidiosa***





**Bioinformática Estrutural Aplicada ao Estudo da Via
Metabólica do Ácido Chiquímico de *Xylella fastidiosa*.**

Helen Andrade Arcuri

Dissertação apresentada para obtenção do
título de Mestre em Biofísica Molecular, área
de concentração Biofísica Molecular do
Departamento de Física do Instituto de
Biociências, Letras e Ciências Exatas da
Universidade Estadual Paulista "Julio de
Mesquita Filho" – UNESP.

Orientador: Prof. Dr. Walter Filgueira de Azevedo Júnior

São José do Rio Preto
Estado de São Paulo – Brasil
Março – 2005



577.112

A675b

15/05/2008

DOAÇÃO

AUTOR(A)

Arcuri, Helen Andrade.

Bioinformática estrutural aplicada ao estudo da Via Metabólica do Ácido Chiquímico de Xylella Fasdtidiosa/ Helen Andrade Arcuri
- São José do Rio Preto : [s.n.], 2005
120 f. : il. ; 30 cm.

Orientador: Walter Filgueira de Azevedo Júnior

Dissertação (mestrado) - Universidade Estadual Paulista.
Instituto de Biociências, Letras e Ciências Exatas

1. Biologia molecular. 2. Bioinformática estrutural. 3. Proteínas Estrutura. I. Azevedo Júnior, Walter Filgueira de. II. Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU - 577.112



AGRADECIMENTOS

A Deus por me acompanhar e iluminar sempre o meu caminho.

Ao Prof. Dr. Walter Filgueira de Azevedo Junior, pela orientação, amizade, confiança, paciência, contribuição no meu crescimento profissional, por me ensinar como fazer pesquisa e sua importância e pelo apoio durante todo o desenvolvimento desse trabalho.

A minha grande e amada família: meus pais, minha irmã, minha avó Maria, meu avô Antônio, meus padrinhos e madrinhas, meus tios e tias, meus primos e primas pelo carinho, apoio, conselhos e por sempre estarem ao meu lado nesta caminhada.

Aos meus pais por serem os meus exemplos de caráter, luta, amor, por estarem sempre ao meu lado nos momentos mais difíceis me dando apoio, palavras de carinho e incentivo, por serem pessoas maravilhosas e batalhadoras. Agradeço a vocês a minha existência e tudo que construímos juntos.

Aos colegas do Laboratório de Sistemas Biológicos pela amizade, companheirismo e pelas trocas de experiências.

Eu sei *“Quando se planta com fé e esperança se colhe com graça e abundância”*. esposa, filha, Netos José, Karina e seu marido André, Márcio, João Renato, André Martins, Sandro, Hugo, Fernanda, Papou, Ivo, Juliana, Sandro e Jaim pela amizade, auxílio, sugestões, discussões e troca de experiências durante todo o desenvolvimento do meu trabalho e pela ajuda e apoio num período especialmente difícil da minha vida.

Aos professores Dr. Roberto de Almeida, Dra. Fernanda e Dr. Walter pela valiosa contribuição e orientação de qualificação.

Aos professores Dr. João César, Dr. Paulo S. L. de Oliveira, Dra. Fernanda e ao Prof. Dr. Walter pelas sugestões durante a defesa para obtenção do título de mestre.

Aos funcionários e professores do Departamento de Física pelo ensinamento, amizade, atenção e convivência que ajudaram a facilitar o desenvolvimento do meu trabalho.

Ao CNPq pela bolsa de mestrado.

A FAPESP pelo apoio financeiro para manutenção do Laboratório de Sistemas Biomoleculares.

A todos os que contribuíram de forma direta ou indireta para a realização deste trabalho.

**Aos meus pais Maria Helena e Roonei,
exemplos de dedicação aos filhos de quem
tive carinho, incentivo e apoio. Gostaria de
compartilhar mais essa vitória,
DEDICO.**



AGRADECIMENTOS

A Deus por me acompanhar e iluminar sempre o meu caminho.

Ao Prof. Dr. Walter Filgueira de Azevedo Junior, pela orientação, amizade, confiança, paciência, contribuição no meu crescimento profissional, por me ensinar como fazer pesquisa e sua importância e pelo apoio durante todo o desenvolvimento desse trabalho.

A minha grande e amada família: meus pais, minha irmã, minha avó Maria, meu avô Antônio, meus padrinhos e madrinhas, meus tios e tias, meus primos e primas pelo carinho, apoio, conselhos e por sempre estarem ao meu lado nesta caminhada.

Aos meus pais por serem os meus exemplos de caráter, luta, amor, por estarem sempre ao meu lado nos momentos mais difíceis me dando apoio, palavras de carinho e incentivo, por serem pessoas maravilhosas e batalhadoras. Agradeço a vocês a minha existência e tudo que construímos juntos.

Aos colegas do Laboratório de Sistemas Biomoleculares pela amizade, companheirismo e pelas trocas de experiências.

Em especial aos meus amigos José Carlos e sua esposa Marilene, Walter e sua esposa Elma, Nelson José, Karisa e seu marido André, Márcio, José Renato, André Martins, Sandro, Diogo, Fernanda, Pepeu, Irene, Solange, Sandro e Jaim pela amizade, auxílio, sugestões, discussões e troca de experiência durante todo o desenvolvimento do meu trabalho ou pela ajuda e apoio num momento especial ou difícil da minha vida.

Aos professores Dr. Júlio César, Dr. José Marcio, Dra. Fernanda e ao Prof. Dr. Walter pela valiosa contribuição e sugestões no exame de qualificação.

Aos professores Dr. Júlio César, Dr. Paulo S. L. de Oliveira, Dra. Fernanda e ao Prof. Dr. Walter pelas sugestões durante a defesa para obtenção do título de mestre.

Aos funcionários e professores do Departamento de Física pelos ensinamentos, amizade, atenção e convivência que ajudaram a facilitar o desenvolvimento do meu trabalho.

Ao CNPq pela bolsa de mestrado.

A FAPESP pelo apoio financeiro para manutenção do Laboratório de Sistemas Biomoleculares.

A todos os que contribuíram de forma direta ou indireta para a realização deste trabalho.



ÍNDICE GERAL

	Páginas
ÍNDICE DE FIGURAS	vii
ÍNDICE DE TABELAS	viii
ÍNDICE DE ABREVIÇÕES	ix
ABSTRACT	xi
RESUMO	xii
 1. INTRODUÇÃO	 01
1.1. Via metabólica do ácido chiquímico	09
1.1.1. DAHPS	16
1.1.2. DHQS	18
1.1.3. DHQD	19
1.1.4. SD	21
1.1.5. SK	22
1.1.6. EPSPS	25
1.1.7. CS	29
 2. OBJETIVOS	 31
 3. METODOLOGIA E FERRAMENTAS	 32
3.1. A bioinformática	32
3.2. Métodos de alinhamento de seqüências	32
3.2.1. Programação dinâmica e os tipos de alinhamentos	34
3.3. Método de modelagem molecular por homologia ou comparativa	37
3.3.1. Procura e seleção do modelo	41
3.3.2. Alinhamento e construção do modelo	42



3.3.3. Avaliação do modelo	47
3.4. Métodos de avaliação	49
3.4.1. PROCHECK	49
3.4.2. Diagrama de Ramachandran	50
3.4.3. G-factor	53
3.4.4. VERIFY-3D	54
3.4.5. WHATIF/WHATCHECK	55
3.4.6. RMSD (Desvio médio quadrático)	55
3.4.7. GRASP	57
3.5. Modelagem das regiões de alças	59
3.6. Modelagem das cadeias laterais	61
Figura 9: A reação catalisada pela enzima DAHPS	16
4. RESULTADOS E DISCUSSÃO	63
4.1. Comparação da sequência primária	63
4.2. Qualidade dos modelos	66
4.3. Descrição dos modelos	79
4.4. Interações com ligantes	87
Figura 13: PEP e GLP	38
5. CONCLUSÃO	96
Figura 17: Predição de estrutura de sítio e modelagem comparativa de estrutura	38
6. REFERÊNCIAS	97
Figura 18: Passos da modelagem molecular comparativa	40
APENDÍCE I – MODELLER	110
Figura 19: Modelagem da estrutura de proteínas por restrição de restrição espacial	44

APENDÍCE II – Molecular Models for Shikimate Pathway Enzymes of *Xylella fastidiosa*

Figura 21: Diagrama de Ramachandran	52
Figura 22: Definição dos ângulos torçionais da cadeia principal (ϕ e ψ) e das cadeias laterais (χ)	61
Figura 23: Resultado do alinhamento das sequências primárias	64 e 66
Figura 24: Diagrama de Ramachandran dos resíduos	68 e 69



ÍNDICE DE FIGURAS

	Páginas
Figura 1: Principais sintomas do CVC	04
Figura 2: Gráfico Incidência de CVC por idade no estado de São Paulo e parte do Triângulo Mineiro	04
Figura 3: Comparação da incidência de CVC no período de 1996 a 2004	05
Figura 4: Folhas com sintomas de CVC	08
Figura 5: Estruturas químicas do ácido chiquímico e do ácido corísmico	09
Figura 6: A via metabólica do ácido chiquímico em <i>Xylella fastidiosa</i>	11
Figura 7: Biossíntese dos aminoácidos aromáticos a partir do corismato	12
Figura 8: O corismato e os compostos essenciais	14
Figura 9: A reação catalisada pela enzima DAHPS	16
Figura 10: A reação catalisada pela enzima DHQS	19
Figura 11: A reação catalisada pela enzima DHQD	20
Figura 12: A reação catalisada pela enzima SD	21
Figura 13: Reação catalisada pela enzima SK	22
Figura 14: Reação catalisada pela enzima EPSPS	25
Figura 15: PEP e GLP	26
Figura 16: Reação catalisada pela enzima CS	30
Figura 17: Predição de estrutura <i>de novo</i> e modelagem comparativa de estrutura de proteína	38
Figura 18: Passos da modelagem molecular comparativa	40
Figura 19: Modelagem comparativa de estrutura de proteína por satisfação de restrição espacial	44
Figura 20: Parte de uma cadeia polipeptídica indicando os graus de liberdade torcional de cada unidade peptídica e os ângulos de rotação ϕ e φ	50
Figura 21: Diagrama de Ramachandran	52
Figura 22: Definição dos ângulos torcionais da cadeia principal (Φ e Ψ) e das cadeias laterais (χ_1)	61
Figura 23: Resultado do alinhamento das seqüências primárias	64 a 66
Figura 24: Diagrama de Ramachandran dos modelos	68 e 69



Figura 25: Gráfico do programa VERIFY 3D dos modelos finas	72 a 74
Figura 26: Sobreposição dos modelos (linha grossa) com seus moldes (linha fina)	76
Figura 27: Figuras das estruturas secundárias das enzimas (alvos)	80 a 87
Figura 28: Alinhamento múltiplo entre as enzimas SKs dos organismos <i>M. tuberculosis</i> (MtSK), <i>E. chrysanthemi</i> (EchSK), <i>X. fastidiosa</i> (XfSK) ...	85
Figura 29: Sobreposição da estrutura secundária entre as enzimas SKs dos organismos <i>M. tuberculosis</i> (MtSK), e <i>X. fastidiosa</i> (XfSK)	85
Figura 30: Superfície de potencial eletrostático para os complexos	90
Figura 31: Sobreposição dos sítios de ligação dos complexos dos modelos (linha grossa) e moldes (linha fina)	91
Tabela 3: Resultado dos valores do VERIFY 3D dos moldes e modelos de <i>X. fastidiosa</i>	71
Tabela 4: Valor do RMSD da sobreposição para os modelos e moldes de <i>X. fastidiosa</i>	75
Tabela 5: Valor do RMSD da geometria ideal para os modelos e moldes de <i>X. fastidiosa</i>	77
Tabela 6: Informações sobre inibidores, co-fatores, íons e substratos	78 a 79
Tabela 7: Interações de van der Waals, áreas de contato e ligações intermoleculares de hidrogênio	89
Tabela 8: Ligações de Hidrogênio entre o sítio ativo de XDHQ5 e CRB e entre o sítio ativo XDHQ5 (molde) e CRB	92
Tabela 9: Ligações de Hidrogênio entre o sítio ativo de XDHQ5 e CRB e entre o sítio ativo XDHQ5 (molde) e CRB	93
Tabela 10: Ligações de Hidrogênio entre o sítio ativo de XDHQ5 e CRB e entre o sítio ativo XDHQ5 (molde) e CRB	94
Tabela 11: Ligações de Hidrogênio entre o sítio ativo de XEPSPS e GLP e entre o sítio ativo XEPSPS (molde) e GLP	95

ÍNDICE DE TABELAS

	Páginas
Tabela 1: Informações sobre as enzimas da via metabólica do ácido chiquímico de <i>X. fastidiosa</i>	15
Tabela 2: As estruturas cristalográficas usadas como molde e a porcentagem de identidade entre as seqüências primárias dos modelos (alvo) e moldes ...	67
Tabela 3: Resultado das análises do diagrama de Ramachandran dos moldes e modelos de <i>X. fastidiosa</i>	68
Tabela 4: Resultado dos valores do G-factor dos moldes e modelos de <i>X. fastidiosa</i>	70
Tabela 5: Resultado dos valores do VERIFY 3D dos moldes e modelos de <i>X. fastidiosa</i>	71
Tabela 6: Valor do RMSD da sobreposição para os modelos e moldes de <i>X. fastidiosa</i>	75
Tabela 7: Valor do RMSD da geometria ideal para os modelos e moldes de <i>X. fastidiosa</i>	77
Tabela 8: Informações sobre inibidores, co-fatores, íons e substratos	78 e 79
Tabela 9: Interações de van der Waals, área de contato e ligações intermoleculares de hidrogênio	89
Tabela 10: Ligações de Hidrogênio entre o sítio ativo de <i>Xj</i> DAHPS e PGL e entre o sítio ativo <i>Ec</i> DAHPS (molde) e PGL	92
Tabela 11: Ligações de Hidrogênio entre o sítio ativo de <i>Xj</i> DHQS e CRB e entre o sítio ativo <i>En</i> DHQS (molde) e CRB	93
Tabela 12: Ligações de Hidrogênio entre o sítio ativo de <i>Xj</i> DHQP e FA1 e entre o sítio ativo <i>Sc</i> DHQP (molde) e FA1	94
Tabela 13: Ligações de Hidrogênio entre o sítio ativo de <i>Xj</i> EPSPS e GLP e entre o sítio ativo <i>Ec</i> EPSPS (molde) e GLP	95



ÍNDICE DE ABREVIACÕES

3D – Tridimensional

ADP – Adenosina difosfato

ATP – Adenosina trifosfato

Co²⁺ - Íons cobalto

CRB - [1R-(1 α , 3 β , 4 α , 5 β)-5-(fosfometil)-1,3,4-trihidroxíciclohexano-1-ácido carboxílico

CS – Corismato sintase

CVC – Clorose variegada de citros

DAHP – 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato

DAHPS – 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato sintase

DHQ – Desidroquinato

DHQD – Desidroquinato desidratase

DHQS – 3-Desidroquinato sintase

DP – Doença de Pierce

E4P – Eritrose-4-fosfato

Ec – *Escherichia coli*

Ech – *Erwinia chrysanthemi*

En – *Emericella nidulans*

EPSP – 5-Enoilpiruvilchiquimato-3-fosfato

EPSPS - 5-Enoilpiruvilchiquimato 3-fosfato sintase

FA1 – Ácido 2,3-anidro-quinico

FAD – Flavina adenina dinucleotídeo

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo

Fe – Ferro

Fe²⁺ - Íons ferro

FMN – Flavina mononucleotídeo

Fundecitrus – Fundo de Defesa da Citricultura

GLP – Glifosato

K – Potássio

K_m – Constante de Michaelis-Menten



Mn^{2+} - Íons manganês (II)

Mt - *Mycobacterium tuberculosis*

NAD^+ - Nicotinamida adenina dinucleotídeo

NMP - Monofosfato de nucleosídeo

PABA - Ácido *p*-aminobenzóico

pb - Pares de bases

PDB - Protein Data Bank (Banco de Dados de Proteínas)

PDF - Função de densidade de probabilidade

PEP - Fosfoenolpiruvato

PGL - 2-fosfoglicolato

Phe - Aminoácido fenilalanina

P_i - Fosfato inorgânico

RMN - Ressonância magnética nuclear

RMSD - Desvio médio quadrático

S3P - Chiquimato 3-fosfato

SB - Chiquimato-ligado

Sc - *Streptomyces coelicolor*

SD - Chiquimato-5-desidrogenase

SK - Chiquimato quinase

Sp - *Streptococcus pneumoniae*

Trp - Aminoácido triptofano

Try - Aminoácido tirosina

Zn - Zinco

Zn^{2+} - Íons zinco



ABSTRACT

Xylella fastidiosa is a fastidious and gram-negative, xylem-limited bacterium that causes several economically important plant diseases as Pierce's disease and Citrus Variegated Chlorosis (CVC). CVC has been a major problem for citrus growers in the São Paulo State jeopardizing thousands of jobs and incomes. Based on the fact that plants may show disease symptoms a long time after infection in *X. fastidiosa* – citrus interaction. Certain strains of *X. fastidiosa* affect other commercially important products such as coffee, nuts, almond, and other fruits.

The genome sequence of *X. fastidiosa* has revealed the presence of genes of shikimate pathway that is of pivotal importance for production of a plethora of aromatic compounds in plants, bacteria, and fungi. The shikimate pathway consists of seven enzymes that catalyze the sequential conversion of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) to chorismate. Putative structural differences in the enzymes from the shikimate pathway, between the proteins of bacterial origin and those of plants could be used for the development of a drug for the control CVC.

From the structural information, obtained using molecular modeling, it will be possible to improve inhibitor drawing based in the structural models. These inhibitors have the potential to be used as drug against the *X. fastidiosa* and other microorganisms; however any new inhibitor for these enzymes of the shikimate pathway must have high specified for the bacterium of *X. fastidiosa*, since these enzymes also are present in plants. The detailed structural information on these enzymes will help to identify these structural differences between the plant and the bacterium; here we describe the molecular modeling of seven enzymes of the shikimate pathway of *X. fastidiosa*.



The analysis of the structural models of shikimate pathway enzymes of *X. fastidiosa* in complex with inhibitors possible, strongly indicate that the previously identified inhibitors may also inhibit the *X. fastidiosa* enzymes. Especially interesting are the enzymes: 5-Enolpyruvylshikimate-3-phosphate synthase (EPSPS) in complex with glyphosate (GLP) and shikimate-3-phosphate (S3P) and 3-Dehydroquinate dehydratase (DHQD) in complex with 2,3-anhydro-quinic acid (FA1) of *X. fastidiosa* that indicate a contact area more than the enzymes EPSPS and DHQD of templates.

RESUMO

Xylella fastidiosa é uma bactéria fastidiosa e gram-negativa, que vive limitada ao xilema de plantas, causando várias doenças economicamente importantes como a doença de Pierce e a Clorose Variegada dos Citros (CVC). A CVC vem sendo um grande problema para os citricultores do estado de São Paulo colocando em risco milhares de empregos e a geração de divisas. Um fato observado na interação *X. fastidiosa* – citros, é que plantas infectadas por *X. fastidiosa* podem demonstrar sintomas da doença em um longo tempo após a infecção. Certas linhagens de *X. fastidiosa* afetam outros produtos comercialmente importantes como café, nozes, amêndoas e outras frutas.

O sequenciamento do genoma de *X. fastidiosa* revelou a presença de genes da via metabólica do ácido chiquímico, que tem uma importância fundamental na produção de uma plethora de compostos aromáticos em plantas, bactérias e fungos. A via metabólica do ácido chiquímico consiste de sete enzimas que catalisam a conversão sequencial de eritrose-4-fosfato (E4P) e fosfoenolpiruvato (PEP) a corismato. Diferenças estruturais putativas nas enzimas da via metabólica do ácido chiquímico, entre a proteína de origem bacteriana e da planta podem ser usadas para o desenvolvimento de uma droga de controle para o CVC.

A partir das informações estruturais, obtidas da modelagem molecular, será possível efetuar o desenho de inibidores baseado nos modelos estruturais. Estes inibidores têm o potencial de serem usados como droga contra a *X. fastidiosa* e outros microrganismos, contudo qualquer novo inibidor para estas enzimas deve ter alta especificidade com a bactéria de *X. fastidiosa*, pois estas enzimas também estão presentes em plantas. A informação estrutural detalhada sobre estas enzimas irá ajudar a identificar estas diferenças



1. INTRODUÇÃO

estruturais entre a planta e a bactéria, descrevemos a modelagem molecular das sete enzimas da via metabólica do ácido chiquímico da *X. fastidiosa*.

A análise dos modelos estruturais das enzimas da via metabólica do ácido chiquímico de *X. fastidiosa* em complexo com possíveis inibidores, indicam que estes inibidores estão fortemente ligados a estas enzimas e que podem ser possíveis inibidores contra *X. fastidiosa*. Especialmente interessante, são as enzimas: 5-Enoilpiruvilchiquimato 3-fosfato sintase (EPSPS) em complexo com glifosato (GLP) e chiquimato-3-fosfato (S3P) e Desidroquinato desidratase (DHQD) em complexo com ácido 2,3-anidro-quinico (FA1) de *X. fastidiosa* que indicam uma área de contato maior que as enzimas EPSPS e DHQD dos moldes.

No Brasil, a clorose variegada do citros (CVC) vem comprometendo a produção de laranjas, posto em risco o sustento de várias famílias e prejudicando seriamente a economia do Brasil, principalmente no estado de São Paulo e região do triângulo mineiro. De acordo com a Fundação (Fundo de Defesa da Citricultura), a incidência de CVC em laranjas teve um aumento aproximado de 22,09% em 1996 para 43,84% em 2004, tornando-se um dos maiores problemas para os citricultores. Em decorrência desse panorama, em 1997 iniciou-se um projeto visando sequenciar o genoma da *X. fastidiosa*, a fim de desenvolver estratégias para o controle da CVC, pois atinge todas as variedades de citros comerciais.



1. INTRODUÇÃO

O Brasil, impulsionado pelo crescimento das exportações e pelo desenvolvimento da indústria citrícola é hoje o maior produtor mundial de laranjas. O estado de São Paulo é responsável por 70% da produção nacional, com um volume que supera 400 milhões de caixas, portanto a cultura da laranja pode ser considerada uma grande fonte econômica quando comparadas às demais produções agrícolas. Mais de 70% das laranjas produzidas no estado de São Paulo abastecem as companhias produtoras de suco de laranja, representando 28% da produção de suco *in natura* (não concentrado) no mercado interno. A citricultura no estado de São Paulo é uma das mais importantes atividades econômicas, gerando 400 mil empregos e milhões de dólares em divisas. Para o Brasil, esse fato é muito relevante visto que o suco de laranja representa aproximadamente 10% das divisas geradas pelas exportações (<http://www.laranjabrasil.com.br>). Mesmo sendo um grande produtor de laranjas, o Brasil tem uma pequena presença no mercado de exportação de suco de laranja (*in natura*), esta baixa produtividade de suco de laranja (*in natura*) se dá por doenças, déficit de água e fatores nutricionais que comprometem as plantações e safras (<http://www.fundecitrus.com.br>).

No Brasil, a clorose variegada do citros (CVC) vem comprometendo a produção de laranjas, pondo em risco o sustento de várias famílias e prejudicando seriamente a economia do Brasil, principalmente no estado de São Paulo e região do triângulo mineiro. De acordo com a Fundecitrus (Fundo de Defesa da Citricultura), a incidência de CVC em laranjais teve um aumento aproximado de 22,09% em 1996 para 43,84% em 2004, tornando-se um dos maiores problemas para os citricultores. Em decorrência desse panorama, em 1997 iniciou-se um projeto visando seqüenciar o genoma da *X. fastidiosa*, a fim de desenvolver estratégias para o controle da CVC, pois atinge todas as variedades de citros comerciais.



A *X. fastidiosa* é uma bactéria gram-negativa e fastidiosa, que vive limitada ao xilema de plantas (WELL *et al.*, 1987; MOLLENHAUER & HOPKINS, 1976), que somente cresce em meios especiais, pois exige vários aminoácidos, micro e macronutrientes para garantir o seu desenvolvimento pleno (ALVES, 2003). A *X. fastidiosa* é o agente causal da CVC (LEE *et al.*, 1993; PURCELL & HOPKINS, 1996) que provoca o entupimento nos vasos do xilema responsáveis por levar água e nutrientes da raiz para a copa da planta. A produção do pomar afetado cai rapidamente, os frutos ficam duros, pequenos e amadurecem precocemente. A perda de peso do fruto pode chegar a 75% (<http://www.fundecitrus.com.br>). Outra característica importante de *X. fastidiosa*, e que deu origem ao nome “fastidiosa”, é a lenta velocidade de crescimento e a necessidade de um meio especial para seu desenvolvimento.

Embora todas as bactérias fastidiosas, gram-negativas e limitadas ao xilema tenham sido incluídas na espécie *X. fastidiosa*, existe variabilidade suficiente de linhagens que justifica a separação em subespécies. Essas diferentes linhagens atacam plantações causando doenças importantes do ponto de vista econômicos como: videiras (DP) (PURCELL & FINLAY, 1979), mirta, amêndoa, café (HENDSON *et al.*, 2001; VAN SLUYS *et al.*, 2003), alfafa anã, olmo plátano, amora (MONTEIRO *et al.*, 2001) e crucíferas, como repolho, rabanete, nabo (MUNIZ, 2003) e citros (CVC) (PURCELL & HOPKINS, 1996). Como ainda não há uma forma específica de combate a *X. fastidiosa*, os citricultores utilizam em seus pomares as estratégias de manejo para o controle da CVC (<http://www.fundecitrus.com.br>).

A CVC ou “amarelinho” foi primeiramente detectada em 1987, em pomares da região norte e sudoeste dos Estados de São Paulo e Minas Gerais (CHANG *et al.*, 1993), hoje está disseminada pelas principais regiões produtoras de citros do país.

Um dos principais sintomas da CVC é a clorose foliar na parte mediana e superior da copa, tomando posteriormente toda a planta, folhas com sintomas de deficiências nutricionais, principalmente de íons de zinco (Zn^{2+}) e potássio (K^+). Folhas com pontuações pardas



visíveis em ambos os lados do limbo foliar, estas áreas cloróticas podem tornar-se necróticas, chegando muitas vezes, a coalescer (ROSSETTI *et al.*, 1990).

Plantas severamente atacadas apresentam atrofia e diminuição da copa, muitas vezes adquirindo aspecto amarelado, podendo ainda ocorrer queda de folhas, morte de ponteiros, encurtamento de internódios e aumento de brotações nas gemas dos ramos afetados, causando a diminuição do porte das plantas (LARANJEIRA, 1997). Outro sintoma importante da CVC é a redução do tamanho e o enrijecimento dos frutos (BERETTA *et al.*, 1996; HENDSON *et al.*, 2001), pois apresentam alto teor de açúcar, tornando-os impróprios e sem valor para o consumo *in natura* ou para produção de suco na indústria (ROSSETTI *et al.*, 1990; LARANJEIRA, 1997). A figura 1 ilustra os principais sintomas causados pela CVC. Durante os meses de fevereiro a setembro, os sintomas foliares ficam mais evidentes devido ao estresse hídrico das plantas, provocado pela falta de chuva (DE NEGRI & GARCIA JUNIOR, 1994).



Figura 1 – Principais sintomas do CVC (<http://www.fundecitrus.com.br/doencas/cvc.html>).

As plantas em sua fase inicial de contaminação mostram sintomas da doença apenas em uma região específica da árvore, aquela onde houve contato com o vetor, mas após um período de infecção de cerca de três anos, a CVC já se espalhou por toda a árvore. A CVC ocorre em plantas de qualquer idade (MUNIZ, 2003; DE NEGRI & GARCIA JUNIOR, 1994). A figura 2 mostra um gráfico que apresenta os diferentes níveis de sintomas

observados em plantas com idades entre 0 a 2 anos, 3 a 5 anos, 6 a 10 anos e mais que 10 anos, de acordo com o levantamento realizado em 2004 pela Fundecitrus.

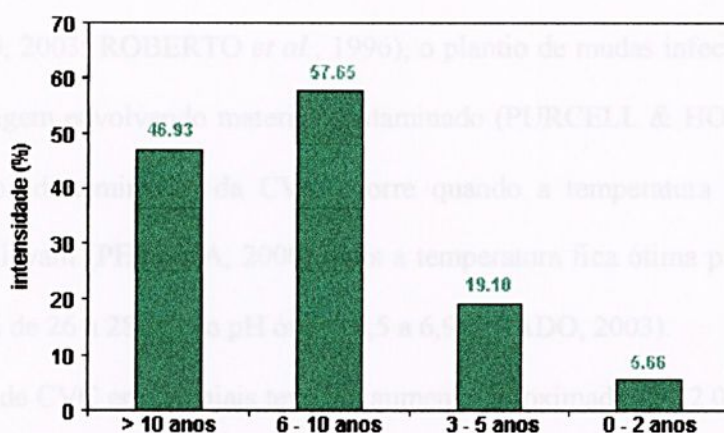


Figura 2 – Incidência de CVC por idade no estado de São Paulo e parte do Triângulo Mineiro

A CVC ataca todas as variedades comerciais de laranja doce (Pêra, Natal, Hamlin, Valência, Folha Murcha, Baianinha e Barão) sobre diferente portas-enxerto (Limão Cravo, Trifoliata, Tangerinas Cleópatra e Sumli, Laranja Caipira, etc). Entretanto, não foram encontrados sintomas nas tangerinas comerciais (Cravo e Poncan), Tangor Murcote, limões verdadeiros (Siciliano e Eureka) e Lima Ácida Galego, mesmo quando localizados em áreas altamente infectadas (SOUZA, 2002).

A disseminação da CVC no Brasil é feita através das cigarrinhas ou gafanhotos (PURCELL & HOPKINS, 1996) e até o momento foram identificadas 11 espécies destes insetos vetores de *X. fastidiosa* em citros no Brasil (HOPKINS, 1989; PURCELL & HOPKINS, 1996). A *X. fastidiosa* é encontrada no lúmen do canal alimentar dos insetos vetores, que com seu aparelho bucal sugador alimentam-se do xilema das plantas contaminadas e adquirem a bactéria, que passa a alojar no intestino e se reproduzir no interior do tubo digestivo do inseto vetor (BRLANSKY *et al.*, 1983; ROBERTO *et al.*, 1996). Uma vez infectados, os insetos vetores podem transmitir a *X. fastidiosa* até o fim de suas vidas

(HOPKINS, 1989; PURCELL & HOPKINS, 1996). Outro meio de transmissão é devido à comercialização de mudas e borbulhas contaminadas com a bactéria *X. fastidiosa*. O que contribui muito para aumento da CVC dentro dos pomares foi falta de controle dos insetos vetores (PRADO, 2003; ROBERTO *et al.*, 1996), o plantio de mudas infectadas e através de enxertia de garfagem envolvendo material contaminado (PURCELL & HOPKINS, 1996). O período de maior disseminação da CVC ocorre quando a temperatura média anual e a precipitação se elevam (PEREIRA, 2000), pois a temperatura fica ótima para o crescimento de *X. fastidiosa* é de 26 a 28 °C e o pH ótimo 6,5 a 6,9 (PRADO, 2003).

A incidência de CVC em laranjais teve um aumento aproximado de 22,09% em 1996 para 43,84% em 2004 nas regiões do estado de São Paulo e parte do triângulo mineiro, como podemos ver na figura 3.

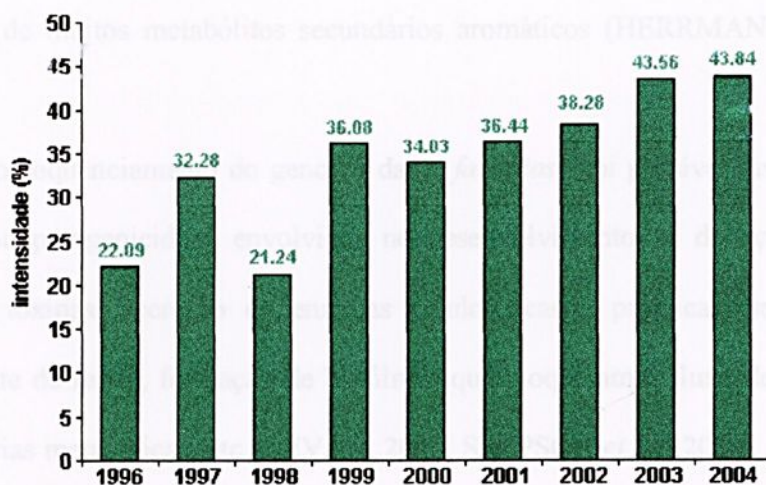


Figura 3 - Comparação da incidência de CVC no período de 1996 a 2004.

A CVC se tornou um dos maiores problemas para os citricultores, pondo em risco o sustento de várias famílias e prejudicando seriamente a economia do estado de São Paulo, triângulo mineiro e do Brasil. Outros dados sobre a CVC podem ser encontrados no site da Fundecitros: http://www.fundecitrus.com.br/est_cvc_br.html. Atualmente, o controle da CVC tem sido através de um programa de manejo adotado que envolve a redução do inoculo da

CVC, através de poda e/ou erradicação de plantas sintomáticas, o controle químico das cigarrinhas transmissoras, melhorias no sistema de irrigação em pomares de citros com sintomas iniciais da CVC e, principalmente, a utilização de mudas certificadas e sadias para a instalação dos pomares (<http://www.fundecitrus.com.br>).

A presente situação do ataque de CVC nos laranjais indica a necessidade do desenvolvimento de uma droga eficaz contra *X. fastidiosa*. Entre os possíveis alvos para o desenho de drogas contra a bactéria *X. fastidiosa* estão as enzimas da via metabólicas do ácido chiquímico, pois a via é essencial para sobrevivência de algas, plantas, bactérias, fungos e apicomplexos de parasitas, mas é ausente em animais (COGGINS *et al.*, 2003). A via metabólica do ácido chiquímico conecta o metabolismo de carboidratos a biossíntese de compostos aromáticos, numa sequência de sete passos metabólicos, eritrose-4-fosfato (E4P) e fosfoenolpiruvato (PEP) são convertidos a corismato, o precursor dos aminoácidos aromáticos e de muitos metabólitos secundários aromáticos (HERRMANN & WEAVER, 1999).

Através do sequenciamento do genoma da *X. fastidiosa* foi possível investigar genes de virulência e/ou patogenicidade envolvidos no desenvolvimento da doença CVC, como a produção de toxinas, secreção de enzimas celulolíticas e protéicas, sequestro de íons (principalmente de ferro), formação de biofilmes que bloqueiam o fluxo de seiva nos vasos xilemáticos, vias metabólicas, etc (BEVAN, 2000; SIMPSON *et al.*, 2000). Um aspecto que parece ser importante na patogenicidade da *X. fastidiosa* é a sua capacidade de movimentação através dos vasos do xilema, e colonização dos diferentes tecidos distantes do ponto de infecção (HOPKINS, 1989; PURCELL & HOPKINS, 1996). O genoma da *X. fastidiosa* foi totalmente sequenciado em estudos realizados por diversos laboratórios de pesquisa do Estado de São Paulo e financiado pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo – www.fapesp.com.br) e pela Fundecitrus (SIMPSON *et al.*, 2000).



O mecanismo pelo qual a bactéria *X. fastidiosa* causa a doença CVC ainda não é totalmente conhecido, mas muitos genes relacionados a patogenicidade foram encontrados no genoma da *X. fastidiosa* quando comparados aos de outras bactérias (SIMPSON *et al.*, 2000; LAMBAIS *et al.*, 2000; VAN SLUYS, 2003). Entre os fatores relacionados a patogenicidade é possível destacar os processos de interação célula-célula, degradação da parede celular da planta, homeostase de íons, resposta antioxidante e síntese de toxinas (MUNIZ, 2003).

Antes da publicação do genoma, as hipóteses sobre os mecanismos pelos quais a *X. fastidiosa* causava doenças em seus vários hospedeiros eram vagas e pouco exploradas. Hopkins (1989) cita a oclusão dos vasos xilemáticos pelas bactérias, tiloses e gomas; a produção de fitotoxinas e o desbalanço de reguladores como possíveis mecanismos de patogenicidade dessa bactéria. Nenhum experimento conseguiu provar nem excluir qualquer dessas teorias (SOUZA, 2002).

O bloqueio dos vasos xilemáticos por um biofilme de *X. fastidiosa* é um dos mecanismos de patogênese mais sugerido e pesquisado. A bactéria *X. fastidiosa* freqüentemente forma agregados no interior dos vasos do xilema, constituídos de uma matriz extracelular produzida pela bactéria *X. fastidiosa*, a qual lembra uma massa de fibras de polissacarídeos e é denominada glicocálice, desempenhando um papel importante na sobrevivência da *X. fastidiosa* (MUNIZ, 2003; SOUZA, 2002; ALVES, 2003; SCARPARI, 2001).

A partir da sequência genômica de *X. fastidiosa* foi traçado todo o perfil básico para a sobrevivência dessa bactéria como o seu metabolismo energético, a síntese de aminoácidos, nucleotídeos e lipídeos e a maquinaria de transcrição, tradução e reparo (SOUZA, 2002).

A *X. fastidiosa* é capaz de sintetizar uma variedade extensa de enzimas, co-fatores e grupos protéticos, inclusive biotina, ácido fólico, pantotenato e coenzima A, ubiquinona, glutatona, tiorredoxina, glutarredoxina, riboflavina, FMN, FAD, nucleotídeos de pirimidina, porfiria, tiamina, piridoxal 5'-fosfato e lipase (SIMPSON *et al.*, 2000).



O sequestro de íons ou homeostase de íons está presente no processo regulatório da expressão de genes envolvidos na patogenicidade em bactérias. Diversos genes responsáveis pela codificação de proteínas seqüestradoras de íons ferro que, posteriormente serão transportados através da membrana, sugerindo que estas proteínas têm um importante papel na sobrevivência da bactéria nos vasos xilemáticos, pois a concentração de íons ferro na bactéria é rigorosamente controlada na bactéria (SCARPARI, 2001).

Alguns íons, especialmente ferro, magnésio e manganês são essenciais para a sobrevivência de todos os organismos vivos, a redução da disponibilidade desses íons para a planta pode contribuir significativamente para o aparecimento das lesões nas folhas dos hospedeiros, como pode ser observado na figura 4. (SIMPSON *et al.*, 2000; MEIDANIS *et al.*, 2002).



Figura 4 - Folhas com sintomas da redução da disponibilidade dos íons
(<http://acg.lbi.ic.unicamp.br/xf/home/mmachado.html>).

Com o término do seqüenciamento do genoma de *Xylella fastidiosa*, as pesquisas passaram a exibir grandes avanços nos últimos anos, principalmente no Brasil. Com as informações (2,7 milhões de pb mapeados) geradas através do projeto genoma de *Xylella fastidiosa* novos conhecimentos estão disponíveis (SIMPSON *et al.*, 2000), podendo levar ao desenvolvimento de estratégias de controle da CVC minimizando assim as perdas causadas por esse patógeno.

1.1. Via Metabólica do Ácido Chiquímico

A via metabólica do ácido chiquímico é uma das rotas biossintéticas que leva à produção de alguns dos metabólitos primários e secundários importantes na natureza (GANEM, 1978; DE OLIVEIRA, 2001). O ácido chiquímico (Figura 5) foi primeiramente isolado por Eykmann, em 1885, a partir da fruta da planta oriental *Illicium religiosum*, sendo seu nome derivado do nome em japonês para esta planta, shikimi-no-ki (CAMPBELL *et al.*, 1993). Entretanto, sua estrutura permaneceu desconhecida por aproximadamente 50 anos. Somente na década de 30, graças aos esforços pioneiros de Fisher, Freudenberg e Karrer, a estrutura completa e a estereoquímica absoluta deste ácido carboxílico foi determinada (GANEM, 1978).

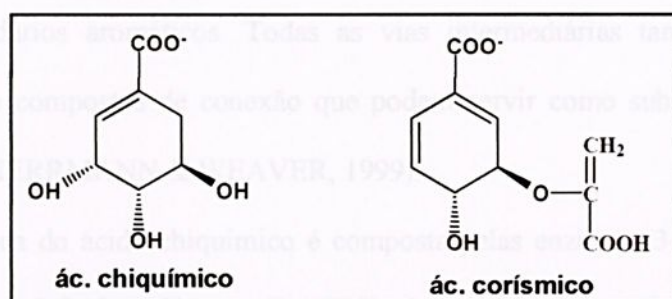


Figura 5 – Estruturas químicas do ácido chiquímico e do ácido corísmico.

Estudos subsequentes de Bernhard Davis (1950) e de outros pesquisadores como David Sprinson e Frank Gibson nos 25 anos seguintes (GIBSON, 1999) revelaram uma rota bem definida de oito metabólitos intermediários que conduziam, a partir da glicose, primeiro para o ácido chiquímico, e depois ao ácido corísmico (Figura 5). Estas substâncias constituíam uma rota biossintética comum que seria re-direcionada para a formação de vários metabólitos aromáticos mono e policíclicos em plantas e microrganismos por diferentes caminhos (GANEM, 1978).

A via metabólica do ácido chiquímico conecta o metabolismo de carboidratos a biossíntese de compostos aromáticos, numa sequência de sete etapas metabólicas. A etapa inicial é a condensação do PEP com E4P. Na segunda etapa o 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato (DAHP) perde seu grupamento fosforila e cicliza-se para 3-desidroquinato. A desidratação fornece então 3-desidrochiquimato, que é reduzido usando como co-fator nicotinamida adenina dinucleotídeo (NADH) a D-chiquimato. Na quinta etapa ocorre uma específica fosforização do grupo 3-hidroxil da D-chiquimato usando adenosina trifosfato (ATP), como co-fator. Na sexta etapa uma segunda molécula de PEP condensa-se com chiquimato 3-fosfato (S3P) para produzir 5-enoilpiruvil 3-fosfato. Na etapa final 5-enoilpiruvil 3-fosfato perde seu grupamento fosforila, produzindo corismato (Figura 6), o precursor dos aminoácidos aromáticos (Phe, Trp e Tyr) (Figuras 7) e de muitos metabólitos secundários aromáticos. Todas as vias intermediárias também podem ser consideradas como compostos de conexão que podem servir como substrato para outras vias metabólicas (HERRMANN & WEAVER, 1999).

A via metabólica do ácido chiquímico é composta pelas enzimas: 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato Sintase (DAHPS), 3-Desidroquinato Sintase (DHQS), 3-Desidroquinato Desidratase (DHQD), Chiquimato-5-Desidrogenase (SD), Chiquimato Quinase (SK), 5-Enoilpiruvilchiquimato 3-fosfato Sintase (EPSPS) e Corismato Sintase (CS).

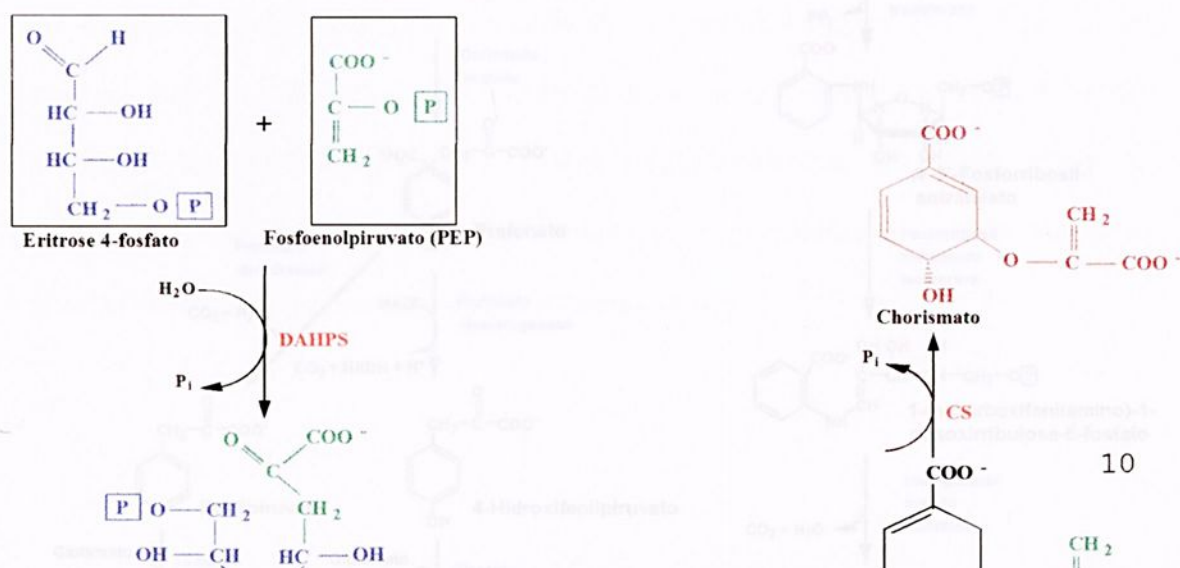
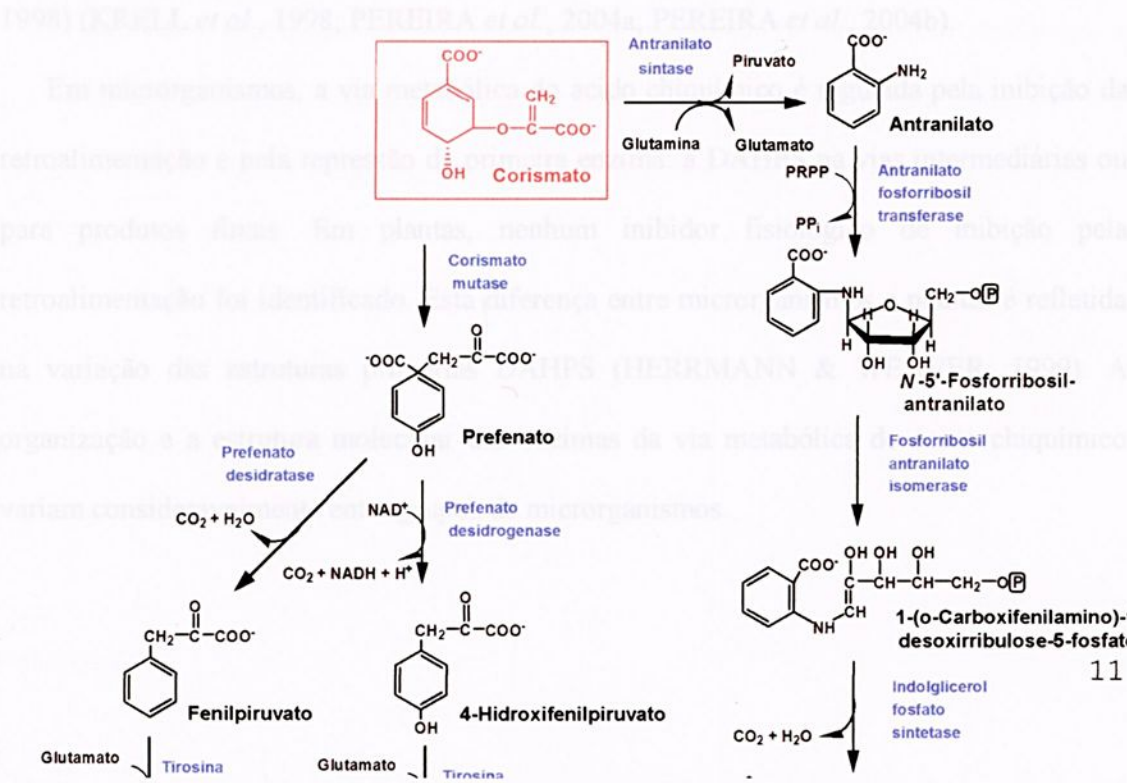


Figura 7 – Biosíntese dos aminoácidos aromáticos a partir do corismato. O corismato e os aminoácidos aromáticos estão indicados por quadros, e as enzimas estão indicadas em azul. (DE OLIVEIRA, 2001).

A via metabólica do ácido chiquímico é essencial para algas, plantas, bactérias e fungos, mas é ausente em animais, que adquirem os aminoácidos aromáticos através da sua alimentação, que são aminoácidos essenciais para os animais (COGGINS *et al.*, 2003).

Figura 6 – A via metabólica do ácido chiquímico. As sete enzimas estão indicadas à esquerda em vermelho.



As bactérias têm seis polipeptídeos individuais, cada um possui uma única atividade enzimática, e são codificados por genes separados. As plantas têm um arranjo molecular similar às bactérias, com a exceção do DHQS e SD, que possuem domínios separados em um polipeptídeo bifuncional (CAMPBELL *et al.*, 2004; HERRMANN & WEAVER, 1999). Em contraste, todos os fungos examinados têm as enzimas DAHPS e CS como polipeptídeos monofuncionais e o restante das enzimas como um polipeptídeo pentafuncional denominado AROM. O polipeptídeo AROM tem os domínios análogos às enzimas bacterianas: DHQS, EPSPS, SK, DHQS e SD (CAMPBELL *et al.*, 2004). As enzimas DHQS e EPSPS formam o domínio N-terminal; SK, DHQS e SD formam o domínio C-terminal.

Figura 7 – Biossíntese dos aminoácidos aromáticos a partir do corismato. O corismato e os aminoácidos aromáticos estão indicados por quadros, e as enzimas estão indicadas em azul. (DE OLIVEIRA, 2001).

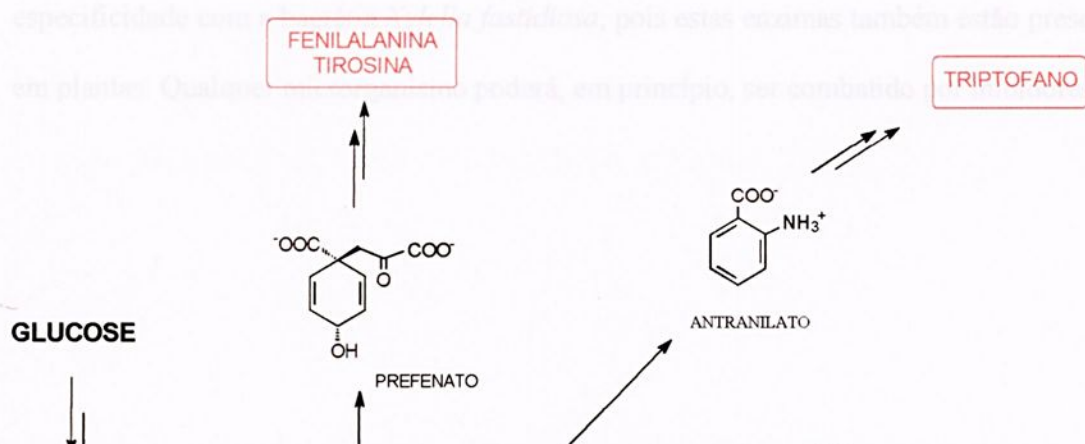
A via metabólica do ácido chiquímico é essencial para algas, plantas, bactérias e fungos, mas é ausente em animais, que adquirem os aminoácidos aromáticos através da sua alimentação, que são aminoácidos essenciais da dieta dos animais (COGGINS *et al.*, 2003). Por causa disso as enzimas pertencentes a esta via representam importantes alvos para o desenvolvimento de herbicidas (KISHORE & SHAH, 1988), antimicrobianos potencialmente não tóxicos (DAVIES *et al.*, 1994) e drogas antiparasitas (ROBERTS *et al.*, 1998) (KRELL *et al.*, 1998; PEREIRA *et al.*, 2004a; PEREIRA *et al.*, 2004b).

Em microrganismos, a via metabólica do ácido chiquímico é regulada pela inibição da retroalimentação e pela repressão da primeira enzima: a DAHPS nas vias intermediárias ou para produtos finais. Em plantas, nenhum inibidor fisiológico de inibição pela retroalimentação foi identificado. Esta diferença entre microrganismos e plantas é refletida na variação das estruturas primárias DAHPS (HERRMANN & WEAVER, 1999). A organização e a estrutura molecular das enzimas da via metabólica do ácido chiquímico variam consideravelmente entre grupos de microrganismos.

As bactérias têm sete polipeptídios individuais, cada um possui uma única atividade enzimática, e são codificados por genes separados. As plantas têm um arranjo molecular similar às bactérias, com a exceção do DHQD e SD, que possuem domínios separados em um polipeptídio bifuncional (CAMPBELL *et al.*, 2004; HERRMANN & WEAVER, 1999). Em contraste, todos os fungos examinados têm as enzimas DAHPS e CS como polipeptídios monofuncional e o restante das enzimas como um polipeptídio pentafuncional denominado AROM. O polipeptídio AROM tem os domínios análogos às enzimas bacterianas: DHQS, EPSPS, SK, DHQD e SD (CAMPBELL *et al.*, 2004). As enzimas DHQS e EPSPS formam o domínio N-terminal e SK, DHQD e SD formam o domínio C-terminal da proteína. A proteína de AROM está relacionada à regulação da transcrição eucariótica (CARPENTER *et al.*, 1998; HERRMANN & WEAVER, 1999).

Um aspecto evolucionário muito interessante do metabolismo aromático em fungos é uma duplicação do gene *arom* que resulte nos genes que codificam proteínas regulatórias para a degradação do quinato (HERRMANN & WEAVER, 1999). Os complexos AROM foram estudados de *Neurospora crassa*, *Aspergillus nidulans* (CARPENTER *et al.*, 1998), *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, e *Pneumocystis carinii* (HERRMANN & WEAVER, 1999).

O corismato é o último produto comum da via metabólica do ácido chiquímico e precursor chave para a biossíntese de PABA, ácido *p*-hidroxibenzóico, micobactinas e dos aminoácidos aromáticos essenciais (Figuras 7 e 8) (ROBERTS *et al.*, 2002; DOSSELAERE & VANDERLEYDEN, 2001).



enzimas da via metabólica do ácido chiquímico que impossibilitam a produção do composto.

Como mostrado acima, a via se difere de um organismo para outro, então foi colocada abaixo uma tabela com algumas informações sobre as enzimas da via metabólica do ácido chiquímico de *X. fastidiosa* como massa molecular, número de aminoácidos, gene de codificação e o número de classificação (E.C.), pois foi o organismo estudado neste trabalho.

Tabela 1 – Informações sobre as enzimas da via metabólica do ácido chiquímico de *X. fastidiosa*.

Enzimas	Massa Molecular KDa	Número de Aminoácidos	Gene	E.C.	Código Acesso Swiss-Prot®
DAHPS	41,4	389	—	4.1.2.13	Q9PHB8
DHQS	42,1	394	<i>aroH</i>	4.2.3.4	Q9PDPS
DHQS	15,9	148	<i>aroQ</i>	4.2.1.10	Q9PH07
SD	39,2	382	<i>aroE</i>	1.1.1.25	Q9PBN3
SK	29,3	480	<i>aroK</i>	2.7.1.71	Q9PD04
EPSPS	48,2	454	<i>aroA</i>	2.5.1.19	Q9PH21

Figura 8 – O corismato e os compostos essenciais sintetizados a partir do corismato estão destacados por quadros em vermelho (DE OLIVEIRA, 2001).

As bactérias gastam mais de 90 % da sua energia metabólica total na biossíntese de proteína. Conseqüentemente, a via metabólica do ácido chiquímico em bactérias serve quase exclusivamente para sintetizar aminoácidos aromáticos (HERRMANN, 1995a). O bloqueio de qualquer uma das sete etapas da via metabólica do ácido chiquímico por meio da inibição de uma de suas enzimas específicas poderia, em princípio, levar a bactéria à morte pela não disponibilização de metabólitos essenciais para a sua sobrevivência (HERRMANN, 1995a). Contudo qualquer novo inibidor para estas enzimas deve ter alta especificidade com a bactéria *Xylella fastidiosa*, pois estas enzimas também estão presentes em plantas. Qualquer microrganismo poderá, em princípio, ser combatido por inibidores das (*aroH*) (HERRMANN, 1995; HERRMANN & WEAVER, 1999).



enzimas da via metabólica do ácido chiquímico que impossibilitarão a produção do corismato.

Como mostrado acima, a via se difere de um organismo para outro, então foi colocada abaixo uma tabela com algumas informações sobre as enzimas da via metabólica do ácido chiquímico de *X. fastidiosa* como massa molecular, número de aminoácidos, gene de codificação e o número de classificação (E.C.), pois foi o organismo estudado neste trabalho.

Tabela 1 – Informações sobre as enzimas da via metabólica do ácido chiquímico de *X. fastidiosa*.

Enzimas	Massa Molecular KDa	Número de Aminoácidos	Gene	E.C.	Código Acesso Swiss-Prot*
DAHPS	41,4	380	-----	4.1.2.15	Q9PHB8
DHQS	42,3	394	<i>aroB</i>	4.2.3.4	Q9PDP5
DHQD	15,9	148	<i>aroQ</i>	4.2.1.10	Q9PH97
SD	30,2	282	<i>aroE</i>	1.1.1.25	Q9PFN3
SK	20,3	180	<i>aroK</i>	2.7.1.71	Q9PDP4
EPSPS	48,2	454	<i>aroA</i>	2.5.1.19	Q9PB21
CS	39,8	372	<i>aroC</i>	4.2.3.5	Q9PDL0

*(Bairoch *et al.*, 2004).

1.1.1. DAHPS

A enzima DAHPS (Figura 9) é uma metaloenzima que catalisa a primeira etapa na biossíntese de aminoácidos aromáticos. A enzima DAHPS também é conhecida como 2-deidro-3-deoxifosfoheptonato sintase e catalisa a condensação do PEP com o quarto carbono de açúcar de D-eritrose-4-fosfato (E4P) para produzir DAHP e fosfato inorgânico (P_i) (HERRMANN & WEAVER, 1999; HERRMANN, 1995b). A enzima DAPHs de bactéria mais investigada foi a de *Escherichia coli* (*Ec*), que codifica três isoenzimas do DAHPS, uma sensível a Phe (*aroG*), uma sensível a Tyr (*aroF*) e uma sensível a Trp (*aroH*) (HERRMANN, 1995b; HERRMANN & WEAVER, 1999).



A natureza do íon do metal pode variar e pode depender das circunstâncias sob as quais as bactérias foram crescidas. Os co-fatores metálicos preferenciais parecem ser Fe^{2+} e talvez Zn^{2+} (HERRMANN, 1995b). O íon de metal tem função na catálise e pode também afetar a conformação da enzima. O domínio de sítio dirigido se encontra no Cys-X-X-His, perto do N-terminal do polipeptídeo, e foi identificado como um ligante do íon metálico para isoenzima sensível a Phe da *E. coli* (STEPHENS & BAUERLE, 1992; HERRMANN, 1995b).

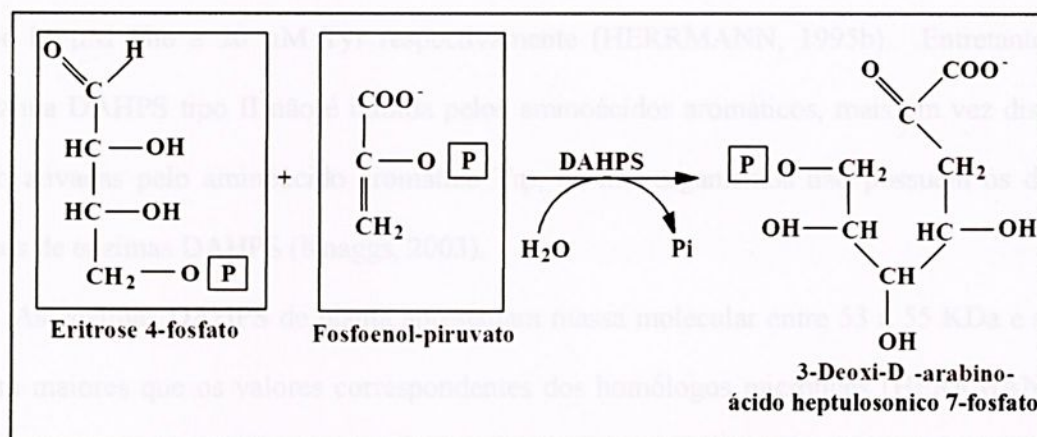


Figura 9 – A reação catalisada pela enzima DAHPS.

A comparação da sequência primária de aminoácidos entre a enzima DAHPS de planta e DAPHIS de bactéria mostra surpreendentemente uma baixa identidade de 20%. De fato, PRETTYBOX, um programa usado para fazer alinhamentos múltiplos de sequência, identificou somente 24 resíduos invariantes para todas as enzimas de DAHPS conhecida (HERRMANN & WEAVER, 1999; HERRMANN, 1995b).

Existem dois tipos de enzimas DAHPS (tipo I e II), Walker e seus colaboradores (1960) fizeram uma distinção entre uma enzima DAHPS de bactéria do tipo I de 39 KDa e de uma enzima DAHPS do tipo II de planta 54 KDa. Entretanto, uma enzima DAHPS do tipo II foi encontrada também na bactéria *Streptomyces rimosus* e os dados parciais

desta sequência para esta proteína mostram similaridade substancial à enzima DAHPS da planta, então foi verificado que a enzima DAPHS tipo II não era exclusiva de planta sendo encontrada também nos procariotos e DAHPS tipo I de bactéria em microrganismos eucarióticos (HERRMANN & WEAVER, 1999).

A enzima DAHPS tipo I é inibida pelos aminoácidos aromáticos (Phe, Trp, Tyr), vai depender da sensibilidade de cada enzima DAHPS e de microorganismo para microorganismo, isto não acontece para enzimas DAHPS em plantas (HERRMANN, 1995a; Knaggs, 2003), por exemplo, isoenzimas sensíveis a Phe e Trp são 50% inibidas por 13 μM Phe e 20 μM Tyr respectivamente (HERRMANN, 1995b). Entretanto a enzima DAHPS tipo II não é inibida pelos aminoácidos aromáticos, mais em vez disso, são ativadas pelo aminoácido aromático Trp, os microrganismos não possuem os dois tipos de enzimas DAHPS (Knaggs, 2003).

As enzimas DAHPS de planta apresentam massa molecular entre 53 a 55 KDa e são bem maiores que os valores correspondentes dos homólogos microbiais (HERRMANN, 1995b), por exemplo, o DAHPS da bactéria *Xylella fastidiosa* possui massa molecular 41,4 KDa. Os valores de K_m (constante de Michaelis-Menten) para o PEP e E4P são 30 e 70 μM , respectivamente, similares aos valores obtidos para a enzima DAHPS de microrganismos (HERRMANN, 1995a). Este trabalho descreve a modelagem da enzima DAHPS de *Xylella fastidiosa* (XfDAHPS) em complexo com o íon manganês (II) (Mn^{2+}) e o 2-Fosfoglicolato (PGL), um possível inibidor.

1.1.2. DHQS

A segunda enzima da via metabólica do ácido chiquímico, DHQS (Figura 10), é ativada por íons P_i , um dos produtos da reação, e requer também quantidades catalíticas de Nicotinamida Adenina Dinucleotídeo (NAD^+) e de um cátion divalente para a



atividade, mesmo que a reação catalisada pela enzima seja um redox neutro (HERRMANN & WEAVER, 1999; HERRMANN, 1995a). A enzima DHQS é uma metalenzima dependente de NAD^+ que converte DAHP ao desidroquinato (DHQ) (CARPENTER *et al.*, 1998). Em bactérias o Co^{2+} e Zn^{2+} são os íons metálicos mais ativos. A enzima DHQS de planta é a única enzima da via que ainda não possui nenhum dado estrutural disponível (HERRMANN, 1995b). Em fungos esta enzima é codificada em um único polipeptídeo pentafuncional chamado AROM (CARPENTER *et al.*, 1998; HERRMANN & WEAVER, 1999). A segunda reação da via metabólica do ácido chiquímico é catalisada pela enzima DHQS onde há a eliminação de um fosfato do DAHP para gerar o DHQ.

A conversão do DAHP a DHQ prossegue por uma troca intramolecular de oxigênio do anel do DAHP com carbono sete, dirigida pela clivagem do fosfoéster. A reação envolve uma oxidação e uma redução, na β -eliminação do P_i , uma abertura do anel, e uma condensação intramolecular do aldol.

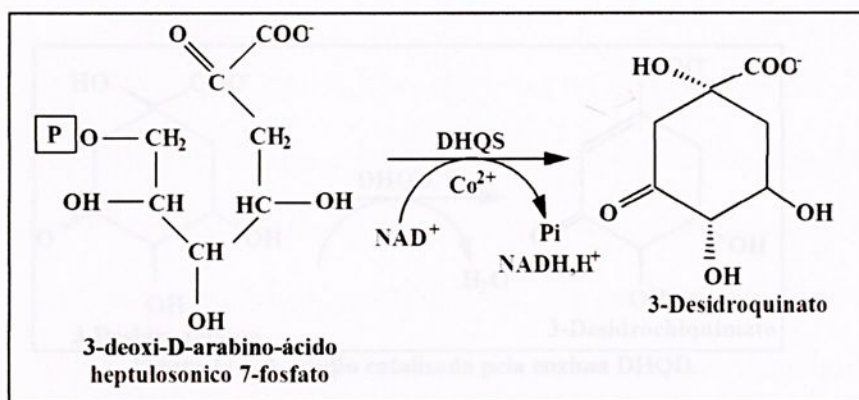


Figura 10 – A reação catalisada pela enzima DHQS.

A enzima *X*/DHQS foi modelada em complexo com [1R-(1 α , 3 β , 4 α , 5 β)]-5-(fosfometil)-1,3,4-trihidroxíciclohexano-1-ácido carboxílico (CRB) um possível inibidor e NAD⁺.

1.1.3. DHQD

A terceira etapa da via metabólica do ácido chiquímico, a desidratação do DHQ resultando na síntese 3-desidrochiquimato (Figura 11), é catalisada pela enzima DHQD. Esta enzima existe em dois tipos diferentes (tipo I e II) (HERRMANN & WEAVER, 1999). Esta reação de desidratação do DHQ resultando na síntese 3-desidrochiquimato é comum também para a via catabólica quinato (COGGINS *et al.*, 2003).

A enzima DHQD tipo I opera na via metabólica do ácido chiquímico e o DHQD tipo II tem as duas funções: de biossíntese e catabólica. Os dois tipos de enzima DHQD têm arquiteturas completamente diferentes uma da outra (COGGINS *et al.*, 2003). Em plantas esta enzima é bifuncional (DHQD – SD) (HERRMANN, 1995a) e os fungos têm duas enzimas DHQD com baixa similaridade aparente de sequência (GILES *et al.*, 1985).

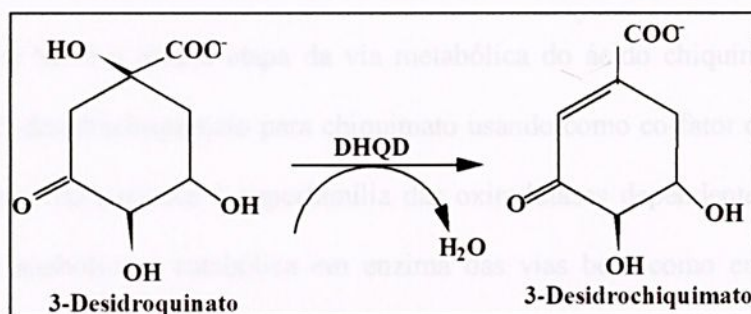


Figura 11 – A reação catalisada pela enzima DHQD.

A enzima DHQD tipo I é geralmente encontrada em bactérias e plantas (HERRMANN, 1995a), sua topologia consiste em um barril α/β de 8 fitas enquanto a

enzima DHQD tipo II consiste em um núcleo folha β com 5 fitas flanqueado por 4 α -hélices. A enzima DHQD tipo I catalisa uma reação *cis* (*syn*)-desidratação que envolve a perda do hidrogênio 2-pro-R, enquanto a enzima DHQD do tipo II catalisa uma reação *trans* (*anti*)-desidratação que resulta na perda do hidrogênio 2-pro-S (COGGINS *et al.*, 2003). As diferenças entre estes mecanismos estão refletidas nas estruturas destas proteínas; não há similaridade de sequência entre as enzimas DHQD tipo I e II, um exemplo raro da evolução convergente (HERRMANN & WEAVER, 1999). No geral, a enzima DHQD do tipo II é que participa da degradação do 3-desidroquinato (HERRMANN, 1995a).

A ocorrência de duas formas distintas da enzima DHQD (mecanismo e estrutura) viabiliza a possibilidade de desenvolvimento de inibidores seletivos para os dois tipos de enzimas DHQD (COGGINS *et al.*, 2003). A enzima *Xj*DHQD é do tipo II, segundo a avaliação da estrutura secundária. Assim, foi modelada a enzima *Xj*DHQD em complexo com o ácido 2,3-anidro-quinico (FA1), um possível inibidor.

1.1.4. SD

A enzima SD é a quarta etapa da via metabólica do ácido chiquímico e catalisa a redução do 3-desidrochiquimato para chiquimato usando como co-fator o NADH (Figura 12). A enzima SD pertence à superfamília das oxiredutases dependentes de NAD(P)H, com função anabólica e catabólica em enzima das vias bem como em destoxificação xenobiótica. Esta superfamília é geralmente subdividida em diversas famílias, incluindo desidrogenases de cadeias curtas, médias e longas, redutases aldo-queto e desidrogenases de álcool ferro-ativado. O mecanismo de reação usado pela enzima SD envolve a catálise da transferência de hidreto para o co-fator NAD(P)H, baseada na estabilização do



acúmulo de carga negativa no substrato aceitador de hidreto, por uma cadeia lateral de um aminoácido agindo como uma base geral doadora de prótons ou um cátion prostático no sítio ativo (BENACH *et al.*, 2003).

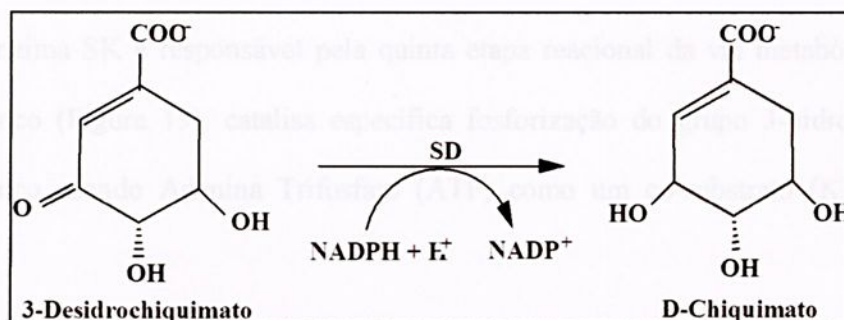


Figura 12 – A reação catalisada pela enzima SD.

Existem dados da existência de duas famílias independentes de desidrogenases de quinato/chiquimato identificadas. A primeira consiste em desidrogenases dependentes de NAD, e a segunda consiste nas desidrogenases associadas à membrana que utilizam o pirrolo-quinolina-quinona como um co-fator (ROSZAK *et al.*, 2002). Ambos os tipos de desidrogenases são envolvidas na via catabólica do quinato, que permite o crescimento dos microrganismos com quinato como a única fonte do carbono por sua conversão no protocatequato e no metabolismo subsequente pela via do β -quetoadipato (ROSZAK *et al.*, 2002).

As desidrogenases geralmente formam oligômeros, mas na maioria das bactérias está na forma monomérica, como por exemplo, em *Escherichia coli*. Em plantas a enzima SD está associada com a enzima DHQD tipo I e assim formam uma enzima bifuncional, ao passo que, em fungos como *Neurospora crassa*, esta enzima SD forma um dos cinco domínios de um polipeptídeo pentafuncional (AROM), o qual catalisa 5 das 7 etapas da via metabólica do ácido chiquímico. Entretanto, a base molecular para reconhecimento e a

redução enzimática não são ainda conhecidas (MICHEL *et al.*, 2003). A enzima SD de *Xylella fastidiosa* (XfSD) foi modelada junto com o co-fator NAD^+ .

1.1.5. SK

A enzima SK é responsável pela quinta etapa reacional da via metabólica do ácido chiquímico (Figura 13), catalisa específica fosforização do grupo 3-hidroxil do ácido chiquímico usando Adenina Trifosfato (ATP) como um co-substrato (KRELL *et al.*, 1998).

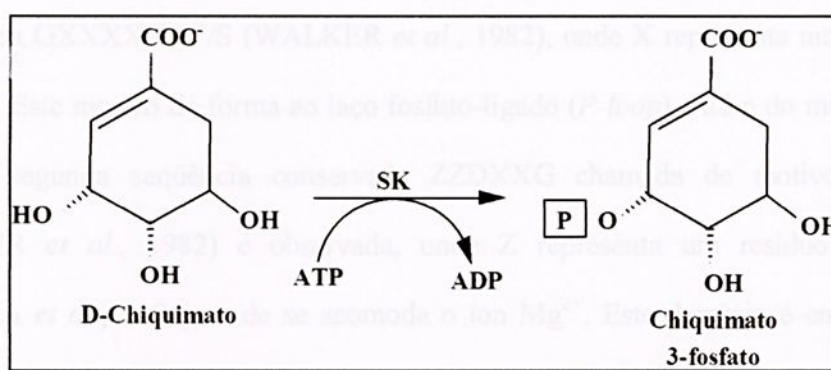


Figura 13 – Reação catalisada pela enzima SK.

As três estruturas de SK de *Erwinia chrysanthemi* (EchSK) resolvidas por cristalografia (Krell *et al.*, 1998; Krell *et al.*, 2001), ao qual é codificada pelo *aroL*, mostra que a SK pertence à família de quinases do monofosfato de nucleosídeo (NMP). Uma característica das quinases de NMP é que se submetem a mudanças conformacionais grandes durante a catálise (VONRHEIN *et al.*, 1995). Há duas regiões flexíveis na estrutura que são responsáveis pelos movimentos: o sítio de ligação NMP e o domínio LID. Na SK, o domínio chiquimato-ligado (SB) corresponde ao domínio de NMPB de quinases NMP (PEREIRA *et al.*, 2004b).

As quinases de NMP (YSAI & YAN, 1999) são compostas de três domínios: o NÚCLEO (*CORE*), a TAMPA (*LID*), e o domínio de NMP-ligado (NMPB). O domínio *CORE* contém um laço fosfato-ligado (*P-loop*) altamente conservado, o domínio *LID* fecha o sítio de ATP e carrega resíduos funcionalmente essenciais, e o domínio NMPB é responsável pelo reconhecimento e ligação específica de NMP (GU *et al.*, 2002).

Os três motivos funcionais do nucleotídeo-ligado das enzimas são reconhecidos para SK de *M. tuberculosis* (PEREIRA *et al.*, 2004; GU *et al.*, 2002), incluindo o motivo Walker A, um motivo Walker B e um laço adenina-ligado. O motivo Walker A é localizado entre a primeira folha β ($\beta 1$) e a primeira hélice α ($\alpha 1$), contendo a sequência conservada GXXXXGKT/S (WALKER *et al.*, 1982), onde X representa um aminoácido qualquer. Este motivo dá forma ao laço fosfato-ligado (*P-loop*). Além do motivo Walker A, uma segunda sequência conservada ZZDXXG chamada de motivo Walker B (WALKER *et al.*, 1982) é observada, onde Z representa um resíduo hidrofóbico (PEREIRA *et al.*, 2004), onde se acomoda o íon Mg^{2+} . Este domínio é encontrado em todas as enzimas SK conhecidas e também em diversas proteínas que ligam ATP e GTP (KRELL *et al.*, 1998). Todas estas proteínas que contêm a laço de ligação do fosfato (*P-loop*) apresentam uma estrutura em seu domínio central formada de quatro fitas β paralelas e por dois pares de hélices α -flanqueadas (MILNER-WHITE *et al.*, 1991).

Em *Escherichia coli*, a SK é encontrada em duas isoformas: a enzima SK tipo I, codificada pelo gene *aroK*, e a enzima SK tipo II, codificada pelo gene *aroL*, apresentando uma identidade de sequência de 30% e similaridade de 46% entre elas. A ocorrência de duas isoformas não é comum para enzimas no meio de vias biossintéticas e sugere que a SK pode ser um ponto de ramificação para duas diferentes vias (KRELL *et al.*, 1998).

A grande diferença de afinidade observada das duas isoformas em relação ao chiquimato sugere que a enzima SK tipo II seja a isoenzima principal na biossíntese do corismato, e tem um K_m de 0,2 mM, enquanto que a enzima SK tipo I tem um K_m de 20 mM e possivelmente exerça uma função biológica alternativa, só fosforilando o chiquimato esporadicamente (DE FEYTER & PITTARD, 1986).

A maioria das SKs em bactérias parecem ser codificadas pelo *aroK*, porque suas seqüências de aminoácido têm maior identidade com a SK de *E. coli* tipo I (GU *et al.*, 2002). A função fisiológica da enzima SK tipo I não está claro, mas a presença de um sítio de ligação de fosfato, necessário para suas funções fosfatase/fosfotransferase, observado tanto na enzima SK tipo II quanto na enzima SK tipo I. Ambas as enzimas SK tipo I e II apresentam o motivo Walker A (DE FEYTER & PITTARD, 1986).

A atividade enzimática apresenta dependência por cátions divalentes como co-fatores. O Mg^{+2} é o cátion mais efetivo, mas significantes atividades são obtidas na presença de outros cátions divalentes. Na presença dos íons Fe^{+2} e Ca^{+2} , a enzima SK tipo II apresenta somente 50% da atividade obtida com Mg^{+2} (DE FEYTER & PITTARD, 1986). A enzima SK de *Xylella fastidiosa* é classificada como do tipo I por apresentar maior similaridade com a enzima SK tipo I do que com a enzima SK tipo II. Foi modelada a enzima SK de *Xylella fastidiosa* (XfSK) com Adenina Difosfato (ADP) e o Ácido Chiquímico e o íon magnésio (Mg^{+2}) no sítio ativo.

1.1.6. EPSPS

A enzima EPSPS catalisa o penúltimo passo na parte pré-corismato da via metabólica do ácido chiquímico. A reação envolve a transferência reversível de um grupo enolpiruvil do PEP para o chiquimato-3-fosfato (S3P) produzindo 5-enoilpiruvilchiquimato-3-fosfato (EPSP), mais P_i (BENTLEY, 1990; DE OLIVEIRA, 2001) (Figura 14). A reação



catalisada pela EPSPS foi a melhor estudada de todas as enzimas da via metabólica do ácido chiquímico. Este enzima foi purificada de fontes microbiais e de planta (DUNCAN *et al.*, 1984; MOUSEDALE & COGGINS, 1984) e o mecanismo de reação enzima-catálise foi estudado em detalhe.

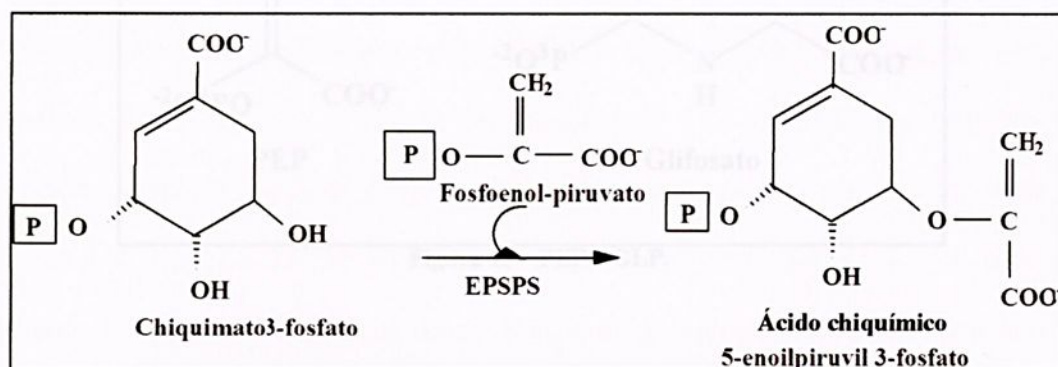


Figura 14 – Reação catalisada pela enzima EPSPS.

A EPSPS é inibida pelo glifosato (GLP) (Figura 15), substância extensamente usada como herbicida de largo espectro (STEINRÜCKEN & AMRHEIN, 1980). O GLP inibe a reação catalisada pela EPSPS atuando como um inibidor competitivo reversível em relação ao PEP e um inibidor não competitivo em relação ao S3P (ESCHENBURG *et al.*, 2002). Este intermediário metabólico (PEP) é substrato não somente para a EPSPS, mas também para DAHPS que não é inibida pelo glifosato (HERRMANN, 1995a). A tolerância do glifosato nas plantas foi obtida pela expressão do alelo mutante codificando uma enzima bacteriana menos sensível (STALKER *et al.*, 1985) ou pela superprodução de EPSPS (SMART *et al.*, 1985; SHAH *et al.*, 1986; GOLDSBROUGH *et al.*, 1990) (HERRMANN, 1995a).

O mecanismo de ação do glifosato na enzima EPSPS foi observado por meio de estudos de cinética enzimática em estado estacionário das enzimas EPSPS de diferentes organismos (ESCHENBURG *et al.*, 2002; DE OLIVEIRA, 2001). A enzima EPSPS da

planta tem os valores de K_i para o glifosato de aproximadamente uma ordem de grandeza mais baixa do que os valores para a enzima EPSPS de microrganismos (HERRMANN, 1995a).

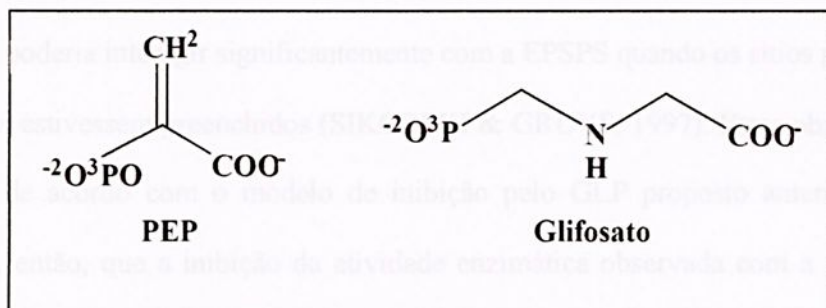


Figura 15 – PEP e GLP.

Os primeiros estudos sobre o mecanismo de inibição sugeriram que o glifosato deveria funcionar como um análogo ao estado de transição (ESCHENBURG *et al.*, 2002). Estudos de cinética enzimática da EPSPS de *E. coli* e de *Neurospora crassa* (BOOCOCK & COGGINS, 1983), medidas de fluorescência em equilíbrio (ANDERSON *et al.*, 1988) e calorimetria de varredura diferencial (MERABET *et al.*, 1993) sustentaram esta teoria, sugerindo que o GLP formaria, preferencialmente, um complexo ternário muito estável com a enzima (EPSPS) e o substrato (S3P), ocupando o sítio de ligação do PEP (EPSPS:S3P:GLP).

No entanto, estudos posteriores de cinética em estado estacionário (SAMMONS *et al.*, 1995), análise de análogos sintéticos do complexo ternário EPSPS:S3P:GLP (MARZABADI *et al.*, 1996), e mutagênese sítio-direcionada de possíveis resíduos do sítio ativo (PADGETTE *et al.*, 1991; MAJUMDER *et al.*, 1995) levaram à conclusão de que, na presença de S3P, o sítio de ligação do GLP não seria exatamente o mesmo do PEP. Além disto, o GLP poderia formar um complexo ternário EPSPS:EPSP:GLP,

demonstrando que o grupo carboxivinil no EPSP não impediria a ligação do inibidor e, portanto, não ocuparia o sítio de ligação deste (DE OLIVEIRA, 2001).

O GLP também não impediria a ligação de Pi ao complexo binário EPSPS:EPSP. Um complexo quaternário EPSPS:EPSP:GLP:Pi seria formado nesta situação, indicando que o glifosato poderia interagir significativamente com a EPSPS quando os sítios para ambos os substratos estivessem preenchidos (SIKORSKI & GRUYS, 1997). Estas observações não estavam de acordo com o modelo de inibição pelo GLP proposto anteriormente. Foi proposto, então, que a inibição da atividade enzimática observada com a formação dos complexos ternários EPSPS:S3P:GLP e EPSPS:EPSP:GLP deveria ser causada pela ligação do inibidor a resíduos de aminoácidos que não estivessem diretamente envolvidos na catálise ou na associação com os substratos, e que parte da molécula de GLP deve ligar próxima, mas fora do sítio-ativo da EPSPS (SAMMONS *et al.*, 1995). Desta forma, o glifosato ao que parece funcionaria como um inibidor alostérico que competiria indiretamente com a ligação do PEP, por causar mudanças conformacionais na EPSPS quando a ela ligado (SIKORSKI & GRUYS, 1997).

Com a determinação da estrutura cristalina da EPSPS de *E. coli* (STALLINGS *et al.*, 1991) não ligada aos substratos, com uma resolução média de 3 Å, mostrou que a estrutura apresenta dois domínios globulares simétricos separados por uma fenda. Cada um dos domínios apresenta um eixo de simetria C3 devido à presença de três subdomínios. A ocorrência de uma conformação fechada para a enzima EPSPS (SKARZYNSKI *et al.*, 1996) foi observada com a cristalização da enzima EPSPS em complexo binário com o substrato S3P (SCHÖNBRUNN *et al.*, 2001) e da enzima EPSPS em complexo ternário com o substrato S3P e o inibidor GLP (SCHÖNBRUNN *et al.*, 2001). Como a estrutura da EPSPS complexada somente com o S3P (EPSPS:S3P) é virtualmente idêntica à estrutura da enzima complexada com o S3P e o glifosato



(EPSPS:S3P:GLP), foi proposto que o S3P é que promove a transição do estado aberto da enzima para o estado fechado e não os íons negativamente carregados do glifosato (ESCHENBURG *et al.*, 2002). O fechamento dos dois domínios levaria ao acúmulo de resíduos com cargas positivas na fenda, atraindo moléculas carregadas negativamente para a cavidade do sítio.

As distâncias entre enzima-ligante dentro do complexo ternário contendo glifosato foram medidas por RMN, e agora os dados de diversos laboratórios indicam que o complexo não pode ser um análogo do estado da transição, porque a ligação do PEP e do glifosato não é aparentemente idêntica. Na sustentação destas descobertas, foi observada por algum tempo que as enzimas de diferentes organismos mostram grandes variação em suas sensibilidades ao herbicida. As constantes de inibição podem variar perto de três ordens de magnitude quando enzimas de algumas plantas e bactérias estão sendo comparadas, enquanto as variações de K_m para PEP são bem menores (HERRMANN & WEAVER, 1999).

Entretanto, análogos recentemente descobertos do estado da transição em que o anel do chiquimato é substituído por um anel do benzeno são inibidores muito eficazes da EPSPS (MILLER *et al.*, 1994; MILLER *et al.*, 1995; SHAH *et al.*, 1997). Desde de concentrações em nM de alguns destes compostos inibem as enzimas eficazmente, e desde que estes compostos são síntese orgânica direta prontamente acessível, podem representar a próxima geração de herbicidas comerciais que funcionam como inibidores para a via. O grau de tolerância do herbicida é correlacionado diretamente aos níveis da enzima (HERRMANN & WEAVER, 1999). A EPSPS de *Xylella fastidiosa* (XfEPSPS) foi modelada na forma de complexo ternário EPSPS:S3P:GLP.



1.1.7. CS

A CS (Figura 16), a sétima e última enzima na via metabólica do ácido chiquímico, catalisa a transformação da EPSP para corismato que é o último precursor comum na biossíntese de um grande número de compostos aromáticos, incluindo os aminoácidos aromáticos e metabólitos secundários (Figuras 9-11) em bactérias, fungos, plantas e apicomplexos de parasitas. A reação da CS requer um nucleotídeo de flavina reduzido (FMNH₂) como um co-fator e o substrato EPSP para produzir corismato e libera uma molécula de P_i e o FMN reduzido, embora a eliminação do 1,4-*anti* do fosfato e o C(6-*proR*)-hidrogênio não envolva uma mudança no redox (BALASUBRAMANIAN *et al.*, 1990; HAWKES *et al.*, 1990; MACHEROUX *et al.*, 1999). É uma reação incomum em que o estado redox do co-fator funcional (FMN) permanece sem mudança. Uma outra característica incomum da reação catalisada pela CS é uma eliminação anti-1,4 do 3-fosfato e C-(6-*pro-R*)-hidrogênio não combinado (VIOLA *et al.*, 2004). O mesmo acontece com a segunda enzima da via (DQHS) que requer NAD⁺ para ativar sua atividade catalítica. Análises espectroscópicas da função do co-fator sugerem a formação de um radical como um intermediário na reação catalisada pela CS (RAMJEE *et al.*, 1992).

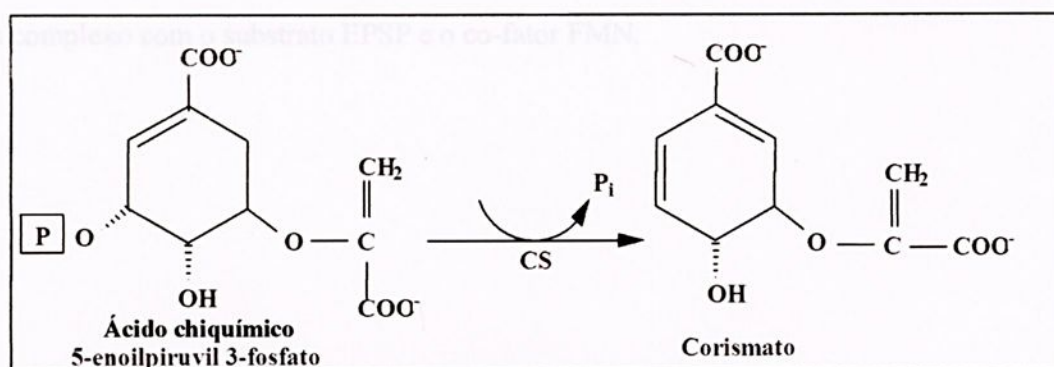


Figura 16 – Reação catalisada pela enzima CS.

O complexo binário enzima-flavina e os complexos ternários da enzima-flavina-substrato foram caracterizados em detalhe. Uma análise cinética mostrou que o intermediário da reação do FMN está dando forma, depois que EPSP é ligado, mas antes que seja consumido. O FMN pode servir como um doador de elétron para EPSP, iniciando desse modo a clivagem da ligação C-O (HERRMANN & WEAVER, 1999). Existem duas explicações para a função do co-fator reduzido de flavina: (i) a flavina têm meramente uma função estrutural e não um papel funcional, (ii) a flavina restaura a forma reduzida de uma sulfidril sensível à oxidação na enzima que é crucial para o ligamento do substrato ou possivelmente importante cataliticamente (HASAN E NESTER 1978; MACHEROUX *et al.*, 1999).

As enzimas de fungos como *N. crassa* (WELCH *et al.*, 1974) e *Euglena gracillis* (Schaller *et al.*, 1991a) são bifuncionais, em contraste as de *E. coli* (WHITE *et al.* 1988) e plantas (SCHALLER *et al.*, 1990) são monofuncionais e requerem FMNH₂ e estritamente ensaios de condições anaeróbicos. Schaller *et al.* (1991b) notou uma diferença entre as enzimas monofuncionais e bifuncionais, que é, a falta do domínio de redutase da flavina na enzima monofuncional, isto é refletido também nos tamanhos da subunidade nas enzimas oligoméricas. A enzima corismato sintase de *X. fastidiosa* (XfCS) foi modelada em complexo com o substrato EPSP e o co-fator FMN.



2. OBJETIVOS

A presente dissertação aplicou métodos de modelagem molecular comparativa para o estudo estrutural das sete enzimas da via metabólica do ácido chiquímico: 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato Sintase (DAHPS), 3-Desidroquinato Sintase (DHQS), 3-Desidroquinato Desidratase (DHQD), Chiquimato-5-Desidrogenase (SD), Chiquimato Quinase (SK), 5-Enoilpiruvilchiquimato 3-fosfato Sintase (EPSPS) e Corismato Sintase (CS), identificadas no genoma de *Xylella fastidiosa*. A partir das informações estruturais, obtidas da modelagem molecular, é esperado que seja possível efetuar o desenho de inibidores baseado nos modelos estruturais. Qualquer microrganismo poderá, em princípio, ser combatido por inibidores das enzimas da via metabólica do ácido chiquímico que impossibilitarão a produção do ácido corísmico. A utilização das enzimas da via metabólica do ácido chiquímico como alvos para o desenvolvimento de inibidores pode ser plenamente justificada e validada pela inibição específica da 5-Enoilpiruvilchiquimato 3-Fosfato Sintase (EPSPS) pelo glifosato (N-(fosfonometil) glicina), o qual é um herbicida efetivo e específico.



3. METODOLOGIA E FERRAMENTAS

3.1. A bioinformática

A bioinformática é a aplicação da tecnologia de informação ao gerenciamento de dados biológicos, com a finalidade de responder perguntas mais complexas neste campo da ciência. A bioinformática é uma área de investigação multidisciplinar que pode ser definida como a interface de duas ciências, a biologia e a computação, ou seja, um subconjunto de um campo maior da biologia computacional, a aplicação de técnicas analíticas e quantitativas a modelagem de sistemas biológicos. A bioinformática também está envolvida no agrupamento, organização, armazenamento e recuperação das informações biológicas que se encontram em bancos de dados públicos disponíveis na *web*, tais como seqüências de nucleotídeos, de aminoácidos, estruturas e domínios específicos das mais variadas proteínas e famílias.

Dentre os diversos objetivos da bioinformática está o desenvolvimento e implementação de ferramentas que permitem o acesso, uso e manejo de vários tipos de informação, alinhamentos seqüenciais, estruturas, visualização de enovelamentos protéicos, etc. Essas novas ferramentas desenvolvidas através da bioinformática oferecem a oportunidade de interpretar maior quantidade de dados. A comparação seqüencial é a ferramenta computacional mais útil para os biólogos moleculares. A crescente automação da biologia molecular e experimental e a aplicação da tecnologia da informação nas ciências biológicas conduzem a uma mudança fundamental na maneira como a pesquisa biológica é realizada.

3.2. Métodos de Alinhamento de Seqüências

Dentre as ferramentas mais utilizadas na bioinformática encontram-se aquelas relacionadas à comparação de seqüência de DNA e proteínas. O alinhamento de seqüência



está distante de ser uma tarefa comum em bioinformática (NOTREDAME, 2002). O alinhamento é uma comparação aproximada entre duas seqüências, a qual tem como objetivo evidenciar padrões comuns (similaridade) que podem ocorrer globalmente ou localmente entre as seqüências comparadas, e que refletem a existência de estruturas espaciais e funções biológicas compartilhadas entre as moléculas que elas representam. A comparação de pares de seqüências é utilizada para pesquisa de bancos de dados baseada em seqüência, construção de árvores evolutivas e identificação de características próprias de famílias de proteínas, criação de modelos por homologia e transmitir informações conhecidas de um genoma para outro.

A identidade de uma seqüência refere-se à presença do mesmo ácido nucléico ou aminoácido na mesma posição em duas seqüências alinhadas. A homologia de seqüência aplica-se à relação evolutiva entre as seqüências. Duas seqüências são consideradas homólogas se derivarem da mesma seqüência ancestral. Similaridade refere-se à presença de locais similares e idênticos em duas seqüências, enquanto homologia indica que duas seqüências possuem uma probabilidade alta de compartilharem o mesmo ancestral (GIBAS & JAMBECK, 2001). Ao comparar duas ou mais seqüências, sejam de aminoácidos ou de nucleotídeos, o problema básico é encontrar segmentos similares entre cada uma das seqüências (ALTSCHUL *et al.*, 1994).

Para fazer um alinhamento entre pares de seqüências de aminoácidos, é necessário entender os critérios estatísticos para a execução dos alinhamentos (matrizes PAM e BLOSUM). As matrizes de substituição de aminoácidos são complicadas, pois refletem a natureza química e a frequência de ocorrência de aminoácidos. Entre as matrizes de substituição mais usadas estão a BLOSUM (for BLOcks SUBstitution Matrix) (HENIKOFF & HENIKOFF, 1992) e a PAM (for "Point Accepted Mutation") (DAYHOFF *et al.*, 1978). A matriz PAM é construída utilizando as cadeias de Markov e caminhos randômicos



simples. Uma vantagem da matriz BLOSUM em relação a matriz PAM é detectar os parentescos biológicos entre seqüências moderadamente divergentes. A BLOSUM62 é uma matriz padrão para alinhamentos contínuos, enquanto a matriz BLOSUM50 é mais usada para geração de alinhamentos com espaços (*gaps*). Um alinhamento de seqüências pode permitir introdução de espaços para compensar inserções ou deleções de uma seqüência em relação à outra. A inclusão de espaços mantém, em registro, seqüências obviamente similares, o que não seria possível com alinhamento direto, sem espaços (GIBAS & JAMBECK, 2001).

Os programas mais utilizados nos processos de alinhamentos são o BLAST (ALTSCHUL *et al.*, 1990; ALTSCHUL *et al.*, 1997), o FastA (WILBUR & LIPMAN, 1983; LIPMAN & PEARSON, 1985), o CLUSTAL (THOMPSON *et al.*, 1994; HIGGINS *et al.*, 1992), e o MULTALIN (CORPET, 1988), todos disponíveis como servidores na *Internet*.

3.2.1. Programação Dinâmica e os Tipos de Alinhamentos

Needleman e Wunsch conceituaram o problema de alinhamento como sendo um problema de Programação Dinâmica (NEEDLEMAN & WUNSCH, 1970). Conhecido por algum tempo como algoritmo heurístico de homologias, a importância do método proposto por Needleman e Wunsch é que este método foi o primeiro a introduzir a noção de uma matriz de percurso, idéia central na versão moderna do algoritmo de alinhamento global de seqüências por Programação Dinâmica. Existem alguns tipos possíveis de alinhamentos de seqüências:

- Alinhamento Global: incluem todos os caracteres de cada uma das duas seqüências que estão sendo comparadas; o alinhamento ótimo é aquele com o mais alto escore possível;



- Alinhamento Local: incluem apenas regiões de maior similaridade, com ótimo alinhamento;
- Alinhamento Múltiplo, de membros de famílias protéicas;

Em cada tipo, os alinhamentos serão validados de acordo com o esquema de valores determinados pela matriz de substituição utilizada a fim de se estimar o grau de similaridade e/ou identidade entre as seqüências. A maior parte dos métodos de alinhamento baseia-se em técnicas de programação dinâmica (FILHO & ALENCASTRO, 2003).

Um algoritmo de programação dinâmica encontra uma boa solução dividindo o problema original em menores problemas e solucionando-os depois, por isso é um algoritmo ideal para se usar em paralelização computacional, pois não existe a dependência seqüencial. O objetivo do algoritmo de programação dinâmica é maximizar a pontuação geral para o alinhamento. Para isso, o número de pares de resíduos de alta pontuação deve ser maximizado e o número de espaços e pares de baixa pontuação deve ser minimizado.

A programação dinâmica apresenta como problema importante o fato de o número de operações a realizar crescer com o produto do comprimento das duas seqüências a comparar. No caso da pesquisa em bases de dados, salvo utilização de material específico, conduz a tempos de cálculo importantes. Por esta razão foram desenvolvidas várias heurísticas. Estes programas são muito mais rápidos, mas acarretam a perda da garantia de optimalidade do alinhamento. O objetivo das heurísticas é a pesquisa da fração menor possível das células da matriz, tentando evitar perder os melhores alinhamentos.

O alinhamento local de seqüências foi tratado inicialmente por Smith e Waterman (1981), que é uma modificação do algoritmo de Needleman e Wunsch, exceto pela possibilidade de haver uma opção adicional ao percorrer a matriz e através de duas



modificações principais. Um alinhamento local não precisa se estender do começo ao fim de duas seqüências a serem alinhadas. Os alinhamentos locais podem ser representados como blocos desprovidos de espaços. A formação de blocos é facilitada pela baixa penalidade imposta aos espaços de abertura e para os espaços de extensão. Logo, uma seqüência de resíduos poderá ter uma maior “mobilidade” e deslocar um grande número de resíduos através da inserção ou deleção de espaços (MUNIZ, 2003).

O alinhamento global é útil para comparar duas seqüências homólogas. Mas quando as duas seqüências apenas possuem certos domínios em comum, ou quando é necessário comparar uma seqüência com todas as entradas de uma base de dados, o mais interessante passa a ser os alinhamentos locais entre duas subseqüências. Tanto o alinhamento local quanto o global têm dependência quadrática, ou seja, para duas seqüências de comprimentos n e m , o tempo gasto na execução do alinhamento é proporcional ao produto $n*m$.

O alinhamento global considera a seqüência completa de bases ou aminoácidos. Nesse tipo de alinhamento, as penalidades tanto para os espaços de abertura como para os espaços na extensão são bastante altos. Portanto, formações de blocos durante os alinhamentos são inexistentes e o que se observa são pequenas regiões ou alguns poucos espaços espalhados ao longo da seqüência, preservando assim, o maior número possível de resíduos alinhados. Esse tipo de alinhamento é apropriado para seqüências que possuem grande similaridade em todo seu comprimento, já que o alinhamento é otimizado em toda a sua extensão (MUNIZ, 2003). Quando se pretende caracterizar uma família de seqüências que partilham uma mesma atividade biológica, a utilização de alinhamentos de seqüências duas a duas não é satisfatória, pois não fornece uma comparação do conjunto das seqüências.



O alinhamento múltiplo foi uma das primeiras respostas a estes tipos de problemas e ainda continua a ser a abordagem dominante, pelo menos no que se refere ao estudo da filogenia, identificação de motivos e domínios conservados e predição de estruturas (NOTREDAME, 2002). O algoritmo de programação dinâmica descrito anteriormente para alinhamento global é facilmente generalizável ao alinhamento de N seqüências. No entanto, o tempo e a memória necessários para a sua execução cresce em L^N (L sendo o comprimento característico das seqüências). Para valores típicos de L de 1000, este tipo de abordagem toma-se assim muito rapidamente impraticável.

Algoritmos alternativos têm sido utilizados para contornar este problema como: algoritmo progressivo, algoritmo exato, algoritmo iterativo, algoritmo iterativo estocástico, algoritmo iterativo não estocástico e algoritmo baseado na consistência (NOTREDAME, 2002). O alinhamento múltiplo tem sido utilizado para: validação de ORF, padrões de conservação diferencial, estudos evolucionários, organização de domínios, modelagem de estruturas por homologia e propagação de informações (LECOMPTE *et al.*, 2001).

3.3. Método de Modelagem Molecular por Homologia ou Comparativa

A caracterização funcional de uma proteína é um dos problemas mais freqüentes na biologia (UCHÔA *et al.* 2004). A seqüência de uma proteína determina sua estrutura 3D, contudo o algoritmo que permita, com precisão, determinar a estrutura 3D de uma proteína, partindo de sua seqüência ainda está por ser determinado. As estruturas 3D de proteínas naturais são guiadas por dois distintos princípios: as leis da física e da teoria da evolução. Estes princípios deram ascensão a duas classes de métodos de predição de estrutura de proteína (Figura 17) (FISER *et al.*, 2002). A primeira classe de método de aproximação (*de*



novo ou *ab initio*), prediz a estrutura da sequência sozinha, sem confiar na similaridade do nível de enovelamento entre a sequência modelada e estrutura conhecida (BONNEAU & BAKER, 2001). Os métodos *de novo* supõem que a estrutura nativa corresponde ao mínimo de energia livre global acessível a proteína e tenta encontrar este mínimo por uma exploração de muitas conformações concebíveis de proteínas. Os dois componentes chaves do método *de novo* que são o procedimento para eficientemente realizar a busca conformacional, e a função de energia livre usada para conformações possíveis de avaliação (BONNEAU & BAKER, 2001). A segunda classe de métodos inclui *threading* (DOMINGUES *et al.*, 2000) e modelagem comparativa (BLUNDELL *et al.*, 1987; MARTIRENOM *et al.*, 2000), confia na similaridade detectável que mede a maioria da sequência modelada ao menos de uma estrutura conhecida. A modelagem de uma proteína pelo método da homologia é baseada no conceito de evolução molecular.

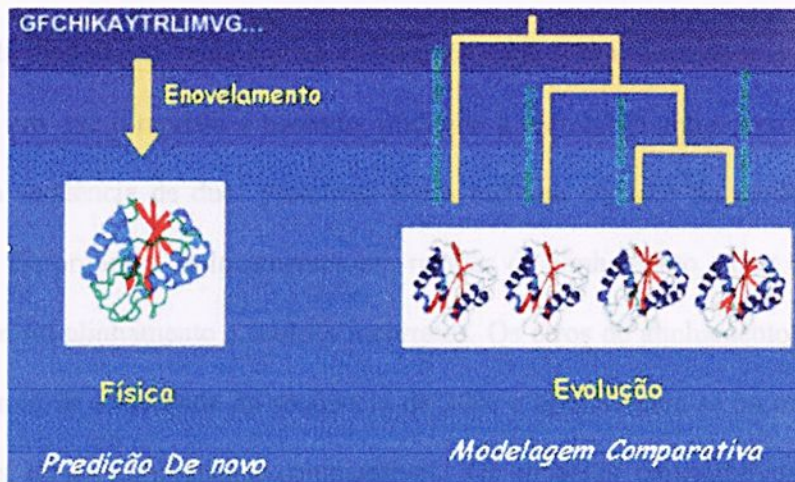


Figura 17 – Predição de estrutura *de novo* e modelagem comparativa de estrutura de proteína (FISER *et al.*, 2002).

Relacionamentos evolucionários permitem o uso de métodos computacionais para prever a estrutura de uma proteína baseada na sua similaridade com estruturas de proteínas conhecidas (ARCURI *et al.*, 2004; PEREIRA *et al.*, 2003; DA SILVEIRA *et al.*, 1998; ESWAR *et al.*, 2000).

2004). Entre todas as aproximações teóricas atuais, a modelagem molecular comparativa ou por homologia de estruturas de proteína é o único método computacional que tem o potencial de construir um modelo 3D para uma proteína de estrutura desconhecida (o alvo) baseada em uma ou mais proteínas relacionadas de estruturas conhecidas (os moldes) com precisão e confiança comparáveis a estruturas experimentais de baixa resolução a partir da sequência de aminoácidos (TRAMONTANO *et al.*, 2001; BLUNDELL *et al.*, 1987; GREER, 1981; JOHNSON *et al.*, 1994; SALI & BLUNDELL, 1993; SALI, 1995; SANCHEZ & SALI, 1997; MARTI-RENOM *et al.*, 2000; FISER *et al.*, 2002; SANCHEZ & SALI, 2000).

A técnica de modelagem, em geral, é cercada por diversas dificuldades, tais como alinhamentos problemáticos que geram os erros na modelagem comparativa, modelagem de Alças (*loops*) existência de múltiplos conformacionais, a modelagem de sítios de ligação e modelagem das cadeias laterais (MUNIZ, 2003).

A precisão da modelagem comparativa é correlacionada com a identidade da sequência da porcentagem em que o modelo é baseado, imitando a correlação entre o estrutural e as similaridades da sequência de duas proteínas. Erros incluem enganos na modelagem das cadeias laterais, distorções e deslocamentos em regiões de alinhamento, erros nas regiões sem molde, erros no alinhamento e moldes incorretos. Os erros de alinhamento aumentam rapidamente abaixo da identidade da sequência de 30% e transformam-se na origem mais significativa dos erros em modelos comparativos. Os modelos comparativos têm sido usados em estudos de mecanismos catalíticos de enzimas, desenho e melhoramento de ligantes, *docking* de macromoléculas, *screening* virtuais e *docking* de pequenos ligantes, etc (FISER *et al.*, 2002).

A modelagem comparativa consiste em quatro etapas principais (SANCHEZ & SALI, 1998; ESWAR *et al.*, 2003): (i) identificação da similaridade entre a sequência alvo de



interesse e ao menos uma estrutura de proteína conhecida (o molde); (ii) alinhamento da sequência alvo e a sequência do molde(s); (iii) construção de um modelo baseado no molde(s) escolhido(s); e (iv) avaliação do modelo para verificar sua precisão (UCHÔA *et al.*, 2004), mostradas na figura 18.

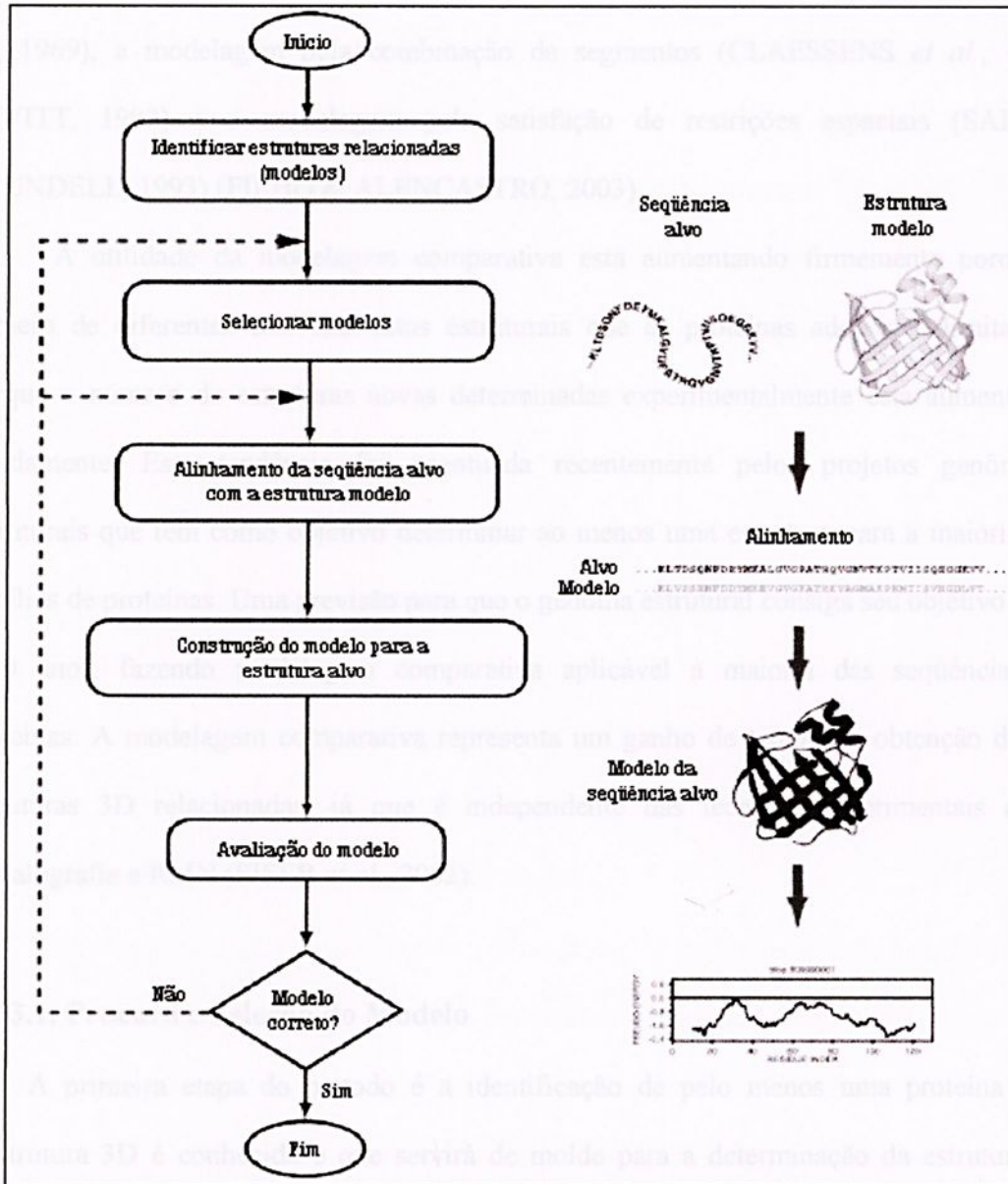


Figura 18 - Passos da modelagem molecular comparativa (Sanchez & Šali, 1997).

A localização no espaço 3D do maior número possível de átomos da proteína (alvo) depende do alinhamento. A construção da parte interna da proteína (alvo) baseia-se na idéia de que a conformação da cadeia principal da estrutura (molde) pode ser transferida para ela. Existem muitos métodos de modelagem dessas regiões conservadas. Os mais importantes são: a modelagem por corpo rígido (GREER, 1990; BLUNDELL *et al.*, 1987; BROWNE *et al.*, 1969), a modelagem pela combinação de segmentos (CLAESSENS *et al.*, 1989; LEVITT, 1992) e a modelagem pela satisfação de restrições espaciais (SALI & BLUNDELL, 1993) (FILHO & ALENCASTRO, 2003).

A utilidade da modelagem comparativa está aumentando firmemente porque o número de diferentes enovelamentos estruturais que as proteínas adotam é limitado e porque o número de estruturas novas determinadas experimentalmente está aumentando rapidamente. Esta tendência foi acentuada recentemente pelos projetos genômicos estruturais que tem como objetivo determinar ao menos uma estrutura para a maioria das famílias de proteínas. Uma previsão para que o genoma estrutural consiga seu objetivo é em 5-10 anos, fazendo modelagem comparativa aplicável à maioria das seqüências de proteínas. A modelagem comparativa representa um ganho de tempo na obtenção dessas estruturas 3D relacionadas, já que é independente das técnicas experimentais como cristalografia e RMN (FISER *et al.*, 2002).

3.3.1. Procura e Seleção do Modelo

A primeira etapa do método é a identificação de pelo menos uma proteína cuja estrutura 3D é conhecida e que servirá de molde para a determinação da estrutura da proteína (alvo). Duas situações são, em geral, possíveis: (i) a família protéica a que pertence à proteína (alvo) é conhecida e (ii) não se sabe a que família a proteína (alvo) pertence (FILHO & ALENCASTRO, 2003).



No primeiro caso, é suficiente selecionar a proteína (molde) diretamente do PDB (BERMAN, *et al.* 2000, WESTBROOK *et al.*, 2002). No segundo caso, o método mais simples é procurar de forma sistemática um ou mais moldes adequados em um banco de dados de estruturas (FILHO & ALENCASTRO, 2003) de proteínas conhecidas (PDB) (<http://www.rcsb.org/pdb>), usando como parâmetro de entrada uma sequência de estrutura não determinada (alvo) para que esta seja comparada com as sequências mais homólogas de estruturas conhecidas (moldes).

Neste trabalho foi utilizado o segundo caso através do programa FastA (PEARSON, 2000; PEARSON, 1995) para estar realizando a identificação dos moldes para as sete enzimas da via metabólica do ácido chiquímico. Os moldes escolhidos foram aqueles que apresentavam identidade elevada e que não fosse inferior a 30% e que tivessem um menor número de espaços no alinhamento.

Depois de escolhido os moldes para cada enzima da via metabólica do ácido chiquímico, as coordenadas atômicas e informações (resolução da estrutura, organismo, se tem ligante ou não,...) sobre o molde foram retiradas do banco de estruturas de proteínas PDB (BERMAN, *et al.* 2000, WESTBROOK *et al.*, 2002).

3.3.2. Alinhamento e Construção do Modelo

Uma vez feita a seleção dos moldes, um método de alinhamento deve ser usado para realizar um segundo alinhamento entre o molde com a sequência designada. O objetivo do alinhamento é alinhar resíduos estruturalmente equivalentes (FILHO & ALENCASTRO, 2003). Após o alinhamento, é possível reconhecer regiões estruturalmente conservadas e regiões variáveis (FILHO & ALENCASTRO, 2003). Para este segundo de alinhamento foi utilizado o alinhamento gerado pelo programa MODELLER (SALI & BLUNDELL,



1993). Este alinhamento será utilizado como entrada para gerar as modelagens no programa MODELLER (SALI & BLUNDELL, 1993).

Neste trabalho de modelagem das enzimas da via metabólica do ácido chiquímico de *X. fastidiosa* foi utilizado uma aproximação automatizada para modelagem comparativa de estrutura de proteína baseada na satisfação da restrição espacial (Figura 19) (SALI & BLUNDELL, 1993; FISER, 2000) que é executada pelo programa MODELLER (<http://guitar.rockefeller.edu>) (SALI & BLUNDELL, 1993; FISER *et al.*, 2002), através do servidor de *web* PARMODEL (UCHÔA *et al.*, 2004).

A modelagem comparativa por satisfação de restrição espacial consiste de três etapas basicamente para construção do modelo (alvo). Na primeira etapa da construção do modelo, as restrições espaciais da sequência alvo são calculadas, a partir do segundo alinhamento gerado. No geral, as restrições são expressas como funções de densidade de probabilidade condicional (pdf) $P(f/I)$ para a característica espacial restrita f , dado diversas variáveis I que foram encontradas para as características a serem preditas. Há três tipos de restrição, depende da origem e da natureza da informação I (SALI & BLUNDELL, 1993; FISER *et al.*, 2002).

Primeiramente, as restrições são obtidas para aqueles resíduos do alvo que são alinhados com os resíduos do molde. Esta homologia derivada das restrições limitada as distâncias entre os átomos da cadeia principal e cadeia lateral, bem como os ângulos diedros da cadeia principal (Φ , Ψ e Ω) e os ângulos diedros da cadeia lateral (χ_i). Estas restrições foram obtidas de uma análise estatística de 105 famílias alinhadas que incluem 416 proteínas definidas estruturalmente (SALI & OVERINGTON, 1994). Por exemplo, uma restrição numa certa distância entre dois $C\alpha$ - $C\alpha$ equivalentes em duas estruturas de proteína relacionadas são bem descritas por uma soma pesada de duas funções Gaussianas



que correspondem às duas distâncias do molde, respectivamente (SALI & BLUNDELL, 1993; FISER *et al.*, 2002).

A segunda classe de restrições reflete as preferências estatísticas extraídas das estruturas de proteínas conhecidas no geral e são relacionadas aos potenciais estatísticos médios de força (FISER *et al.*, 2002). As restrições dependem somente dos tipos restritos de átomos ou resíduos e não da estrutura do molde. Estas restrições são aplicadas à cadeia principal e aos ângulos diedros da cadeia lateral dos resíduos alvos que não são alinhados com os resíduos do molde e para as distâncias entre todos os pares de átomos não ligados. São usados porque foram encontrados para resultar em modelos mais exatos do que os termos correspondentes do campo de força mecânica molecular (FISER *et al.*, 2000; SALI & BLUNDELL, 1993; FISER *et al.*, 2002).

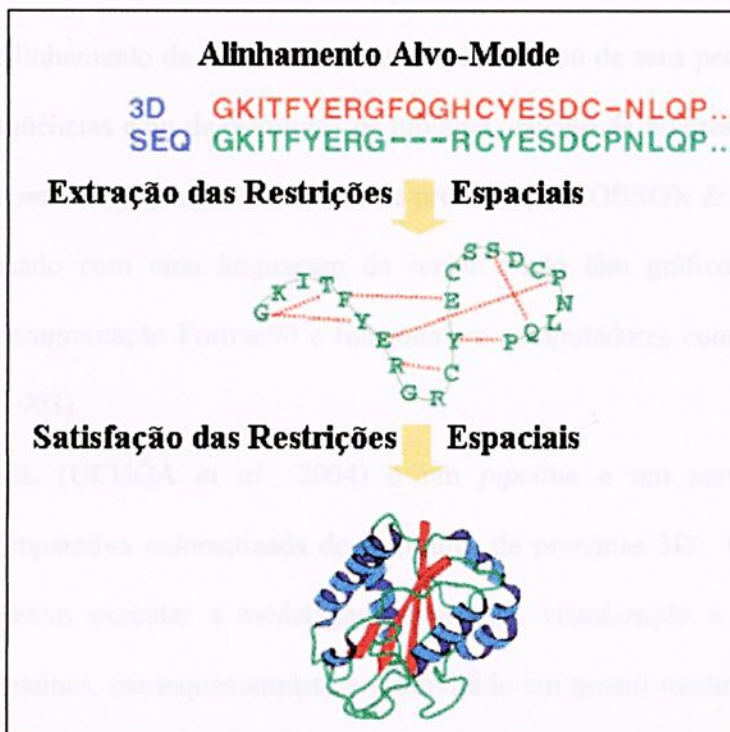


Figura 19 – Modelagem comparativa de estrutura de proteína por satisfação de restrição espacial (FISER *et al.*, 2002). Primeiro, a estrutura 3D conhecida usada como modelo é alinhada com a sequência (alvo) a ser modelada. Segundo características espaciais são transferidas do modelo para o alvo. Terceiro e ultimo passo, o modelo 3D é obtido por satisfação de todas as restrições que são possíveis.

O terceiro tipo de restrição é obtido a partir do campo de força mecânica molecular do CHARMM-22 (MACKERELL *et al.*, 1998) e incluem restrições de ligações químicas. Estas restrições mecânicas moleculares reforçam estereoquimicamente o modelo correto. Depois que todas as pdfs são calculadas, seus logaritmos são somados para obter uma função objetiva que dependa do modelo e dê sua probabilidade. Finalmente, o modelo que contem todos os átomos não hidrogenados é calculado otimizando a função objetiva no espaço cartesiano. A otimização é realizada pelo método da função do alvo variável que emprega gradientes conjugados e a dinâmica molecular com *simulated annealing* (SALI & BLUNDELL, 1993; FISER *et al.*, 2002).

A saída é um modelo 3D para a sequência alvo contendo todas as cadeias principais e cadeias laterais dos átomos sem o hidrogênio. Além da construção do modelo, o MODELLER (SALI & BLUNDELL, 1993) pode executar tarefas auxiliares adicionais, incluindo um alinhamento de duas sequências de proteína ou de seus perfis, alinhamento múltiplo de sequências e/ou de estruturas de proteína, cálculo de árvores filogenéticas, e modelagem *de novo* de alças nas estruturas de proteína (JACOBSON & SALI, 2004). O programa é usado com uma linguagem de *script* e não têm gráficos. É escrito na linguagem de programação Fortran90 e funciona em computadores com Unix (SALI & BLUNDELL, 1993).

PARMODEL (UCHÔA *et al.*, 2004) é um *pipeline* e um servidor *web* para modelagem comparativa automatizada de estruturas de proteínas 3D. O PARMODEL tem como objetivo executar a modelagem, avaliação, visualização e otimização dos modelos de proteínas, conseqüentemente é subdividido em quatro módulos: Modelagem Parmodel, Avaliação Parmodel, Visualização Parmodel, e Otimização Parmodel. O PARMODEL automatiza e integra os principais programas usados na modelagem comparativa como: MODELLER (SALI & BLUNDELL, 1993), PROCHECK



(LASKOWSKI *et al.*, 1993), Raster3D (BACON & ANDERSON, 1988; MERRITT, 1994), MOLSCRIPT (KRAULIS, 1991) e GROMACS (VAN GUNSTEREN *et al.*, 1996; BERENDSEN *et al.*, 1995; LINDAHL *et al.*, 2001).

PARMODEL executa a paralelização do MODELLER em um *cluster Beowulf* de 16 PCs rodando em Linux. O usuário pode construir vários modelos para uma mesma proteína (até 1000) e pode aumentar a possibilidade de obter um bom modelo, porque as mudanças nas restrições espaciais iniciais gerarão diferentes conformações na estrutura final.

O módulo de modelagem do PARMODEL requer algumas informações e arquivos na interface de entrada de dados para começar a modelagem dos modelos, como: (i) a sequência de aminoácido da proteína (alvo), (ii) até oito estruturas de proteínas (moldes), (iii) o número de modelos 3D (alvo) a ser modelado, (iv) escolher o método experimental dos modelos, (v) escolher o número de ligantes que o modelo (alvo) ou se não vai ser modelado com ligantes, (vi) escolher que programas serão usados para avaliar os modelos: PROCHECK (LASKOWSKI *et al.*, 1993) e 3DAnalysis (LUTHY *et al.*, 1992; BOWIE *et al.*, 1991) e (vii) escolher como os modelos finais serão classificados, pela função objetiva do MODELLER ou pela qualidade estereoquímica. Depois de entrar com as informações é só submeter os dados de entrada e aguardar o retorno do melhor modelo gerado. O módulo de modelagem do PARMODEL retoma os resultados da modelagem por e-mail, que contém um arquivo comprimido com o melhor modelo da estrutura de proteína (alvo), o resultado do alinhamento, uma lista com a função objetiva do MODELLER ou da qualidade estereoquímica do modelo para todos os modelos construídos e os resultados da análise para o melhor modelo.



3.3.3. Avaliação do Modelo

Diversas metodologias existem para a checagem e avaliação da qualidade de uma estrutura, seja ela determinada experimentalmente ou por modelagem. O modelo pode ser avaliado como um todo ou por regiões. Como a qualidade do modelo depende muito da estrutura da proteína (molde) utilizada, é preciso que esta última tenha sido obtida em alta resolução ($\leq 2\text{\AA}$) e com um fator-R satisfatório ($\leq 20\%$). A qualidade do modelo construído será proporcional à qualidade da estrutura 3D utilizada como molde (SANCHEZ AND ŠALI, 1998).

A análise geométrica das estruturas de proteínas possui dois objetivos principais: (i) avaliação interna de sua consistência intrínseca que verifica se o modelo satisfaz ou não às restrições calculadas e (ii) avaliação externa que confia em informações que não foram usadas no cálculo do modelo. A análise geométrica pode mostrar onde um modelo determinado experimentalmente ou por modelagem viola as leis da química. Existem leis da física que governam interações intermoleculares. Essas interações limitam não somente os contatos entre pares de átomos em diferentes partes de uma cadeia protéica, como também o tanto que os átomos possam girar livremente ao redor das ligações que os conectam. A estrutura das órbitas atômicas e a natureza das ligações entre átomos impõem limites naturais na posição de átomos ligados entre si, de modo que os ângulos de ligação e os ângulos diedros fiquem, restritos a um grupo limitado de valores. Algumas são executadas de forma independente; outras são incorporadas em pacotes de simulação e de análise de estrutura mais abrangentes (GIBAS & JAMBECK, 2001).

A qualidade das enzimas da via metabólica do ácido chiquímico foram avaliadas através dos programas: WHATCHECK/WHATIF (HOOFT *et al.*, 1996; VRIEND, 1990), VERIFY-3D (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992; KABSCH & SANDER, 1983), RMSD da geometria ideal (ENGH & HUBER, 1991; KEARSLEY, 1989) usando o



programa X-PLOR (SCHWIETERS *et al.*, 2003; BRÜNGER, 1992), o RMSD entre a distância dos átomos de Ca - Ca sobrepostos (KABASH & SANDER, 1983) usando o programa LSQKAB do pacote CCP4 (Collaborative Computational Project N° 4., 1994), o G-factor usando o programa PROCHECK (LASKOWSKI *et al.*, 1994) e o diagrama de Ramachandran (RAMACHANDRAN *et al.*, 1963; RAMAKRISHNAN & RAMACHANDRAN, 1965) usando o programa PROCHECK (LASKOWSKI *et al.*, 1994).

Na avaliação da interação entre enzima-ligante foram verificadas as ligações intermoleculares de hidrogênio, a área de contato do ligante com a proteína através da rotina AEROMOL do pacote CCP4 (Collaborative Computational Project N° 4., 1994), ligações de van der Waals e o potencial eletrostático de superfície gerado pelo GRASP (NICHOLLS *et al.*, 1991).

Para obter a análise do programa VERIFY-3D (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992; KABSCH & SANDER, 1983), e o diagrama de Ramachandran foi utilizado o módulo de avaliação do programa PARMODEL (UCHÔA *et al.*, 2004). O módulo de avaliação PARMODEL permite que o usuário avalie os modelos gerados por modelagem, difração de raios X e RMN. O módulo requer como entrada apenas um arquivo das coordenadas atômicas no formato “pdb” da proteína que será avaliada, os resultados são mostrados na mesma página da web. As saídas geradas podem ser salvas em arquivos com extensão “pdf” para o PROCHECK e para o VERIFY-3D uma página de *web* com os valores e um gráfico.



3.4. Métodos de Avaliação

3.4.1. PROCHECK

O PROCHECK (LASKOWSKI *et al.*, 1994) analisa a geometria global da estrutura ou cada resíduo individualmente, utilizando para isso parâmetros estereoquímicos derivados de estruturas de alta resolução ou bem refinadas (MORRIS *et al.*, 1992) que constituem sua base de dados. É um pacote popular de programas para verificação da qualidade da estrutura 3D da proteína, que inclui cinco programas, onde quatro são escritos na linguagem de programação Fortran77 e um escrito em C, pode ser executado em sistemas UNIX e Windows.

Os parâmetros estereoquímicos usados são aqueles descritos em detalhe em MORRIS *et al.* (1992). As verificações também fazem uso do comprimento e ângulo de ligação ideal, derivados de uma análise detalhada (ENGH & HUBER, 1991) de estruturas de pequenas moléculas em Cambridge Structural Database (CSD) (ALLEN *et al.*, 1979). Estes parâmetros utilizados como informações estereoquímicas checados pelo PROCHECK são: ligações covalentes, planaridade de grupos planares (aromáticos, ligações peptídicas, etc.) ângulos diédricos, quiralidade, interações não covalentes, ligações de hidrogênio da cadeia principal e pontes de dissulfeto.

O PROCHECK requer como entrada um arquivo das coordenadas atômicas da estrutura da proteína a ser avaliada no formato "pdb" e produz representações coloridas em PostScript (Adobe System Inc., 1985) facilmente interpretadas, descrevendo a estrutura de uma proteína, e também comparando duas estruturas de proteínas relacionadas, juntamente com uma lista detalhada de resíduo por resíduo. Dando uma avaliação da qualidade total da estrutura, em comparação às estruturas bem-refinadas da mesma definição, e destaca também as regiões que podem necessitar uma investigação



mais adicional. Dentre as varias análises fornecidas pelo PROCHECK foram utilizadas neste trabalho: o Diagrama de Ramachandran e o G-factor.

3.4.2. Diagrama de Ramachandran

G. N. Ramachandran descreveu as conformações disponíveis para os aminoácidos em uma cadeia polipeptídica. A conformação da cadeia peptídica é simplesmente descrita pelos valores dos ângulos diedros na estrutura principal da proteína (ângulo descrito pelo nitrogênio e carbono alfa ($N-C_\alpha$) e o ângulo descrito pelo carbono alfa e carbono $C_\alpha-C$) (Figura 20). Esses ângulos são denominados ângulos de torção ϕ e ψ , respectivamente (GIBAS & JAMBECK, 2001). Por convenção, ambos, ϕ e ψ , são definidos como 0° na conformação na qual as duas ligações peptídicas conectadas a um único carbono α estão em um mesmo plano. Em princípio, ϕ e ψ podem ter qualquer valor entre -180° e $+180^\circ$, porém muitos valores de ϕ e ψ são proibidos pelas interferências estéricas entre átomos pertencentes ao mesmo esqueleto polipeptídico e pelas cadeias laterais dos resíduos de aminoácidos. A conformação na qual ϕ e ψ são, ambos, iguais a 0° é proibida por esta razão; esta convenção é meramente usada como ponto de referência para descrever os ângulos de rotação.

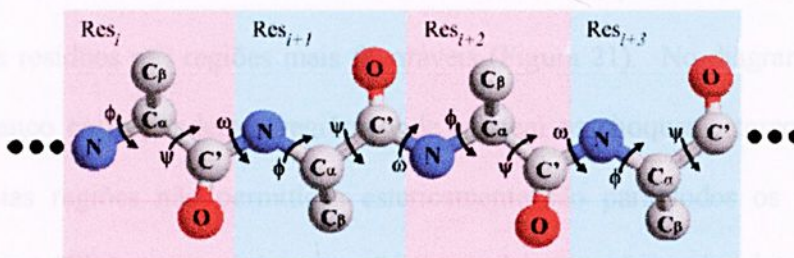


Figura 20 – Parte de uma cadeia polipeptídica indicando os graus de liberdade torcional de cada unidade peptídica e os ângulos de rotação ϕ e ψ .

A cadeia não é livre para girar ao redor do terceiro tipo de ligação na estrutura principal da proteína, a ligação peptídica, por ser uma ligação parcialmente dupla e, portanto quimicamente restrita a ser planar; sendo assim, os valores de ϕ e ψ para cada aminoácido fornecem uma descrição completa da estrutura da cadeia principal da proteína.

G. N. Ramachandran usou modelos de pequenos peptídeos para variar sistematicamente os ângulos de torção ϕ e ψ com o objetivo de encontrar conformações estáveis. Para cada conformação, a estrutura foi examinada para verificar a proximidade existente dos contatos entre os átomos. Os átomos foram tratados como esferas duras com as dimensões que correspondem a seus raios em van der Waals. Conseqüentemente, os ângulos de torção ϕ e ψ que fazem com que as esferas colidam correspondem a conformação estericamente não permitida (região não permitida) para a cadeia principal de um polipeptídeo (RAMAKRISHNAN, 2001).

O diagrama de Ramachandran (RAMACHANDRAN *et al.*, 1963; RAMAKRISHNAN & RAMACHANDRAN, 1965) indica a distribuição dos ângulos Φ e ψ dos resíduos pertencentes a uma determinada estrutura. Seus indicadores de qualidade baseiam na análise de 118 estruturas de resolução de pelo menos 2,0 Å de resolução e R_{factor} menor 20%. Nos seus critérios uma estrutura de boa qualidade deve ter 90% ou mais de seus resíduos nas regiões mais favoráveis (Figura 21). No diagrama abaixo as áreas em branco correspondem a regiões onde existem os choques estereoquímicos na proteína. Estas regiões não permitidas estericamente são para todos os aminoácidos exceto a glicina (Gly) que é a única que não tem cadeia lateral e pode adotar os ângulos de torção ϕ e ψ em todos os quadrantes do diagrama de Ramachandran.



As regiões vermelhas correspondem à conformação onde não há nenhum choque estereoquímico, ou seja, são as regiões permitidas e aonde se encontram as hélices- α e folhas- β . As áreas em amarelo mostram as regiões limites onde um raio ligeiramente mais curto de van der Waals é usado no cálculo. Isto faz com que tenha uma região adicional que corresponda a hélices esquerda e direita.

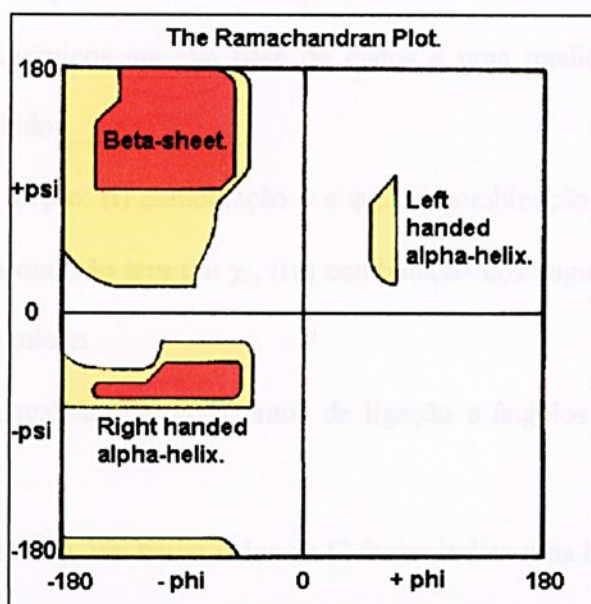


Figura 21 – Diagrama de Ramachandran. A região generosamente permitida em vermelho, a região adicionalmente permitida em amarelo e a não permitida em branco. E as regiões onde estão as hélices- α e folhas- β

Os aminoácidos L não podem formar extensas regiões de hélices esquerdas, mas ocasionalmente os resíduos individualmente podem adotar esta conformação. Estes resíduos são geralmente glicinas (Gly), mas também podem ser resíduos de asparaginas ou aspartatos onde a cadeia lateral forma uma ligação de hidrogênio com a cadeia principal e estabiliza conseqüentemente esta conformação inicialmente desfavorável. A hélice 3(10) ocorre perto da região hélice- α na parte superior a direita e está na borda da região adicionalmente permitida indicando uma estabilidade mais baixa. As regiões não

permitidas envolvem geralmente um obstáculo estérico entre a cadeia lateral do grupo metileno C_{β} e átomos da cadeia principal.

3.4.3. G-factor

O G-factor fornece uma medida do desvio de uma dada propriedade estereoquímica (ângulos de torção e geometria covalente), baseado nas distribuições observadas nos parâmetros estereoquímicos em sua base de dados e uma medida da normalidade da estrutura como um todo.

- Ângulos de torção: (i) combinação ϕ e φ , (ii) combinação $\chi_1 - \chi_2$, (iii) torção χ_1 para aqueles que não tem um χ_2 , (iv) combinação dos ângulos de torção χ_3 e χ_4 e torção do ângulo ω .
- Geometria covalente: comprimentos de ligação e ângulos de ligação da cadeia principal.

Para um dado resíduo, um baixo valor do G-factor indica uma baixa probabilidade de ocorrência do parâmetro que se observa. O cálculo do G-factor global envolve os diversos G-factors para toda a estrutura, num único número. O G-factor global é a melhor medida da qualidade total da estereoquímica de uma molécula. Por exemplo, os resíduos que caem nas regiões não permitidas do diagrama de Ramachandran terão um G-factor baixo (ou muito negativo). Se uma proteína tiver muitos resíduos com G-factor baixo sugere que algo pode estar errado com geometria total da proteína avaliada.

Para que os ângulos de torção do G-factor obtivessem um padrão foram derivados de uma análise de 163 não homólogos, cadeias de proteínas de alta resolução escolhidas das estruturas resolvidas por cristalografia por raios X com uma resolução $\leq 2 \text{ \AA}$ e um R-fator $\leq 20\%$. As análises forneceram as distribuições observadas de ϕ , φ , $\chi_1 - \chi_2$, χ_1 , χ_3 , χ_4 e ω .



para cada um dos 20 tipos de aminoácidos. Estas distribuições foram divididas então em celas. Por exemplo, cada tipo de resíduo do diagrama Ramachandran de valores ϕ e φ foram divididos em celas 45x45. Os números das observações em cada celas foram usados para calcular a probabilidade para cada tipo cela. Para calcular os comprimentos e ângulos de ligação da cadeia principal, o G-factor usa como parâmetro os valores medidos de pequenas moléculas e desvios padrões de Engh & Huber (1991).

3.4.4.VERIFY-3D

O VERIFY-3D (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992; KABSCH & SANDER, 1983) mede a compatibilidade entre a sequência de aminoácidos de uma proteína e o modelo da sua estrutura tridimensional, usando um perfil 3D. Para tanto, o programa utiliza a seguinte metodologia: 1) reduz a estrutura 3D a uma sequência de ambientes químicos (um para cada resíduo) categorizados entre 18 possibilidades de acordo com a área do resíduo exposta ao solvente, com a fração de contatos feitos com átomos polares e a estrutura secundária local (perfil 3D), 2) utilizando uma matriz (chamada matriz 3D-1D (BOWIE *et al.*, 1991)) que descreve a probabilidade de se encontrar cada um dos 20 aminoácidos em cada uma das 18 classes de ambientes químicos, o programa determina a probabilidade, por resíduo, de acordo com o tipo de aminoácidos e a natureza do ambiente químico calculado no passo anterior. A matriz 3D-1D é previamente conhecida e derivada de uma análise de estruturas conhecidas de boa qualidade (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992).

O resultado é uma medida da compatibilidade entre a sequência e sua estrutura 3D descrita pelo seu perfil tridimensional. Lüthy *et al.* (1992) determinaram empiricamente o índice global esperado de $S_{calc} = \exp(-0,83 + 1,008 \times \ln(L))$ de compatibilidade entre a



seqüência e a estrutura 3D onde L é o comprimento da seqüência e S_{calc} representando, então, a soma das probabilidades individuais dos resíduos. Lüthy *et al.* (1992) também sugere um limite de $0,45 \times S_{calc}$ para a confiabilidade da compatibilidade seqüência/estrutura. Valores menores que $0,45 \times S_{calc}$ indicam uma estrutura incorreta, valores em torno de $0,45 \times S_{calc}$ podem estar corretas, mas possuem qualidade questionável e valores acima de $0,45 \times S_{calc}$ indicam uma estrutura correta. O VERIFY-3D requer como entrada um arquivo de coordenadas atômicas no formato .pdb.

3.4.5. WHATIF/WHATCHECK

O WHATIF (VRIEND, 1990) pode calcular quase todas as propriedades de proteínas, desde a área de superfície acessível a solvente até valores de pKa para contato com a dinâmica molecular com GROMOS (VAN GUNSTEREN *et al.*, 1987).

O WHATECHECK (HOOFT *et al.*, 1996) fornece acesso a um subconjunto de verificações de qualidade estrutural WHATIF. WHATECHECK dá informações sobre a formação de regiões centrais hidrofóbicas, a acessibilidade de resíduos e átomos a moléculas de solvente (água), a distribuição espacial de grupos iônicos, a distribuição das distâncias atômicas e das ligações de hidrogênio da cadeia principal. Ou seja, retrata a estereoquímica, comprimentos de ligações, ângulos e diedros, entre outras quantidades na forma de um relatório gerado no formato pdf.

3.4.6. RMSD (Desvio Médio Quadrático)

O RMSD pode ser calculado de duas maneiras: (i) uma função de todos os átomos de uma proteína (ENGH & HUBER, 1991 – Fórmula 1) e (ii) uma função de algum subconjunto dos átomos, como a estrutura principal da proteína ou posições de C_{α} apenas



(KABASH & SANDER, 1983 – Fórmula 2). O parâmetro mais comum que expressa a diferença entre duas estruturas protéicas é o RMSD, ou desvio médio da raiz quadrada, em posições atômicas entre as duas estruturas.

Quando o RMSD é calculado como uma função de todos os átomos de uma proteína, podemos adotar a seguinte definição que o RMSD serve para analisar a medida de quanto à posição de cada átomo de uma estrutura varia com relação a um vetor aceito, e é denominado como RMSD da geometria ideal (KEARSLEY, 1989). O RMSD da geometria ideal utiliza como parâmetros o comprimento e ângulo de ligação ideal, derivados de uma análise detalhada (ENGH & HUBER, 1991) de estruturas de pequenas moléculas em Cambridge Structural Database (CSD) (ALLEN *et al.*, 1979). O programa utilizado para executar os cálculos de comprimento e ângulos de ligação foi o programa X-PLOR (SCHWIETERS *et al.*, 2003; BRÜNGER, 1992).

Quando o RMSD é calculado como uma função de algum subconjunto dos átomos é comum a utilização de um subconjunto de átomos da proteína, quando duas estruturas protéicas são comparadas, elas não serão idênticas entre si em sequência e, por isso, os únicos átomos entre comparação de posição (um-a-um) que podem ser efetuados são os átomos da estrutura principal (C_{α} - C_{α}). Neste contexto abordado a orientação de uma estrutura molecular se torna importante. Uma vez que as estruturas de proteínas são geralmente descritas em coordenadas atômicas cartesianas, e surgem como uma orientação incorporada em relação ao espaço. O RMSD é uma função da distância entre os átomos em uma estrutura e os mesmos átomos em outra estrutura. Portanto, se uma molécula começar em uma posição diferente do sistema de coordenadas de referência, a outra molécula, o RMSD entre as duas proteínas será calculado independentemente.



$$rmsd_{C\alpha-C} = \sqrt{\sum_{j=1}^N (d_j - d_{C\alpha-C})^2} / N$$

Fórmula (1)

$$rmsd = \sqrt{\sum_{j=1}^N d_j^2} / N$$

Fórmula (2)

Para computar RMSDs significativos, as duas estruturas em consideração devem primeiro ser superpostas, desde que possível. A superposição das estruturas de proteínas começa geralmente com uma comparação de seqüências. A comparação de seqüências define as relações um-a-um entre os pares de átomos onde o RMSD é computado. As relações átomo-a-átomo, para fins de comparação de estruturas, podem ocorrer na verdade entre resíduos que não estão na mesma posição relativa na seqüência de aminoácidos. As inserções e deleções da seqüência podem forçar duas seqüências a ficarem sem registro entre si, enquanto a arquitetura central das duas estruturas permanece similar.

Uma vez definida as relações átomo-a-átomo entre duas estruturas, com um programa de sobreposição como o LSQKAB do pacote CCP4 (Collaborative Computational Project N° 4., 1994) alcançamos uma superposição ótima entre as duas estruturas, isto é, a superposição com o menor RMSD possível. A sobreposição de um par de átomos pode perfeitamente deixar outro par de átomos à parte. Os algoritmos de superposição otimizam a orientação e a posição espacial de duas moléculas (translação e rotação) entre si. Uma vez efetuadas as superposições ótimas de todos os pares de estruturas, os valores de RMSD que são computados como resultado podem ser comparados entre si, já que as estruturas foram removidas para a mesma estrutura de referência antes de fazer os cálculos de RMSD (GIBAS & JAMBECK, 2001).

3.4.5. GRASP

Muitas das atividades de um organismo estão baseadas nas atividades das proteínas mediadas pelas suas interações com o meio: ácidos nucleicos, água, sal, outras proteínas e

inúmeras moléculas presentes nos sistemas. Essas interações surgem a partir de forças não covalentes. Entre as forças não covalentes existentes estão a interação eletrostática. Ligações eletrostáticas (iônicas) resultam da atração eletrostática entre dois grupos ionizados com cargas opostas, tais como os grupos carboxila (COO^-) e amina (NH_3^+). Na água, essas ligações são fracas, mas num ambiente hidrofóbico tal como o contato proteína-proteína, elas ficam mais fortes.

Podemos pensar na eletrostática molecular de uma proteína como um grupo de cargas elétricas em um meio dielétrico. No modelo que computa potenciais eletrostáticos para moléculas de proteínas, cada átomo é representado como um ponto com uma carga atômica parcial. A superfície acessível ao solvente da proteína forma o limite entre o meio interior da proteína e o meio exterior que circunda a proteína (GIBAS & JAMBECK, 2001). A eletrostática molecular para modelagem molecular é algo muito interessante a ser estudado na proteína. Porque a especificidade e afinidade entre uma enzima e seu inibidor dependem da direção das ligações de hidrogênio e das interações iônicas, bem como a complementaridade de carga e de forma de contato de superfície.

Na complementaridade de forma o sítio de ligação do substrato ou ligante depende de um entalhe ou sulco na superfície de uma proteína, complementar ao formato do substrato ou ligante. Na complementaridade de cargas, os resíduos de aminoácidos que formam o sítio de ligação estão organizados de modo a formar interações de atração especificamente com substratos ou ligantes. As moléculas que se diferem do substrato no formato ou distribuição de cargas não podem ligar-se produtivamente a proteína.

O GRASP é um programa de análise e visualização de moléculas particularmente útil para a visualização e manipulação das superfícies das moléculas e suas propriedades eletrostáticas. O programa contém algoritmos para a resolução da equação Poisson-Boltzmann que permite uma descrição do campo eletrostático gerado pelas cargas



existentes na molécula. O GRASP é um programa de visualização da superfície molecular de alta qualidade, desenvolvido pelo grupo de Barry Honig na Universidade da Colômbia, escrito para máquinas Silicon Graphics e tem uma linguagem própria, capaz de gerar a superfície da proteína, calcula o potencial eletrostático desta e escreve um arquivo com os valores de potencial eletrostático calculados para cada átomo.

O GRASP também pode ler arquivos de potenciais eletrostáticos e exibi-los como características de uma superfície molecular, além de ter muitas outras opções de exibição para criação de interpretações visuais didáticas bonitas das propriedades eletrostáticas.

3.5. Modelagem das Regiões de Alças

As diferenças funcionais entre membros de uma mesma família protéica são, em geral, consequência de diferenças estruturais na superfície externa das proteínas. Estas diferenças provêm de substituições, eliminações e inserções de resíduos nas cadeias de proteínas homólogas, principalmente nas alças que são as regiões mais expostas da proteína. Por serem regiões estruturalmente variáveis, as alças geralmente determinam a especificidade das proteínas, contribuindo para a estrutura dos sítios de ligação (no caso de enzimas, sítios ativos). Conseqüentemente, a modelagem das alças é uma etapa essencial na geração de modelos protéicos por homologia (FISER, 2000).

Os erros nas alças são o problema dominante em modelagem comparativa acima da identidade da seqüência de 35%. Nesta escala de similaridade total, as alças entre os homólogos variam quando as regiões do núcleo relativamente forem ainda conservadas e alinhadas exatamente (FISER *et al.*, 2002). Os métodos de predição de alças classificam-se em: métodos *ab initio* (FISER, 2000; MOULT & JAMES, 1986; BRUCCOLERI & KARPLUS, 1987) e métodos comparativos (GREER, 1981; JONES & THIRUP, 1986;

3.6. Modelagem das Cadeias Laterais



VAN VLIJMEN & KARPLUS, 1997; CHOTHIA & LESK, 1986; MARTIN & THOMTON, 1996).

O módulo de modelagem de laço do MODELLER executa a otimização baseada na aproximação (FISER, 2000). As principais razões são a generalidade e simplicidade conceitual da minimização de energia, assim como as limitações impostas na aproximação da base de dados por um número relativamente pequeno de estruturas de proteínas conhecidas (FISER *et al.*, 2002, FISER *et al.*, 2000).

O método otimiza as posições de todos os átomos de não hidrogênio de um laço em um ambiente fixo. A otimização confia em um protocolo que consiste na minimização do gradiente conjugado e na simulação molecular da dinâmica com *simulated annealing* (FISER *et al.*, 2003). A função da pseudo-energia tem termos de um campo força mecânica molecular bem como restrições baseadas nas distribuições estatísticas derivadas de estruturas de proteínas conhecidas. Ligações, ângulos, alguns ângulos diedros e ângulos diedros impróprios são contidos pelos termos correspondentes na função potencial do CHARMM-22 (MACKERELL *et al.*, 1998, FISER *et al.*, 2003). Os ângulos diedros da cadeia principal e lateral bem como os pares de átomos não ligados são restringidos pelos potenciais estatísticos extraídos de muitas estruturas de proteínas conhecidas (SALI & BLUNDELL, 1993; SALI & OVERINGTON, 1994; FISER *et al.*, 2000; MELO & FEYTMANS, 1997; FISER *et al.*, 2003). O método de modelagem de alças no MODELLER inclui também o efeito de solvatação apenas indiretamente, com o potencial estatístico da força média usado restringir o contato entre átomos não ligados de uma proteína (FISER *et al.*, 2002).

3.6. Modelagem das Cadeias Laterais



Como no caso das alças, as conformações das cadeias laterais também são modeladas a partir de dados estéricos e energéticos de estruturas semelhantes obtidos experimentalmente. A modelagem é feita com o auxílio de bancos de dados de rotâmeros permitidos de cadeias laterais (χ_n), derivados de estruturas terciárias do PDB em alta resolução (FILHO & ALENCASTRO, 2003). A Figura 21 define esquematicamente os ângulos torcionais de proteínas envolvidos na análise.

As bibliotecas de rotâmeros são conjuntos de parâmetros usados especificadamente para a otimização de posições das cadeias laterais na criação de modelos moleculares. São denominadas bibliotecas de rotâmeros porque contêm informações sobre as rotações permitidas dos átomos das cadeias laterais de aminoácidos remotos ao redor da ligação C α -C β , expressos como os valores permitidos de ângulos diedros de cadeias laterais (GIBAS & JAMBECK, 2001).

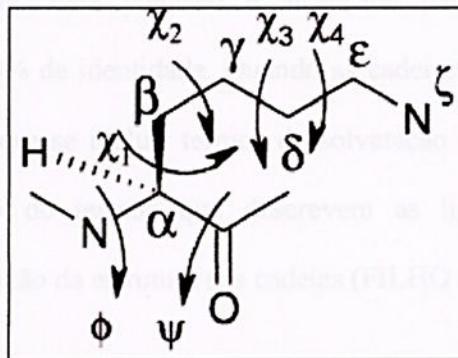


Figura 22 – Definição dos ângulos torcionais da cadeia principal (Φ e Ψ) e das cadeias laterais (χ_n) (FILHO & ALENCASTRO, 2003).

Devido às restrições estéricas na rotação de ligação, as cadeias laterais de aminoácidos em proteínas assumem apenas algumas conformações sem consequências energéticas desfavoráveis. As bibliotecas de rotâmeros podem ser derivadas por meio de restrições de ângulos e de ligações químicas, mas estão mais propensas a serem desenvolvidas pela análise das conformações assumidas pelas cadeias laterais de aminoácidos em estruturas

protéicas conhecidas. As bibliotecas de rotâmeros podem ser dependentes ou independentes da estrutura principal. As bibliotecas independentes da estrutura principal classificam todos os exemplos de um determinado aminoácido como parte do mesmo conjunto, mesmo se uma ocorrência estiver dentro de uma folha- β e outra estiver dentro de uma hélice- α . Por outro lado, as bibliotecas dependentes da estrutura principal ainda classificam os aminoácidos, de acordo com a sua ocorrência em estruturas secundárias específicas (GIBAS & JAMBECK, 2001).

A utilização de bibliotecas de rotâmeros de cadeias laterais é um dos métodos mais eficientes de predição da estrutura das cadeias laterais de proteínas. Elas permitem estimar a probabilidade de cada cadeia lateral assumir uma determinada conformação (χ_1) em função dos ângulos conformacionais da cadeia principal (ψ e ϕ) e do tipo de resíduo. A precisão aproximada do método é de 82% para o ângulo diedro χ_1 e de 72% para os diedros χ_1 e χ_2 , tomados simultaneamente. Esta precisão é alcançada para pares de estruturas que compartilham de 30 a 90% de identidade. Quando as cadeias laterais estão localizadas na superfície da proteína, deve-se incluir termos de solvatação no processo de modelagem. Além disso, a inclusão de termos que descrevem as ligações hidrogênio melhora significativamente a predição da estrutura das cadeias (FILHO & ALENCASTRO, 2003).



6. RESULTADOS E DISCUSSÃO

Foi gerado um total de 1000 modelos para cada enzima da via metabólica do ácido chiquímico de *X. fastidiosa*, e foi escolhido o melhor modelo com o critério de avaliação de qualidade estereoquímica utilizando o PROCHECK (LASKOWSKI *et al.*, 1994) e a função objetiva do MODELLER (SALI & BLUNDELL, 1993).

6.1. Comparação da Seqüência Primária

A porcentagem de identidade da sequência primária (Tabela 2) entre as enzimas da via do ácido chiquímico de *Xylella fastidiosa* (alvos) e dos moldes indica que estas estruturas 3D são bons modelos para usarmos como moldes na modelagem das enzimas da via do ácido chiquímico de *Xylella fastidiosa* (alvos), pois apresentam identidade > 30% (com uma exceção). Os valores encontrados para similaridade entre os moldes e seus respectivos modelos (alvos) ficaram > 60% (Tabela 2).

O alinhamento entre as enzimas da via do ácido chiquímico de *Xylella fastidiosa* (alvos) e dos moldes são mostrados nas figuras 23 A-G. Para gerar o alinhamento da sequência primária entre alvo e o modelo utilizamos o programa FASTA (WILBUR & LIPMAN, 1983; LIPMAN & PEARSON, 1985). Nas figuras de alinhamento, espaços (*gaps*) são representados por linhas tracejadas, os resíduos idênticos são representados por dois pontos, os resíduos similares são representados por um ponto e os resíduos não idênticos são representados por espaços vazios.

A

	10	20	30	40	50	60
1GG1	MNYQNDLRIKEIKELLPFVALLEKFPATENAANTVAHARKAIHKILKGNDRLLVVGPCSIH					
	10	20	30	40	50	60
XfDAHP	MSPPLGLCLNVLRLHMPFFTDLRIRKIEPLTPFNQLLQLPCDEQASLTVATSRNALHDLHGHNDRLLAVVIGPCSIH					
	70	80	90	100	110	120
1GG1	DPVAAKEYATRLALREELKDELEIVMRVYFEKPTTVGWGLINDPHMNSFQINDGLRIARKLLDINDSGLPAGEF					
	70	80	90	100	110	120
XfDAHP	DPVAAKEYATRLALREELKDELEIVMRVYFEKPTTVGWGLINDPHMNSFQINDGLRIARKLLDINDSGLPAGEF					
	130	140	150	160	170	180
1GG1	LDMITPQYLADLMSWGAIGARTTESQVHRELASGLSCPVGKNGTDGTIKVAIDAINAAGAPHCFLSVTKWGHSAIVNTS					
	130	140	150	160	170	180
XfDAHP	LDMITPQYLADLMSWGAIGARTTESQVHRELASGLSCPVGKNGTDGTIKVAIDAINAAGAPHCFLSVTKWGHSAIVNTS					
	190	200	210	220	230	240
1GG1	LDMITPQYLADLMSWGAIGARTTESQVHRELASGLSCPVGKNGTDGTIKVAIDAINAAGAPHCFLSVTKWGHSAIVNTS					
	190	200	210	220	230	240
XfDAHP	LDMITPQYLADLMSWGAIGARTTESQVHRELASGLSCPVGKNGTDGTIKVAIDAINAAGAPHCFLSVTKWGHSAIVNTS					
	250	260	270	280	290	300
1GG1	GNMGDCHIILRGKKEPNYSAKHVAEVKEGLNKAQVNIIDFSHANSSEKQFKKQMDVCADVCCQIAGGEKAIIGVMVESH					
	250	260	270	280	290	300
XfDAHP	GNMGDCHIILRGKKEPNYSAKHVAEVKEGLNKAQVNIIDFSHANSSEKQFKKQMDVCADVCCQIAGGEKAIIGVMVESH					
	310	320	330	340	350	360
1GG1	LVEGNQSLSEGEPLAYGKSITDACIGWEDTDALLRQLANAVKARRG					
	310	320	330	340	350	360
XfDAHP	LVEGNQSLSEGEPLAYGKSITDACIGWEDTDALLRQLANAVKARRG					
	370	380	390	400	410	420
1GG1	IVGGRQELVGLNTLTLYGQSITDACIDWDTSVAMLEQLAQAVRVQTLPTTATINKTSKTQ					
	370	380	390	400	410	420
XfDAHP	IVGGRQELVGLNTLTLYGQSITDACIDWDTSVAMLEQLAQAVRVQTLPTTATINKTSKTQ					

B

	10	20	30	40	50	60	70
1DQS	MSNPTKISILGRESIIADFLWRNYVAKLISDCSSTTVLVTDNIGSIYTPSFEEAF-RKRAAEITPSPRLLIYNRP						
	10	20	30	40	50	60	70
XfDHQS	TLTPRLSRVTNNYPTTIAIGPGLLHDLPLAATIRGRHALILSDSEVAPRYAAQLQETLLRAR----PDLHLNVFTLP						
	80	90	100	110	120	130	140
1DQS	PGEVSKSROTKADIEDWMLSNPPCGRDTVVIALGGVIGDLTGTFVASTYMRGVRYVOVPTTLLAMVDSSIGGKTAIDTP						
	80	90	100	110	120	130	140
XfDHQS	PGEVSKSROTKADIEDWMLSNPPCGRDTVVIALGGVIGDLTGTFVASTYMRGVRYVOVPTTLLAMVDSSIGGKTAIDTP						
	150	160	170	180	190	200	210
1DQS	AGETSKSLENFSAIAQLATLGAT--RDACLFALGGVIGDLGAGTAAACWNRGIDYVQVPTTLLAMVDSSVGGKTAIDTP						
	150	160	170	180	190	200	210
XfDHQS	AGETSKSLENFSAIAQLATLGAT--RDACLFALGGVIGDLGAGTAAACWNRGIDYVQVPTTLLAMVDSSVGGKTAIDTP						
	220	230	240	250	260	270	280
1DQS	LGNLIGAIWQPTKIYIDLEFLETLPVREFINGMAEVIKTAASISSEEEFTALEENAETILKAVRREVTPEHREFEGTEEI						
	220	230	240	250	260	270	280
XfDHQS	LGNLIGAIWQPTKIYIDLEFLETLPVREFINGMAEVIKTAASISSEEEFTALEENAETILKAVRREVTPEHREFEGTEEI						
	290	300	310	320	330	340	350
1DQS	OGKVMVGFHPPRAVIADTDLTATLPLRELRLAGLSEVIKYGAIRDVPVFFHWLOTTREALLA---RD--PA-----A						
	290	300	310	320	330	340	350
XfDHQS	OGKVMVGFHPPRAVIADTDLTATLPLRELRLAGLSEVIKYGAIRDVPVFFHWLOTTREALLA---RD--PA-----A						
	360	370	380	390	400	410	420
1DQS	LKARILASARHKAYVVSADEREGLRLNLLNMGHSIGHAIEAIL---TP---CILHGECVAIGHMVKEAELARHLGILKGVA						
	360	370	380	390	400	410	420
XfDHQS	LKARILASARHKAYVVSADEREGLRLNLLNMGHSIGHAIEAIL---TP---CILHGECVAIGHMVKEAELARHLGILKGVA						
	430	440	450	460	470	480	490
1DQS	LAQATARSCEHKADIVGRDPLEKGERVLLNLGHTFGHAIEAAGYSTPGSNLNLHGEAVAVGMVLAARLNLALSLAPAQD						
	430	440	450	460	470	480	490
XfDHQS	LAQATARSCEHKADIVGRDPLEKGERVLLNLGHTFGHAIEAAGYSTPGSNLNLHGEAVAVGMVLAARLNLALSLAPAQD						
	500	510	520	530	540	550	560
1DQS	VSRIVKCLAAAYGLPTSLKADARIRKLTAGKHCSDVQLMFMNALDKKNDGPKKKIVLLSAIGTPYETRASVAVANEDIRVVLA						
	500	510	520	530	540	550	560
XfDHQS	VSRIVKCLAAAYGLPTSLKADARIRKLTAGKHCSDVQLMFMNALDKKNDGPKKKIVLLSAIGTPYETRASVAVANEDIRVVLA						
	570	580	590	600	610	620	630
1DQS	TETLNLLDAYGLPTVLPSG---LTP-----EMLLRMLDKKNIAGRLRLVLWRGIG---HAEAVSDVDEATVRQILA						
	570	580	590	600	610	620	630
XfDHQS	TETLNLLDAYGLPTVLPSG---LTP-----EMLLRMLDKKNIAGRLRLVLWRGIG---HAEAVSDVDEATVRQILA						
	640	650	660	670	680	690	700
1DQS	P						
	640	650	660	670	680	690	700
XfDHQS	P						

C

	10	20	30	40	50	60	70
1GU1	PRSLANAPIILNNGPNLNLGQROPEIYGSDTLADVEALCVKAAAAGGTVDVFRQSNHEGELVDWIHEARLNHCG-IVIN						
	10	20	30	40	50	60	70
xfdhqd	MAHLLLLHGPNNLLGTREPEIYGRITLPQIDAALAERATAAGHGLSSLOSNAEHVLIERIHATPEDGTAFILIN						
	80	90	100	110	120	130	140
1GU1	PAAYSHTSVAILDALNTCDGLPVVEVHISNIHQREPFRRHSTVSQRADGVVAGCGVQGVFVERIAALAGAGSARA						
	80	90	100	110	120	130	140
xfdhqd	PAAYSHTSVAILDALNTCDGLPVVEVHISNIHQREPFRRHSTVSQRADGVVAGCGVQGVFVERIAALAGAGSARA						
	150	160	170	180	190	200	210
1GU1	PGATHTSVALRDALLAV-ALPFVEIHLNPNYTPREPFRRHSTYLADKALGVVCGFVDSYRIALEGVIAIRLGSVD						
	150	160	170	180	190	200	210
xfdhqd	PGATHTSVALRDALLAV-ALPFVEIHLNPNYTPREPFRRHSTYLADKALGVVCGFVDSYRIALEGVIAIRLGSVD						



D

	10	20	30	40	50	60	70
1NYT	METAYAVFNGPIAHKSPPFIHQFQAQQLNIEHPYGRVLAPINDFINTLNAFFSAGGKGANVTVPFKEEAFARADELTER						
XfSD	MPVSRFAVFGHPIAHSLSPRIHTEFGQMGVALNYLAFDVPDAFVSLERFVAEGGCGANVTLPKKAAPFVCTTLSAR						
	10	20	30	40	50	60	70
	80	90	100	110	120	130	140
1NYT	AALAGAVNTLERLEDGRLLGENTDGVGLLSDL-ERLSFIRPGLRILLIGAGGASRGVLLPLLSLDCAVT---ITNRTVSR						
XfSD	ARRACAVNTLSRV-DGVVHGENTDGTGLVRNLTERHGLDLGRRAILLGAGGAARGVA-PAL-LDAGITENVIVNRSPER						
	90	100	110	120	130	140	150
	160	170	180	190	200	210	220
1NYT	AEELAKLFAHTGSIQALSMDL-EGHEFDLIINATSSGISGDIP--AIPSSLIHPGIYCYDNFYQKGTPTFLAWCEQRGS						
XfSD	ADMLCDALGEPGRVSARYWGDLGDLGNFELIVNATSIGNTSDMRTFSLPRSLDSEHTAAVDLNYGSAAPFLAWAHAVET						
	160	170	180	190	200	210	220
	240	250	260	270			
1NYT	KRNADGLGMLVAQAAHAFLVHGVLPDVEPVIKOLOEELS						
XfSD	RYVIDGLGMLVEQAAEFSLVHGRRPDTPVYTVLHSEYGAPGRS						
	240	250	260	270	280		

E

	10	20	30	40	50	60	70
1WE2	MAPKAVLVGLPGSGKSTIGRRLLAKALGVGLLDTDAIEQRTGRSIADIFATDGEQEFRIEEDVVRAALADHDGVLSL						
XfSHK	MNPAPNLVHIGPMGAKSGSIGRRIAKHFLHFADTDHAIIVERAGTNISAIKYSGEPEFRRLEREVLHDDLHNHENQLIAT						
	10	20	30	40	50	60	70
	80	90	100	110	120	130	140
1WE2	GGGAVTSPGVRAALAGHT-VVYLEISAAEGVRRRTGGNTVRPFLAGPDRAEKYRALMAKRAPLYRRVATHRVDTNRRNPGA						
XfSHK	GGGTILDENRRCHQERGFVVFLKINVNTQLERLAHDYRPLQQIDRKQVLSLYATRQPLQQIADHIVTTDHMSPT						
	90	100	110	120	130	140	150
	160	170					
1WE2	VVRHILSRLQVPSPEAAAT						
XfSHK	ATAQLILDLTAAHWQSSNAA						
	170	180					

F

	10	20	30	40	50	60
1G6S	MESLTLQPIARVDGTINLPKSKSVNRALLAALAHGKTVLTNLLSDDDVRHMLNALTALGVLS					
XfEPSP	MYCRSHLKKPMSHRTHDYVIAHQGTPLHGVLSIPGDKSISHRAVHFAALADGTSRIDGFLEAEDTCSAEILARLGVR					
	10	20	30	40	50	60
	70	80	90	100	110	120
1G6S	YT--LSADRTCEIIGNGGPHAEAGALEFLGNAGTAMRPLAALCLGSDIVLTGEPNKRERPIGHLVDALRLGGAKIT					
XfEPSP	IETPLSTQRI-VHGVGVGDG-LQA-SHPLDCGNAGTGHRLLAGLLVAQPFDSVLVGDASLSKRPVRRVTDPLSQMGARID					
	90	100	110	120	130	140
	150	160	170	180	190	200
1G6S	YLEQENYPPLRLQGGFTGGNVN-VDGSSVSQFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEINQHY					
XfEPSP	-TSDDGTPPLRIYGGQLLHGIDFISPVASAIKSAVLLAGLYARNETVVR----EPHPTRDYTERMLTAFGVDIDVSTG					
	160	170	180	190	200	210
	230	240	250	260	270	280
1G6S	QQFVVKGQSYQSPGTYLVEGDASSASYFLAAAA-IKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICWGDDYISCTRG					
XfEPSP	CARL-RGGQRCLCATDI-TIPADFSSAAFYLVAAASVIPGSDITLRAVGLNPRR--IGLLTVLRLHGANIVESNRH---EQG					
	240	250	260	270	280	290
	300	310	320	330	340	350
1G6S	ELNAIDMDMNHIP-----D--AAHTIATAALF-----AKGTTTTLNIYNURVKETDRLFAMATELRKVGAEVEEGHD					
XfEPSP	GEPVVDLRVRYAPLOGTRVPEDLVADHIDEFPALFVAAAAAEGQTVVSGAAELRVKESDRLAAMVTGLRVLGQVDETAD					
	310	320	330	340	350	360
	370	380	390	400	410	420
1G6S	YIRITPPEKLNFAEIAATYNDHRNMCFLSLVA-LSDTPVTILDPKCTAKTFPDYFEQLARISQAA					
XfEPSP	GATIHGGP-IGHGTINSHGDHRIAMAFSIAGQLSVSTVRIEDVANVATSFDPY-ETLARSAGFGLVEYCDPA					
	390	400	410	420	430	440
	450					

G

	10	20	30	40	50	60	70
1QXO	MRYLTAGESHGPRLTAIIIEGIPAGLPLTAEDINEDLRRRQGGYGRGGRMKIENDQVFTSGVRHGKTTGAP						
XfCS	MGANTFGKLLAVTTFGESHGPAIGCVIDGCPPGLELAAEEFAHDLQRRATGRSRHTSARREADEVEILSGVVEGRTTGTP						
	10	20	30	40	50	60	70
	80	90	100	110	120	130	140
1QXO	ITMDVINKDHOKWLDMSAEDIEDRLKSKRKITHPRFGHADLVGGIKYRFDDLRNSLERSSARETTNRVAAGVAVAKRLLA						
XfCS	IALLRINTDQRS-----KDYATIAARQF---RPGHADYTYWQKYGIRDPRGG--GRSSARETTNRVAAGVAVAKRLLA						
	90	100	110	120	130	140	
	160	170	180	190	200	210	220
1QXO	ELDMEIANHVVFVGGKEIDVPENLTVAEIKQRAAQSEVSIVNQERE-QEIKDYIDQIKRDGDTIGGVVETVVGVPVGLG						
XfCS	QRYGVIVRGFLSQLG-EIR-PEGFA-----WDAVEDNPFVFPQAAQVPELEAYMDALRKSGNSVGARVDVVAEGVPPGWG						
	150	160	170	180	190	200	210
	240	250	260	270	280	290	300
1QXO	STVQWDRKLDARLAQAVVSIKAFKGVFGLGFEAGYRKGSGVMDLWSEKEDGYTERTNNLGGFEGGTMNGQPIVVRGVM						
XfCS	EPIY--GKLDGELAAALMSINAVKGVIEGAGFGSTVQKGTETRDLMT---PLGFL--SNHAGGIIGGITTGQPIIVSIAL						
	220	230	240	250	260	270	280
	320	330	340	350	360	370	380
1QXO	KPIPTLYKPLMSVDIETHEPYKATVERSPTALPAAGHVEAVVATVLAQEILEKFSNDNLEELKEAVAKHRDYTKNY						
XfCS	KPTSSRLRLPGETVDVDGHPVQVITKGRHDPVCGIRAFPIAEANVALVLMQALRHRAQCQGVGEMSPCILENVGFRNADD						
	300	310	320	330	340	350	360
	300	310	320	330	340	350	370

Figura 23 – Resultado do alinhamento das seqüências primárias para (A) DAHPS de *X. fastidiosa* e *E. coli* (1GG1); (B) DHQS de *X. fastidiosa* e *E. nidulans* (1DQS); (C) DHQD de *X. fastidiosa* e *S. celicolor* (1GU1); (D) SD de *X. fastidiosa* e *E. coli* (1NYT); (E) SK de *X. fastidiosa* e *M. tuberculosis* (1WE2); (F) EPSPS de *X. fastidiosa* e *E. coli* (1G6S) e (G) CS de *X. fastidiosa* e *S. pneumoniae*.

6.2. Qualidade dos Modelos

Avaliar a qualidade do modelo é fundamental. A qualidade do modelo construído será proporcional à qualidade da estrutura 3D utilizada como molde (SANCHEZ *et al.*, 1998), por isso deve-se avaliar o modelo como um todo e por região.

A precisão do modelo gerado determina sua utilidade. Como a precisão da modelagem comparativa de proteínas está relacionada à porcentagem de identidade no qual o modelo é baseado na correlação entre as similaridades estrutural e seqüencial das duas proteínas e na qualidade estereoquímica, foi verificado que os modelos finais (alvos) apresentam uma precisão média á alta com base nos valores obtidos nas tabelas 2-7.

Os procedimentos de validação e avaliação que serão descritos a seguir foram realizados para cada modelo final das enzimas da via do ácido chiquímico. Os modelos (alvos) foram construídos baseados nas coordenadas atômicas de estruturas resolvidas por

biocristalografia (Tabela 2) com resolução $\leq 2,0$ Å (com uma exceção). As coordenadas atômicas de todas as moléculas de água foram retiradas dos moldes.

Tabela 2 – As estruturas cristalográficas usadas como molde e a porcentagem de identidade entre as seqüências primárias dos moldes e modelos de *X. fastidiosa*.

Enzimas	Código de Acesso PDB	Resolução (Å)	Organismo	Similaridade (%)	Identidade (%)	Referências
DAHPS	1GG1	2,00	<i>Escherichia coli</i>	83,1	61,1	WAGNER <i>et al.</i> , 2000
DHQS	1DQS	1,80	<i>Emericella nidulans</i>	62,2	31,8	CARPENTER <i>et al.</i> , 1998
DHQD	1GU1	1,80	<i>Streptomyces coelicolor</i>	74,0	42,5	ROSZAK <i>et al.</i> , 2002
SD	1NYT	1,50	<i>Escherichia coli</i>	65,2	42,3	MICHEL <i>et al.</i> , 2003
SK	1WE2	2,30	<i>Mycobacterium tuberculosis</i>	69,3	32,6	PEREIRA <i>et al.</i> , 2004
EPSPS	1G6S	1,50	<i>Escherichia coli</i>	60,3	29,9	SCHÖNBRUN N <i>et al.</i> , 2001
CS	1QXO	2,00	<i>Streptococcus pneumoniae</i>	61,4	30,6	MACLEAN & ALI <i>et al.</i> , 2003

Após a etapa inicial de escolha do melhor modelo, análises mais criteriosas e detalhadas foram efetuadas com a utilização do VERIFY 3D, WHATCHECK, RMSD da geometria ideal, RMSD da sobreposição $\text{Ca} - \text{Ca}$ e PROCHECK, a fim de verificar a qualidade dos modelos construídos. Os resultados obtidos dos modelos finais foram comparados com os valores das estruturas 3D (moldes). O programa PROCHECK foi utilizado para analisar a geometria global dos modelos através do diagrama de Ramachandran (Tabela 3 e Figura 24 A-G) e G-factor (Tabela 4).

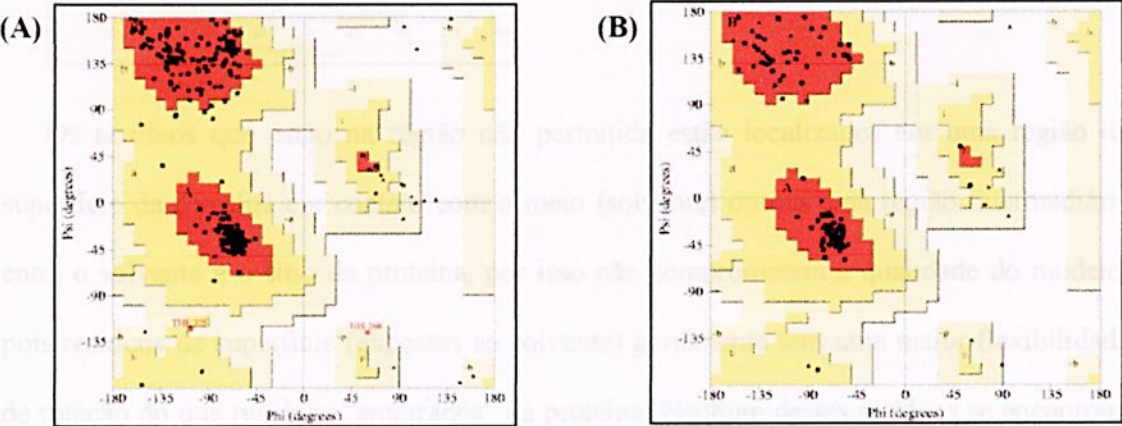
A análise do diagrama de Ramachandran mostrou que mais de 89,0% dos resíduos estão sobre a região mais favorável. Os modelos apresentam uma qualidade estereoquímica ligeiramente inferior aos seus respectivos moldes (Tabela 3), com exceção de três modelos (DAHPS, DHQD e SK).

Tabela 3 – Resultado das análises do diagrama de Ramachandran dos moldes e modelos de *X. fastidiosa*.

Enzimas	Região do Diagrama de Ramachandran			
	Mais Favorável (%)	Adicionalmente Permitida (%)	Generosamente Permitida (%)	Não Permitida (%)
DAHPS	94,4 (90,5)	4,9 (8,1)	0,7 (1,4)	0,0 (0,0)
DHQS	91,2 (92,5)	7,2 (7,5)	1,3 (0,0)	0,3 (0,0)
DHQD	96,8 (92,2)	3,2 (7,0)	0,0 (0,8)	0,0 (0,0)
SD	93,2 (93,3)	5,6 (6,7)	0,9 (0,0)	0,4 (0,0)
SK	93,8 (93,4)	5,6 (5,9)	0,6 (0,7)	0,0 (0,0)
EPSPS	89,0 (90,8)	7,1 (8,9)	2,5 (0,3)	1,4 (0,0)
CS	90,7 (93,1)	7,3 (6,9)	1,7 (0,0)	0,3 (0,0%)

*Entre os parêntesis estão os valores obtidos para as estruturas 3D usadas como moldes.

Nas figuras 24 A-G a região em vermelho é a mais favorável, a região amarelo escuro é a adicionalmente permitida, a região em amarelo claro é a generosamente permitida e a região em branco é a região não permitida. As glicinas são representadas por triângulos.



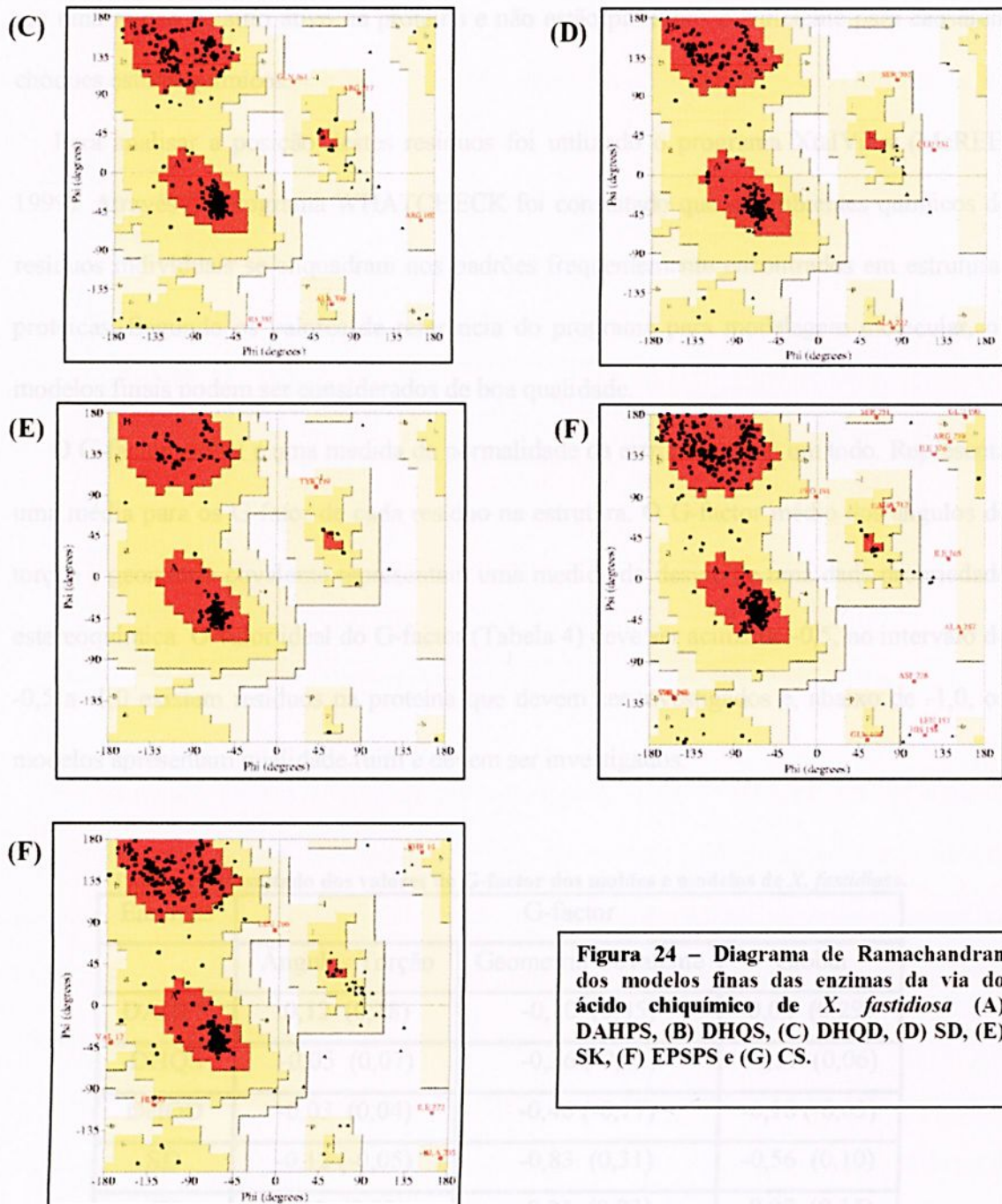


Figura 24 – Diagrama de Ramachandran dos modelos finas das enzimas da via do ácido chiquímico de *X. fastidiosa* (A) DAHPS, (B) DHQS, (C) DHQD, (D) SD, (E) SK, (F) EPSPS e (G) CS.

Os resíduos que estão na região não permitida estão localizados em uma região de superfície da proteína em contato com o meio (solvente) ou em uma região intermediária entre o solvente e o sítio da proteína, por isso não comprometem a qualidade do modelo, pois resíduos de superfície (expostos ao solvente) geralmente tem uma maior flexibilidade de rotação do que resíduos “enterrados” na proteína. Nenhum destes resíduos se encontram

em uma região de sítio ativo da proteína e não estão próximos o suficiente para causarem choques estereoquímicos.

Para analisar a posição destes resíduos foi utilizado o programa XtalView (McREE, 1999). Através do programa WHATCHECK foi constatado que os ambientes químicos de resíduos individuais se enquadram nos padrões freqüentemente encontrados em estruturas protéicas. Segundo os valores de referência do programa para modelagem molecular, os modelos finais podem ser considerados de boa qualidade.

O G-factor global é uma medida da normalidade da estrutura como um todo. Representa uma média para os G-fator de cada resíduo na estrutura. O G-factor médio dos ângulos de torção e geometria covalente representam uma medida do desvio de uma dada propriedade estereoquímica. O valor ideal do G-factor (Tabela 4) deve ser acima de -0,5, no intervalo de -0,5 a -1,0 existem resíduos na proteína que devem ser investigados e, abaixo de -1,0, os modelos apresentam qualidade ruim e devem ser investigados.

Tabela 4 – Resultado dos valores do G-factor dos moldes e modelos de *X. fastidiosa*.

Enzimas	G-factor		
	Ângulos Torção	Geometria Covalente	Global
DAHPS	0,12 (0,18)	-0,10 (0,45)	0,04 (0,29)
DHQS	-0,05 (0,07)	-0,36 (-0,02)	-0,17 (0,06)
DHQD	-0,03 (0,04)	-0,46 (-0,17)	-0,18 (-0,03)
SD	-0,12 (-0,05)	-0,83 (0,31)	-0,56 (0,10)
SK	0,02 (0,08)	-0,21 (0,23)	-0,07 (0,15)
EPSPS	-0,21 (0,08)	-0,87 (0,31)	-0,44 (0,18)
CS	-0,16 (0,00)	-0,73 (-0,16)	-0,37 (-0,04)

*Entre os parêntesis estão os valores obtidos para as estruturas 3D usadas como moldes.

Para verificar confiabilidade na compatibilidade sequência/estrutura foi utilizado o programa VERIFY 3D, os valores obtidos indicam que a estrutura dos modelos fina tem

compatibilidade entre a sequência primária do modelo e a estrutura 3D construída, pois todos os valores gerados para os modelos finais ficaram acima do limite de 0,45 (Tabela 5).

Tabela 5 – Resultado dos valores do VERIFY 3D dos moldes e modelos de *X. fastidiosa*.

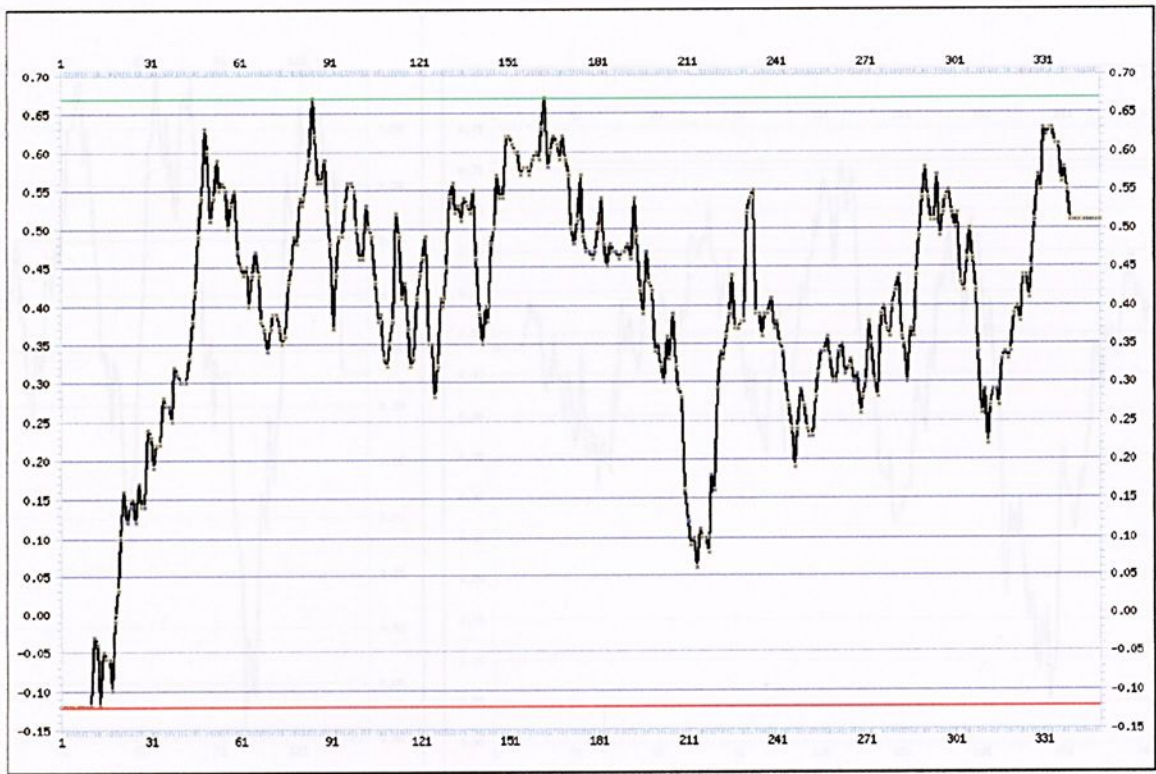
Enzimas	VERIFY 3D		
	Escore Total	Escore Ideal	Escore S_{scale}
DAHPS	135,23 (153,33)	159,94 (154,87)	0,84S (0,99S)
DHQS	150,27 (180,65)	169,15 (173,76)	0,89S (1,04S)
DHQD	58,79 (67,47)	67,17 (67,62)	0,88S (1,00S)
SD	125,98 (131,02)	125,42 (123,58)	1,00S (1,06S)
SK	69,16 (83,33)	81,82 (74,95)	0,85S (1,11S)
EPSPS	175,97 (230,44)	196,36 (195,44)	0,90S (1,18S)
CS	145,42 (163,31)	170,07 (171,92)	0,85S (0,95S)

*Entre os parêntesis estão os valores obtidos para as estruturas 3D usadas como moldes.

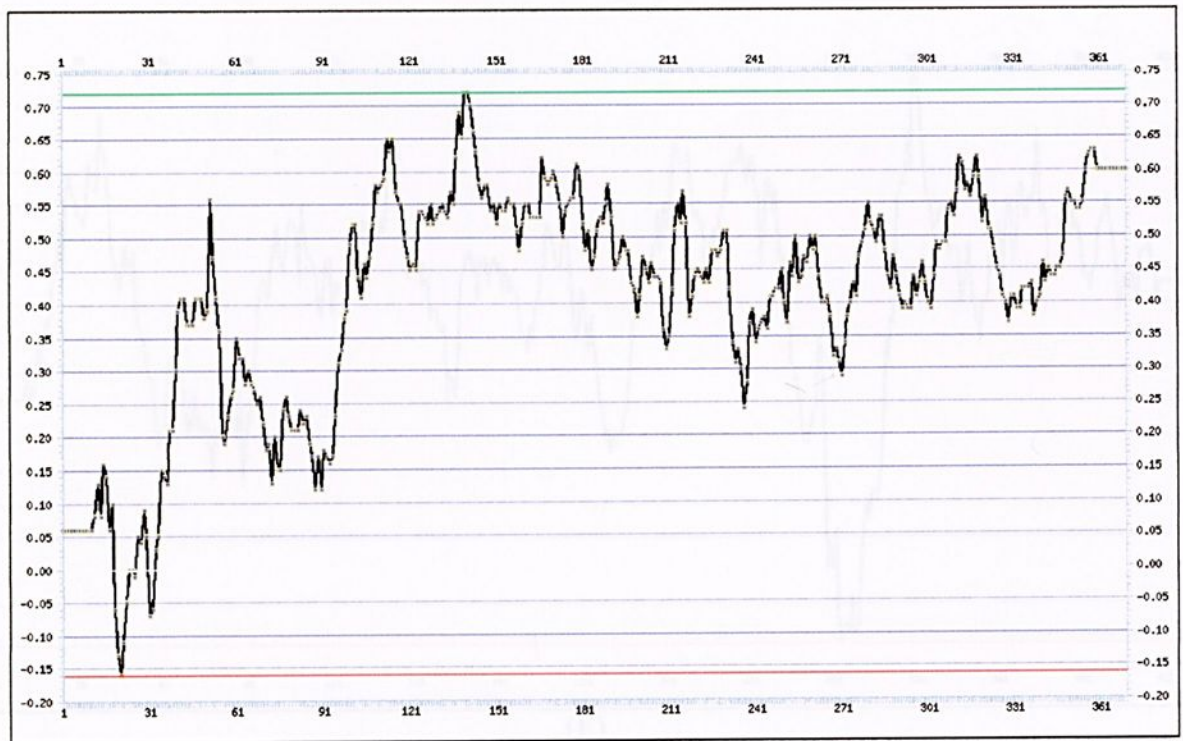
Foi verificado através do gráfico que o programa VERIFY 3D gera que em algumas regiões os aminoácidos dos modelos finais apresentam conformações desfavoráveis, permanecendo abaixo da linha zero. Nenhum desses aminoácidos localizados abaixo da linha zero pertence ao sítio de ligação da proteína, uma possível explicação é que estes aminoácidos podem estar em uma região de alça.

Esses resultados são mostrados nas figuras 25 A-G, onde a linha em vermelho corresponde ao menor ponto que chegou um resíduo e a linha verde o ponto mais alto que o resíduo obteve no gráfico. As regiões que estão abaixo do ponto zero no gráfico possivelmente apresentam problemas quanto ao ambiente no qual os aminoácidos estão contidos. Na faixa de valores na vertical estão representados os valores da média do escore 3D-1D e na faixa horizontal o número de resíduos da proteína.

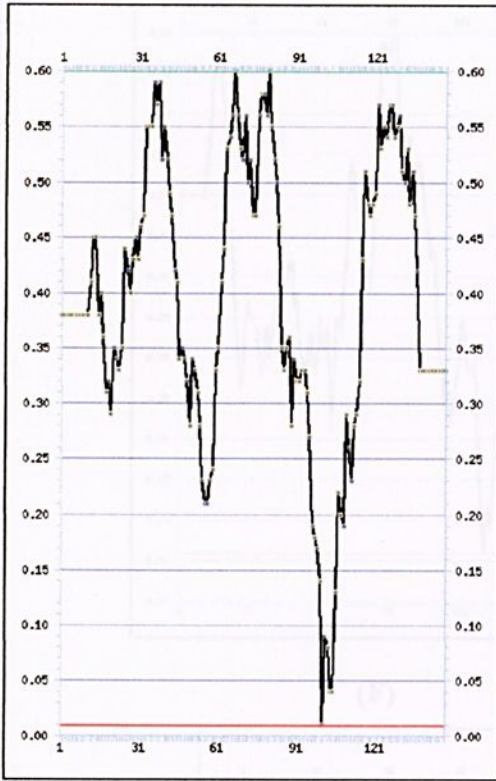




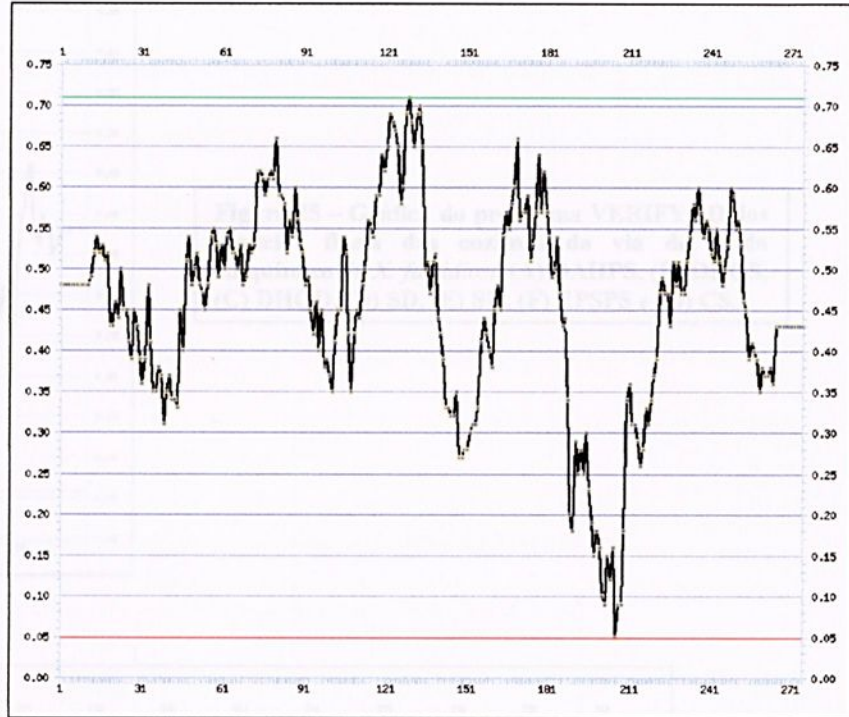
(A)



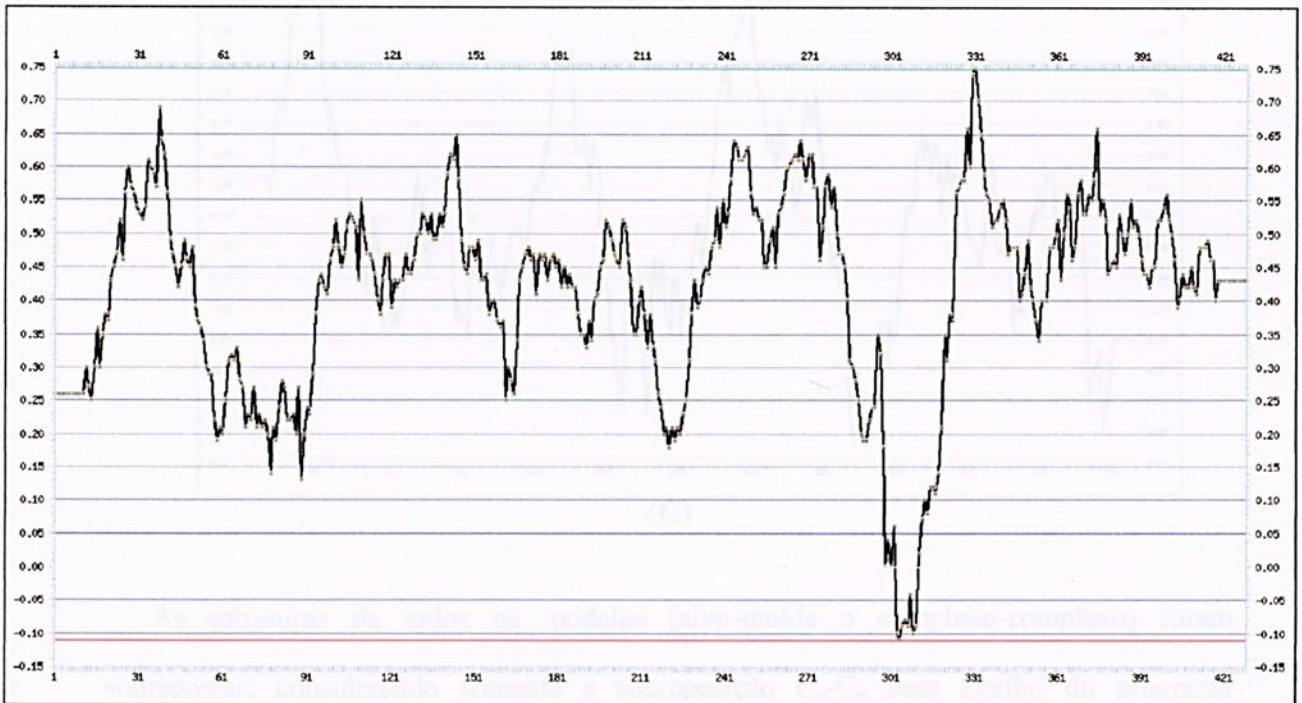
(B)



(C)



(D)



(E)

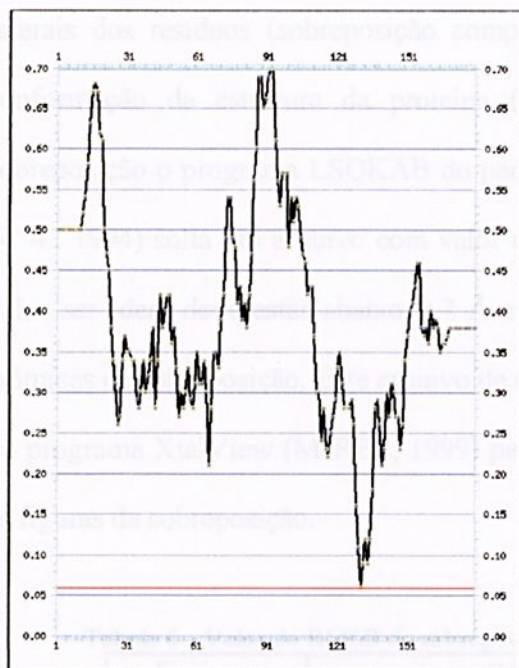
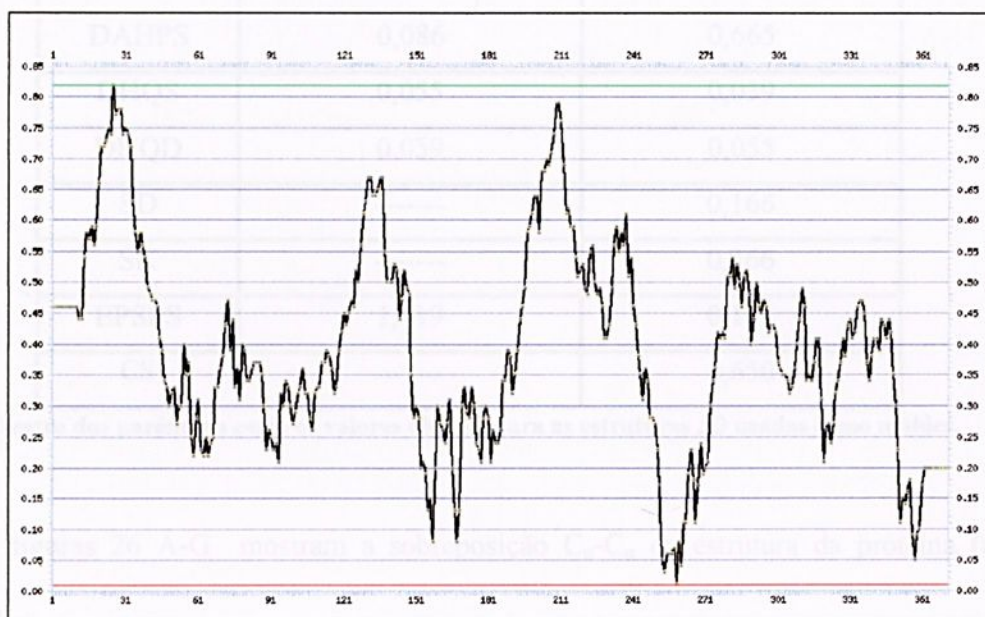


Figura 25 – Gráfico do programa VERIFY 3D dos modelos finais das enzimas da via do ácido chiquímico de *X. fastidiosa* (A) DAHPS, (B) DHQS, (C) DHQD, (D) SD, (E) SK, (F) EPSPS e (G) CS.

(F)



(G)

As estruturas de todos os modelos (alvo-molde e complexo-complexo) foram sobrepostas, considerando somente a sobreposição C_{α} - C_{α} com auxílio do programa LSQKAB do pacote CCP4 (Collaborative Computational Project N° 4., 1994). Através da sobreposição é possível verificar se há possíveis alterações no posicionamento das cadeias

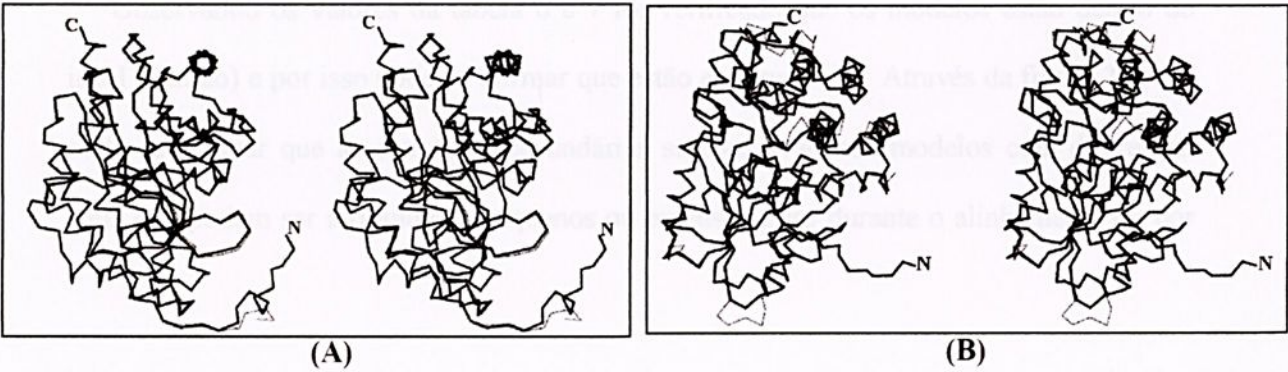
laterais dos resíduos (sobreposição complexo-complexo) ou se há alguma alteração na conformação da estrutura da proteína (sobreposição alvo-molde). Depois de feita a sobreposição o programa LSQKAB do pacote CCP4 (Collaborative Computational Project N° 4., 1994) solta um arquivo com valor do RMSD da sobreposição (Tabela 6), para este valor ser ideal deve estar abaixo 0,7 Å e também gera um arquivo com as coordenadas atômicas da sobreposição. Este arquivo de coordenadas atômicas foi utilizado como entrada no programa XtalView (McREE, 1999) para que fosse possível visualizar, analisar e gerar as figuras da sobreposição.

Tabela 6 – Valor do RMSD da sobreposição para os moldes e modelos de *X. fastidiosa*.

Enzimas	RMSD sobreposição (Å)	
	Sítio	Proteína
DAHPS	0,086	0,665
DHQS	0,055	0,039
DHQD	0,059	0,055
SD	-----	0,166
SK	-----	0,366
EPSPS	1,419	0,118
CS	-----	0,650

*Dentro dos parêntesis estão os valores obtidos para as estruturas 3D usadas como moldes.

As figuras 26 A-G mostram a sobreposição C_α-C_α da estrutura da proteína (alvo-molde).



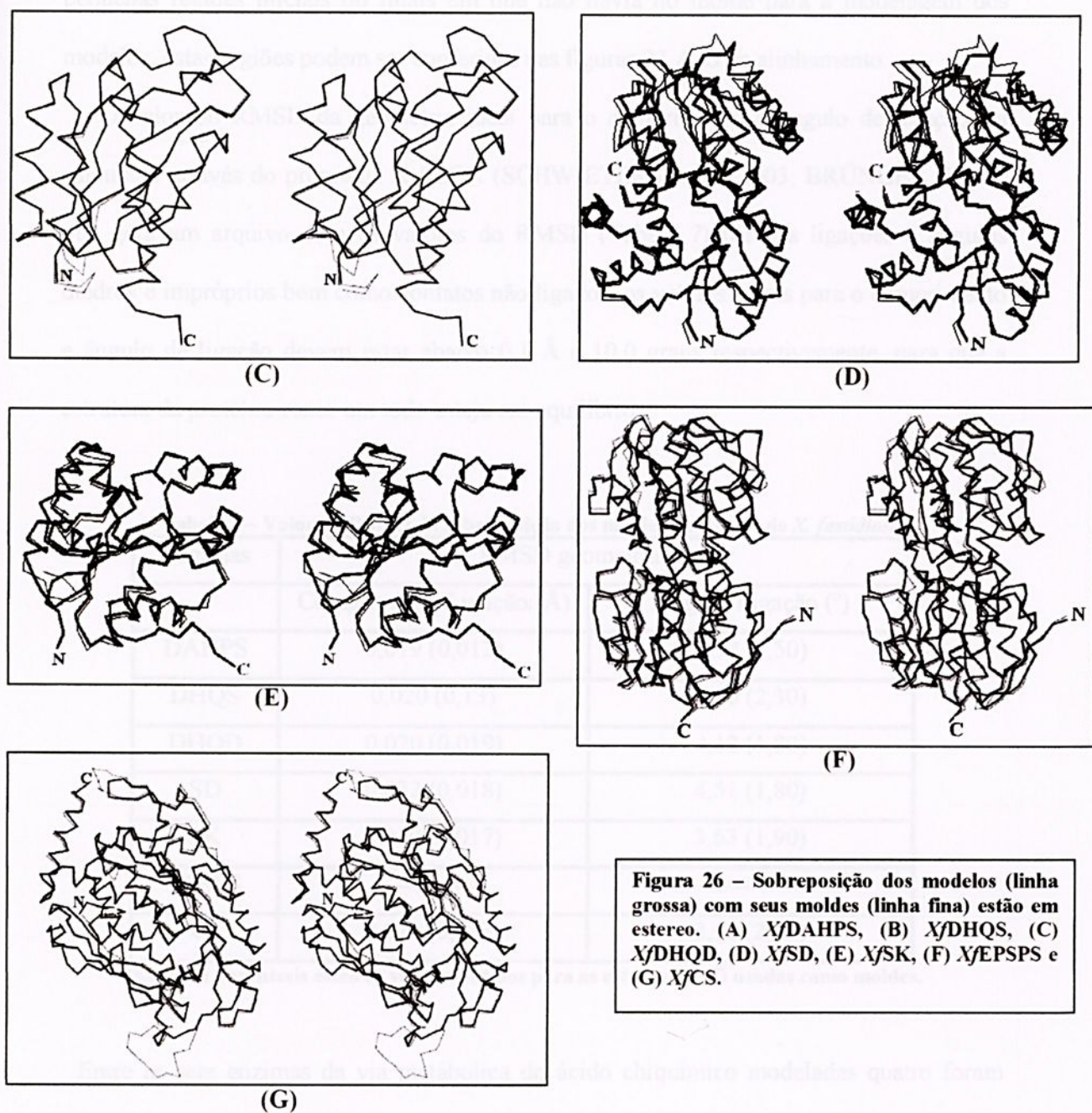


Figura 26 – Sobreposição dos modelos (linha grossa) com seus moldes (linha fina) estão em estereo. (A) X/DAHPS, (B) X/DHQS, (C) X/DHQP, (D) X/SD, (E) X/SK, (F) X/EPSPS e (G) X/CS.

Observando os valores da tabela 6 e 7 foi verificado que os modelos estão dentro do ideal (padrão) e por isso pode se afirmar que estão em equilíbrio. Através da figura 26 A-G pode se afirmar que as estruturas secundárias se conservam nos modelos com diferenças sutis que podem ser atribuídas a pequenos ou extensos *gaps* durante o alinhamento ou por

pequenas regiões iniciais ou finais em que não havia no molde para a modelagem dos modelos, estas regiões podem ser conferidas nas figuras 23 A-G de alinhamento.

O valor do RMSD da geometria ideal para o comprimento e ângulo de ligação foi calculado através do programa X-PLOR (SCHWIETERS *et al.*, 2003; BRÜNGER, 1992), que gera um arquivo com os valores do RMSD (Tabela 7) para as ligações e ângulos diedros e impróprios bem como contatos não ligados, os valores ideais para o comprimento e ângulo de ligação devem estar abaixo 0,1 Å e 10,0 graus respectivamente, para que a estrutura da proteína como um todo esteja em equilíbrio.

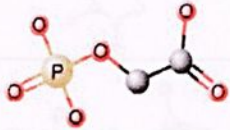
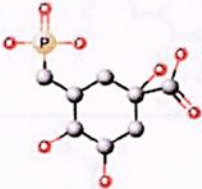
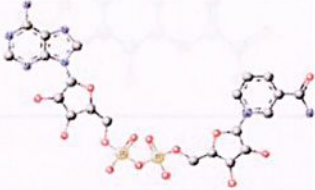
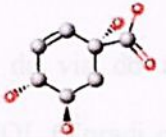
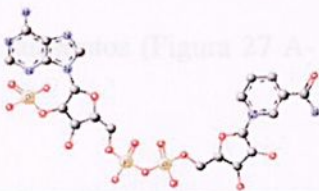
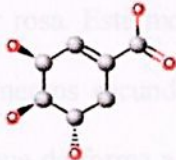
Tabela 7 – Valor do RMSD da sobreposição dos moldes e modelos de *X. fastidiosa*.

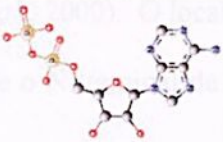
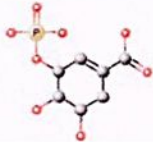
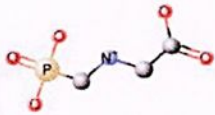
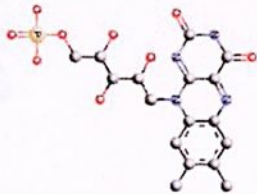
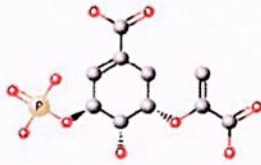
Enzimas	RMSD geometria ideal	
	Comprimento Ligação (Å)	Ângulo Ligação (°)
DAHPS	0,019 (0,012)	2,92 (1,50)
DHQS	0,020 (0,13)	3,70 (2,30)
DHQD	0,020 (0,019)	4,12 (1,80)
SD	0,032 (0,018)	4,51 (1,80)
SK	0,020 (0,017)	3,63 (1,90)
EPSPS	-----	-----
CS	0,022 (0,013)	3,92 (2,58)

*Entre os parêntesis estão os valores obtidos para as estruturas 3D usadas como moldes.

Entre as sete enzimas da via metabólica do ácido chiquímico modeladas quatro foram modeladas com possíveis inibidores, porque somente estes moldes tinham inibidores em sua estrutura cristalográfica (molde). Na tabela 8 foram colocadas as informações como nome, formula molecular, massa molecular e estrutura dos inibidores (escritos em azul), co-fatores, íons, substratos, que foram modelados em complexo com as enzimas.

Tabela 8 – Informações sobre inibidores, co-fatores, ions e substratos.

Enzima	Informações	Estrutura
DAHPS	Nome: ácido 2-fosfoglicolato (PGL) Formula Molecular: C ₂ H ₅ O ₆ P Massa Molecular: 156,0 Da	
	Nome: íon manganês (II) (MN) Formula Molecular: Mn ²⁺ Massa Molecular: 54,9 Da	
	Nome: [1R-(1α, 3β, 4α, 5β)]-5-(fosfometil)-1,3,4-trihidroxíciclohexano-1-ácido carboxílico (CRB) Formula Molecular: C ₈ H ₁₅ O ₈ P Massa Molecular: 270,2 Da	
DHQS	Nome: nicotinamida adenina dinucleotídeo (NAD ⁺) Formula Molecular: C ₂₁ H ₂₇ N ₇ O ₁₄ P ₂ Massa Molecular: 663,4 Da	
	Nome: ácido 2,3-anidro-quinico (FA1) Formula Molecular: C ₇ H ₁₀ O ₅ Massa Molecular: 174,1 Da	
SD	Nome: nicotinamida adenina dinucleotídeo Fosfato (NADP) Formula Molecular: C ₂₁ H ₂₈ N ₇ O ₁₇ P ₃ Massa Molecular: 743,4 Da	
SK	Nome: íon magnésio (MG) Formula Molecular: Mg ²⁺ Massa Molecular: 24,3 Da	
	Nome: ácido (3R,4S,5R)-3,4,5-trihidroxíciclohex-1-ene-1-carboxílico (SKM) Formula Molecular: C ₇ H ₁₀ O ₅ Massa Molecular: 174.153	

	Nome: adenosina 5' difosfato (ADP) Formula Molecular: C ₁₀ H ₁₅ N ₅ O ₁₀ P ₂ Massa Molecular: 427,2 Da	
EPSPS	Nome: chiquimato 3-fosfato (S3P) Formula Molecular: C ₇ H ₁₁ O ₈ P Massa Molecular: 254,1 Da	
	Nome: glifosato (GLP) Formula Molecular: C ₃ H ₉ N O ₅ P ⁺ Massa Molecular: 170,1 Da	
CS	Nome: flavina mononucleotídeo (FMN) Formula Molecular: C ₁₇ H ₂₁ N ₄ O ₉ P Massa Molecular: 456,3 Da	
	Nome: ácido 5-[(1-carboxivinil)oxi]-4-hidroxi-3-(fosfonooxi)ciclohex-1-ene-1-carboxílico (EPSP) Formula Molecular: C ₁₀ H ₁₃ O ₁₀ P Massa Molecular: 324,2 Da	

6.3. Descrição dos Modelos

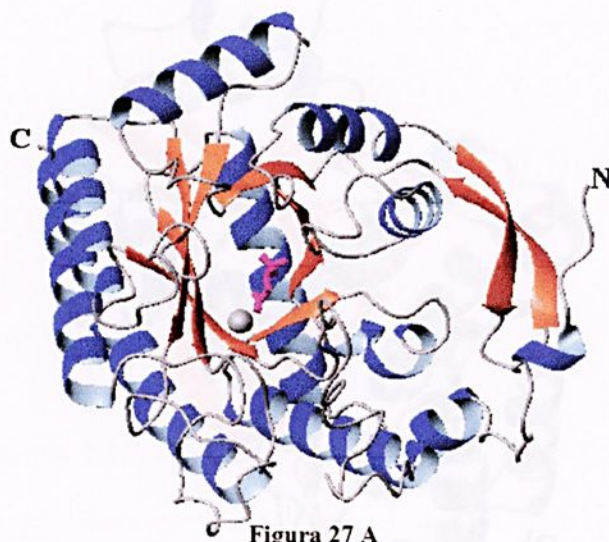
As figuras que mostram as estruturas secundárias das enzimas da via do ácido chiquímico de *X. fastidiosa* foram geradas através do programa MOLMOL (Koradi *et al.*, 1996) juntamente com seus respectivos inibidores, co-fatores ou substratos (Figura 27 A-G).

6.3.1. DAHPS

A figura 27 A mostra a estrutura secundária de *X*/DAHPS em complexo com o íon manganês (II) (Mn²⁺) na cor cinza e o 2-Fosfoglicolato (PGL) na cor rosa. Este modelo mostra que a proteína tem (α/β) barril 8 realçado por diversos elementos secundários adicionais da estrutura, incluindo uma cauda próxima ao N-terminal que dá forma a uma



folha- β anti-paralela seguida de uma fita- β ($\beta 6a$ - $\beta 6b$)(WAGNER *et al.*, 2000). O local do sítio ativo de XfDAHPS consiste de um canal entre o C-Terminal e o N-terminal de um barril com 8 (α/β) com Mn^{2+} e o PGL ligados dentro.



6.3.2. DHQS

A figura 27 B mostra a estrutura secundária de XfDHQS formando um complexo binário com [1R-(1 α , 3 β , 4 α , 5 β)-5-(fosfometil)-1,3,4-trihidrox ciclohexano-1-ácido carboxílico (CRB) um possível inibidor (em rosa) e Nicotinamida Adenina Dinucleotídeo (NAD^+) na cor vermelha.

A estrutura é composta de um N-terminal de domínio α/β e no C-terminal um domínio α -helicoidal (CARPENTER *et al.*, 1998). O domínio do N-terminal inclui uma dobra de Rossmann, mas liga surpreendentemente o NAD^+ em uma orientação invertida àquela observada em todas as outras proteínas clássicas conhecidas que sejam do tipo NAD^+ -ligado de Rossmann (CARPENTER *et al.*, 1998). Este domínio α/β consiste sete fitas folha- β na ordem 1296534. As fitas $\beta 1$ e $\beta 2$ formam um β -grampo, que está conectado a hélice, é conectado pela hélice $\alpha 1$, que leva até as cinco fitas da dobra Rossmann (LESK,

1995; ROSSSMANN *et al.*, 1974). As fitas compreendendo uma unidade do $\beta\alpha\beta$ e uma unidade do $\beta\alpha\beta\alpha\beta$. As fitas $\beta 7$ e $\beta 8$ formam β -grampo separado.

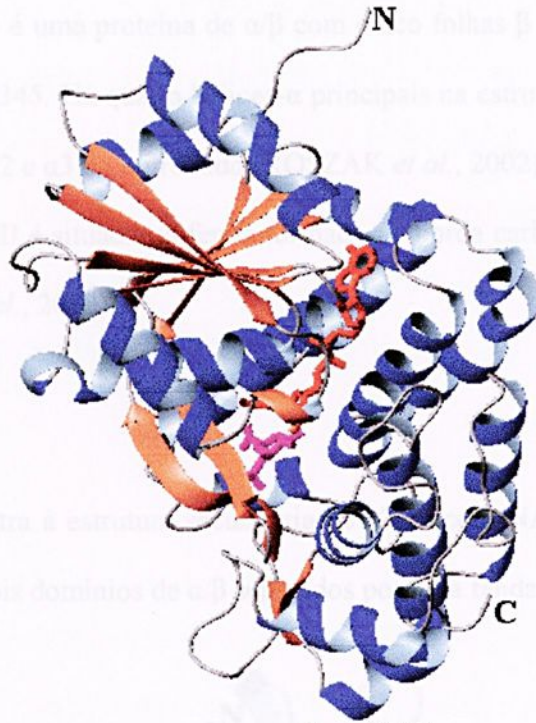


Figura 27 B

6.3.3. DHQD

A figura 27 C mostra a estrutura secundária de X/DHQD em complexo com ácido 2,3-anidro-quinico (FA1) em rosa.



Figura 27 C

A enzima *X/DHQD* é uma proteína de α/β com cinco folhas β em paralelo no centro da enzima na ordem 21345. Há quatro hélices- α principais na estrutura, com a $\alpha 1$ e $\alpha 4$ de um lado das folha- β e $\alpha 2$ e $\alpha 3$ do outro lado (ROSZAK *et al.*, 2002). O local do sítio ativo na enzima DHQD tipo II é situado na fenda formado na borda carboxi entre a folha- $\beta 1$ e folha- $\beta 3$ (ROSZAK *et al.*, 2002).

6.3.4. SD

A figura 27 D mostra a estrutura secundária de *X/SD* com NADP em vermelho e é uma arquitetura com dois domínios de α/β separados por uma fenda larga.

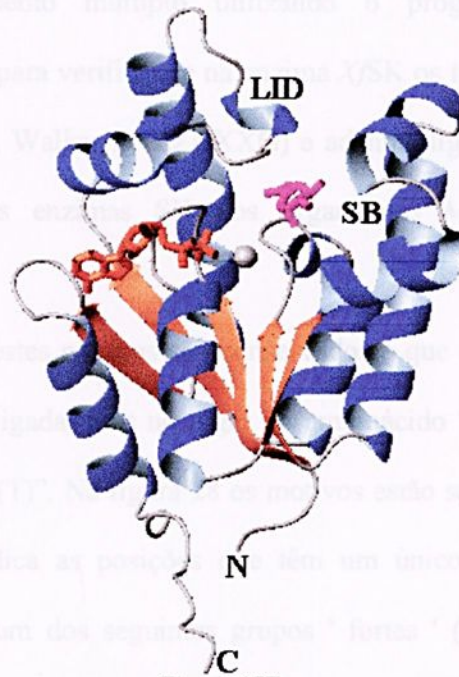


Figura 27 D

As primeiras três fitas- β seguem uma sucessão regular de α/β , com as hélices $\alpha 1$ e $\alpha 2$ paralelas as fitas- β . O domínio é completado por um grampo helicoidal- α no C-terminal ($\alpha 9$ e $\alpha 10$), que empacota o encontro da folha- β do mesmo lado como $\alpha 1$ (MICHEL *et al.*, 2003). Esta estrutura revela uma enzima com uma fenda profunda, que contem o local do sítio ativo na junção dos dois domínios. O domínio do C-terminal é facilmente reconhecido como um domínio obrigatório do dinucleotídeo da dobra de Rossmann, responsável por ligar o co-fator NADP. O domínio N-terminal onde se liga o substrato e a dimerização, um sanduíche do α - β - α , representa uma dobra topológica original (VOGAN, 2003).

6.3.5. SK

A figura 27 E mostra a estrutura secundária de XfSK em complexo com o substrato ácido chiquímico (SKM) em rosa, o co-fator ADP em vermelho e o íon Mg^{2+} em cinza.



A enzima *Xj*/SK é uma proteína α/β que consiste em um núcleo de cinco folhas- β cercada de hélices- α . As cinco folha- β centrais estão em paralelo na ordem 23145 estas folhas- β estão sendo flanqueadas por hélices- α ($\alpha 1$ e $\alpha 8$ em um lado, $\alpha 4$, $\alpha 5$ e $\alpha 7$ no outro) (SCHULZ, 1992; PEREIRA *et al.*, 2004a). Esta ordem estabelecida pelas folha- β 23145 classifica a enzima *Xj*/SK como sendo pertencente à mesma família estrutural quinases NMP.

As quinases fazem movimentos grandes durante a catálise para proteger o seu centro ativo da água a fim evitar a hidrólise do ATP (PEREIRA *et al.*, 2004b). Existe duas regiões flexíveis na estrutura que são responsáveis pelo movimento: chiquimato-ligado (SB) e o LID que estão indicados na figura 27E (PEREIRA *et al.*, 2004b). O domínio do SB é formado pela fita $\beta 2$, hélices $\alpha 1$ e $\alpha 2$ e a região do N-Terminal da hélice $\alpha 4$ (PEREIRA *et al.*, 2004a,b). O sítio de Walker A localizado entre a folha $\beta 1$ e $\alpha 1$ que formam um laço de fosfato (*P-loop*) (KRELL *et al.*, 1998).

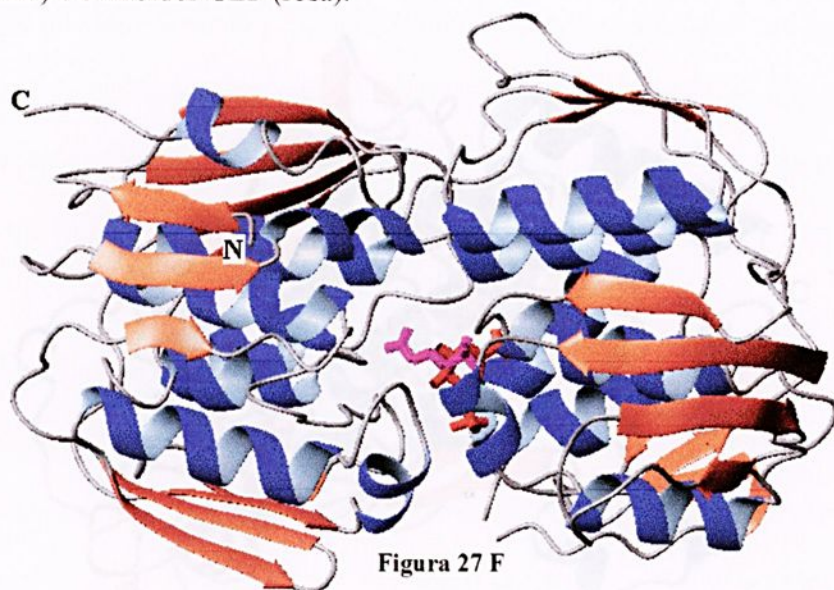
Foi feito um alinhamento múltiplo utilizando o programa CLUSTAL W (THOMPSON *et al.*, 1994), para verificar se na enzima *Xj*/SK os três motivos funcionais: Walker A (GXXXXGKT/S), Walker B (ZZDXXG) e adenina-ligada (I/VDXXX(X)XP) são conservados como nas enzimas SK dos organismos *M. tuberculosis* e *E. chrysanthemi*.

A figura 28 mostra que estes motivos são conservados e que a uma alteração de um resíduo no sítio da adenina-ligada, pois no lugar do aminoácido “Asp (D)” existe outro resíduo de aminoácido “Thr (T)”. Na figura 28 os motivos estão sombreados de cinza, os caracteres especiais: (*) indica as posições que têm um único, resíduo inteiramente conservado; (:) indica que um dos seguintes grupos ' fortes ' (STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW) está conservado inteiramente, (.) indica que um dos seguintes grupos ' mais fracos ' (CSA, ATV, SAG, STNK, STPA, SGND,



6.3.6. EPSPS

A figura 27 F mostra a estrutura secundária de *Xf*EPSPS em complexo com substrato S3P (vermelho) e o inibidor GLP (rosa).



Os dois domínios projetam uma forma em "V" e acomodam o N-terminal das seis hélices em cada domínio. Os efeitos macrodipolar helical criam um poço potencial que facilita a ligação do glifosato e do S3P. Os domínios são compostos de três cópias de $\beta\alpha\beta\alpha\beta\beta$ para cada unidade. As quatro fita folha- β contém fitas paralelas e antiparalelas e as hélices são paralelas a estas fitas (SCHÖNBRUNN *et al.*, 2001).

6.3.7. CS

A figura 27 G mostra a estrutura secundária de *Xf*CS em complexo com o substrato EPSP (vermelho) e co-fator FMN (rosa). A *Xf*CS foi predita inicialmente como um

proteína do tipo barril α/β , mas depois de pesquisas na verificação de sua estrutura secundária chegou-se a conclusão que sua estrutura secundária compreende em um único domínio grande, com nove hélices- α , dezoito fitas- β e cercada de alças e estiramentos discretos da estrutura secundária. O núcleo consiste nas duas hélices- α longas internas, comprimidas entre as quatro folhas- β que formam um sanduíche de enovelamento β - α - β , que estão dispostas em pares antiparalelos (MACLEAN & ALI *et al.*, 2003).

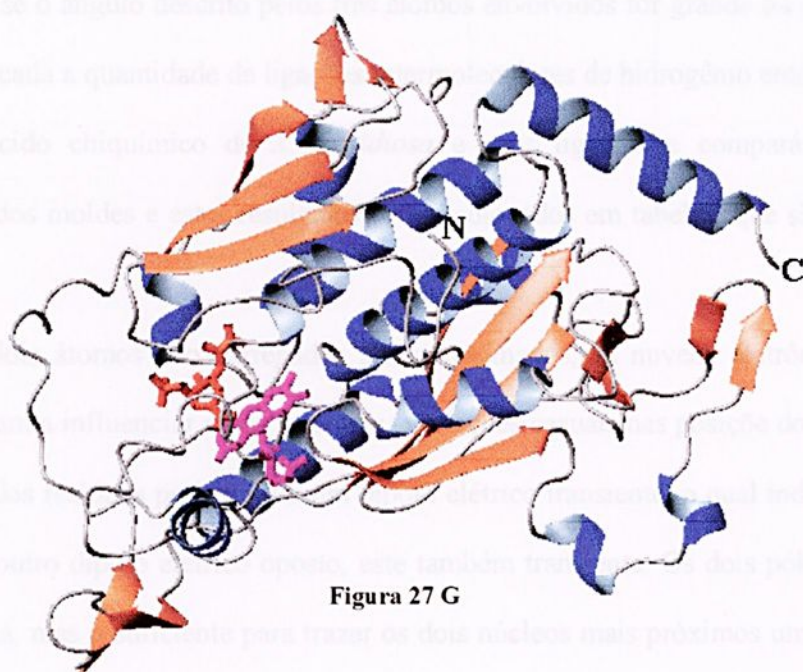


Figura 27 G

6.4. Interações com os Ligantes

As interações não covalentes como ligações intermoleculares de hidrogenio, interações iônicas (eletrostáticas) e interações hidrofóbicas e de van der Waals são mais fracas que as ligações covalentes. Embora estes tipos de interações sejam individualmente fracos quando comparados a ligações covalentes, em razão do grande número de tais interações existentes em uma proteína ou ácido nucléico, o efeito acumulativo se torna significativo, ajudando a estabilizar a estrutura e a definir os níveis de especificidade e afinidade entre proteína-

ligante. Dos quatro tipos de interações comentados acima, três tipos foram observados e avaliados neste trabalho: ligações intermoleculares de hidrogenio, interações iônicas (eletrostáticas) e interações de van der Waals.

As ligações intermolecular de hidrogênio (Tabelas 9 – 13) constituem uma das forças de estabilização mais importantes na estrutura de proteínas. As ligações de hidrogênio são incomuns entre as interações eletrostáticas porque são extremamente direcionais; elas se enfraquecem se o ângulo descrito pelos três átomos envolvidos for grande ou pequena. Por isso foi verificada a quantidade de ligações intermoleculares de hidrogênio entre as enzimas da via do ácido chiquímico de *X. fastidiosa* e seus ligantes e compará-los com as informações dos moldes e estes resultados foram colocados em tabelas, que são mostradas mais abaixo.

Quando dois átomos não carregados são aproximados, as nuvens eletrônicas que os rodeiam passam a influenciar umas as outras. Variações casuais nas posição dos elétrons ao redor de um dos resíduos podem criar um dipolo elétrico transiente, o qual induz no arodo seguinte um outro dipolo elétrico oposto, este também transiente. Os dois pólos atraem-se de forma fraca, mas o suficiente para trazer os dois núcleos mais próximos um do outro. A força desta atração fraca é chamada interação de van der Waals.

À medida que os dois núcleos se aproximam as duas nuvens eletrônicas começam a repelir-se e em algum ponto a atração de van der Waals contrabalança exatamente esta força repulsiva, nesta situação os núcleos não podem ser mais aproximados e são ditos estarem em contato de van der Waals. Foi verificada a quantidade de interações de van der Waals (Tabela 9) existente nas enzimas da via do ácido chiquímico de *X. fastidiosa*. As áreas de contato entre os inibidores e as enzimas foram calculadas através da rotina AEROMOL do pacote CCP4 (Collaborative Computational Project N° 4., 1994), este tipo de análise é



importante para saber o quanto o inibidor está tendo contato com a minha proteína, quanto maior for este valor mais específico é o inibidor para aquela proteína.

Tabela 9 – Interações de van der Waals, área de contato e ligações intermoleculares de hidrogênio para os modelos e moldes.

Enzimas	Quantidade de Interações de van der Waals	Quantidade de Ligações Intermoleculares de Hidrogênio	Área de Contato ($\text{\AA}^2$)
DAHPS	16	12 (12)	81 (96)
DHQS	26	16 (16)	566 (582)
DHQP	14	11 (11)	144 (105)
EPSPS	28	16 (15)	211 (186)

*Dentro dos parêntesis estão os valores obtidos para as estruturas 3D usadas como moldes.

Analisando a tabela acima foi possível verificar que a enzima DHQP de *X. fastidiosa* tem uma área de contato maior do que seu respectivo molde, porque no seu sítio de ligação existem resíduos que não se conservam facilitando assim a afinidade e especificidade entre a enzima e o inibidor. E na enzima EPSPS de *X. fastidiosa* foi verificado também que existem resíduos que não se conservam no sítio de ligação e outro fator importante é a ligação intermolecular de hidrogênio a mais que no molde, estes fatores fazem com que a afinidade e especificidade do inibidor se torne ainda maior para EPSPS de *X. fastidiosa*. Tabelas mais detalhadas de ligações intermoleculares de hidrogênio, são mostradas mais abaixo.

A análise das superfícies de potencial eletrostático foi feita com o auxílio do programa GRASP (NICHOLLS *et al.*, 1991), que é mostrada na figura 37 A-D. Tal análise revelou a presença de complementaridade de cargas e forma na região de contato entre as enzimas de *X. fastidiosa* e seus respectivos inibidores.



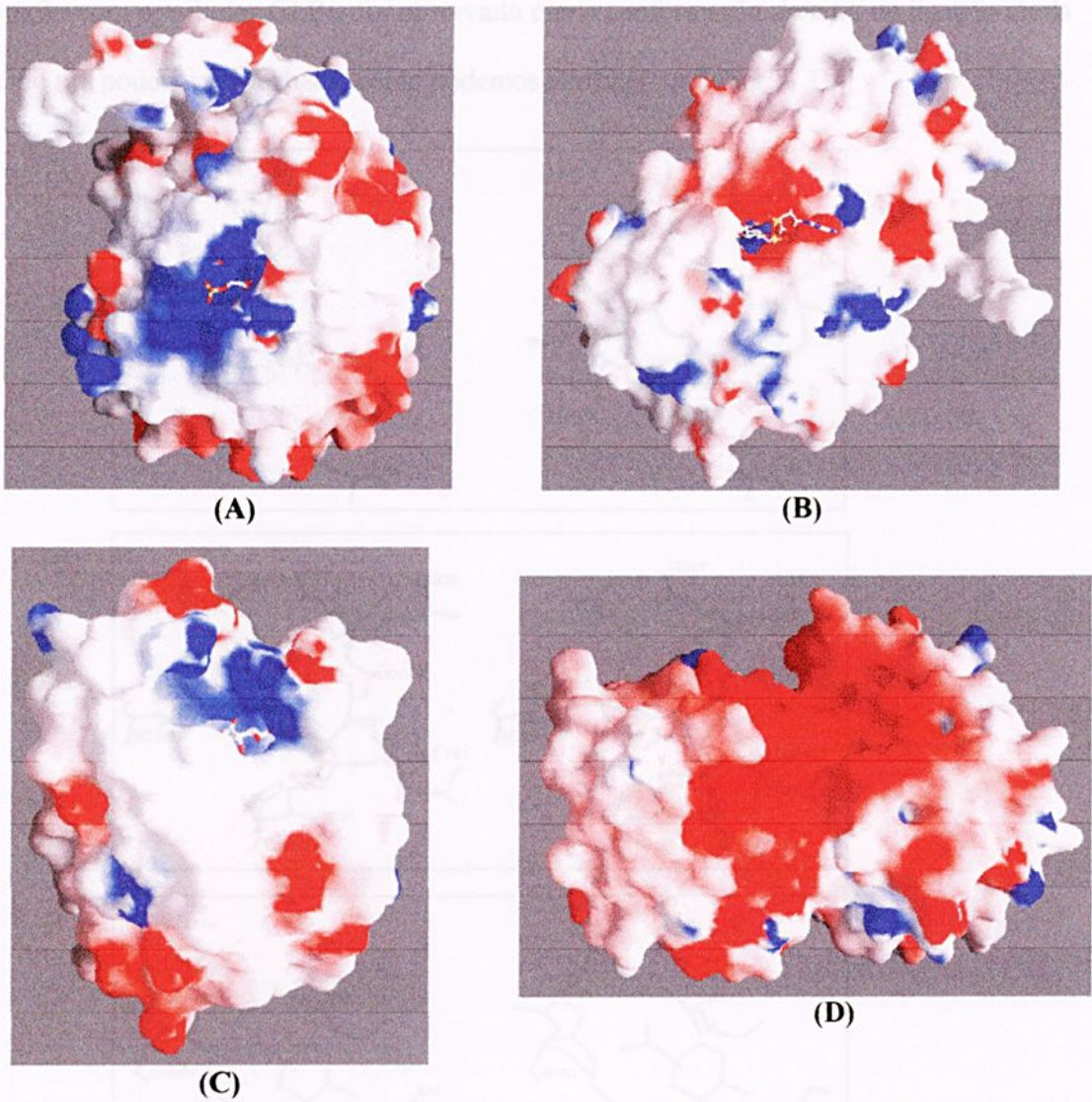


Figura 30 –Superfície de Potencial Eletrostática para os complexos (A) *Xf*DAHPS:PGL, (B) *Xf*DHQS:CRB, (C) *Xf*DHQD:FA1 e (D) *Xf*EPSPS:GLP.

Através da sobreposição dos complexos (Figura 31 A-D) que foram sobrepostos com auxílio do programa LSQKAB do pacote CCP4 (Collaborative Computational Project N° 4., 1994), foi observado alterações no posicionamento das cadeias laterais dos resíduos em relação ao posicionamento e interação com o inibidor na enzima *Xf*EPSPS por apresentar uma ligação intermolecular de hidrogênio a mais que o molde e alguns resíduos do sítio não são conservados (Tabela 9 e 13; Figura 31D). Com isso a enzima *Xf*EPSPS tem uma maior

Figura 31 – Sobreposição dos sítios de ligação dos complexos dos resíduos (linhas verdes) e moléculas (linhas brancas): (A) *Xf*DAHPS:PGL, (B) *Xf*DHQS:CRB, (C) *Xf*DHQD:FA1 e (D) *Xf*EPSPS:GLP.

interação com o inibidor GLP e foi observado que a conformação do sítio de ligação muda ficando um pouco mais fechada, como podemos verificar na figura 31 D.

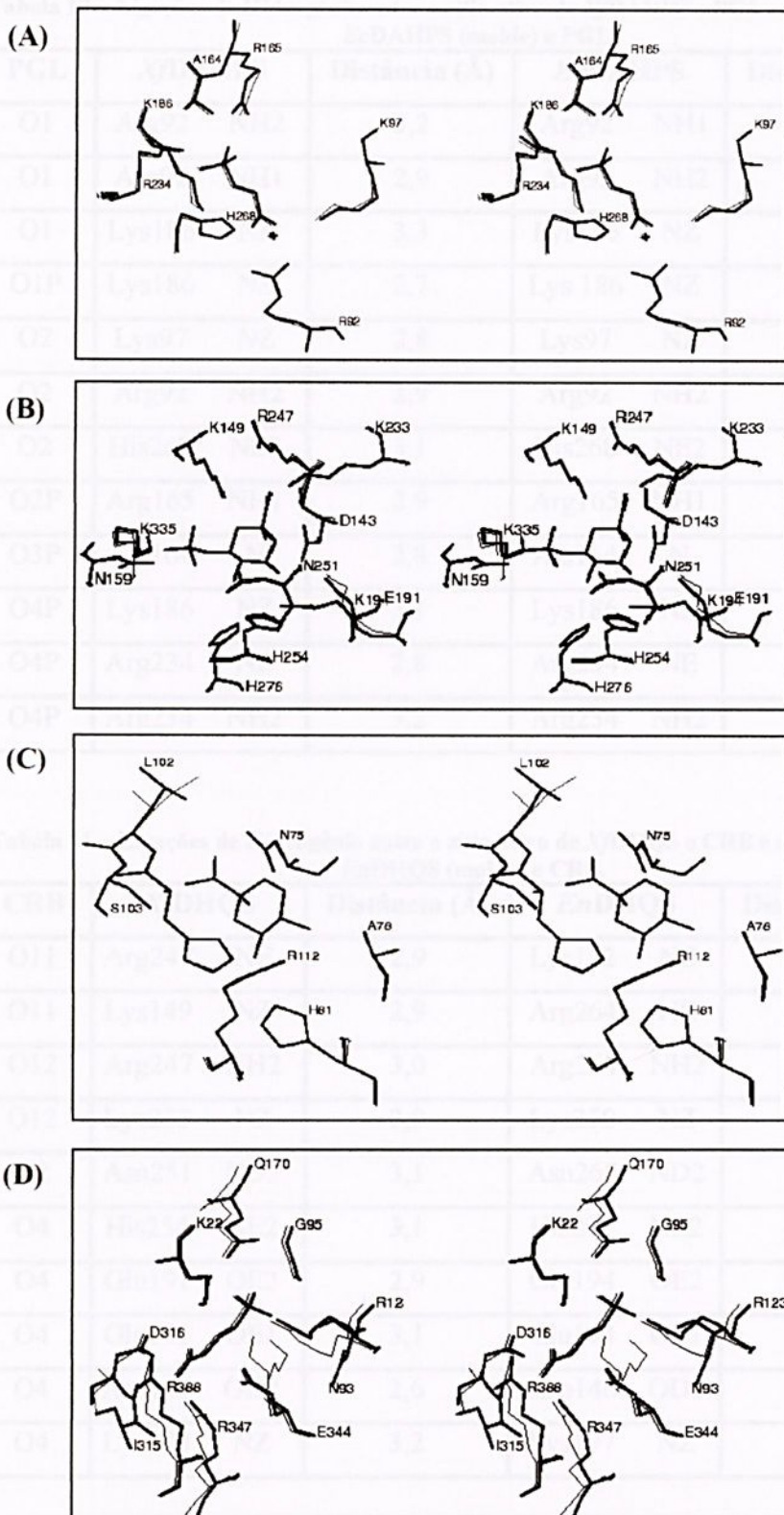


Figura 31 – Sobreposição dos sítios de ligação dos complexos dos modelos (linha grossa) e moldes (linha fina): (A) *Xy*DAHPS:PGL, (B) *Xy*DHQS:CRB, (C) *Xy*DHQD:FA1 e (D) *Xy*EPSPS:GLP

As tabelas 10 a 13 mostram as ligações intermoleculares de hidrogênio detalhadamente.

Tabela 10 – Ligações de Hidrogênio entre o sítio ativo de *Xf*DAHPS e PGL e entre o sítio ativo *Ec*DAHPS (molde) e PGL

PGL	<i>Xf</i> DAHPS	Distância (Å)	<i>Ec</i> DAHPS	Distância (Å)
O1	Arg92 NH2	3,2	Arg92 NH1	2,8
O1	Arg92 NH1	2,9	Arg92 NH2	3,2
O1	Lys186 NZ	3,3	Lys186 NZ	3,3
O1P	Lys186 NZ	2,7	Lys 186 NZ	2,9
O2	Lys97 NZ	2,8	Lys97 NZ	2,8
O2	Arg92 NH2	2,9	Arg92 NH2	3,0
O2	His268 NE2	3,1	His268 NE2	3,1
O2P	Arg165 NH1	2,9	Arg165 NH1	2,9
O3P	Ala164 N	2,8	Ala164 N	2,8
O4P	Lys186 NZ	3,1	Lys186 NZ	2,9
O4P	Arg234 NE	2,8	Arg234 NE	2,8
O4P	Arg234 NH2	3,2	Arg234 NH2	3,1

Tabela 11 – Ligações de Hidrogênio entre o sítio ativo de *Xf*DHQS e CRB e entre o sítio ativo *En*DHQS (molde) e CRB.

CRB	<i>Xf</i> DHQS	Distância (Å)	<i>En</i> DHQS	Distância (Å)
O11	Arg247 NE	2,9	Lys152 NZ	3,0
O11	Lys149 NZ	2,9	Arg264 NE	2,9
O12	Arg247 NH2	3,0	Arg264 NH2	2,9
O12	Lys233 NZ	2,9	Lys250 NZ	2,9
O2	Asn251 ND2	3,1	Asn268 ND2	3,1
O4	His254 NE2	3,1	His271 NE2	3,1
O4	Glu191 OE2	2,9	Glu194 OE2	2,9
O4	Glu191 OE1	3,1	Glu194 OE1	3,0
O4	Asp143 OD2	2,6	Asp146 OD2	2,5
O4	Lys194 NZ	3,2	Lys197 NZ	3,1



O5	His254	NE2	3,1	His287	NE2	3,1
O5	His276	NE2	3,1	His271	NE2	3,1
O91	Lys149	NZ	2,8	Lys152	NZ	2,8
O91	Lys335	NZ	2,9	Lys356	NZ	2,7
O92	Asn159	ND2	3,0	Asn162	ND2	3,0
O93	Asn251	ND2	2,8	Asn268	ND2	2,7

Tabela 12 – Ligações de Hidrogênio entre o sítio ativo de *Xf*DHQD e FA1 e entre o sítio ativo *Sc*DHQD (molde) e FA1.

FA1	<i>Xf</i> DHQD	Distância (Å)	<i>Sc</i> DHQD	Distância (Å)
O1	Leu102 N	3,2	Ile107 N	3,2
O1	Leu103 N	2,8	Ser108 N	2,8
O1	Ser103 OG	2,7	Ser108 OG	2,6
O2	Leu102 N	2,9	Ile107 N	2,9
O2	Asn75 ND2	2,9	Asn79 ND2	2,9
O3	Arg112 NH1	2,9	Arg117 NH1	2,9
O3	His81 NE2	2,8	His85 NE2	2,8
O3O	Asn75 OD1	2,9	Asn79 OD1	2,9
O4	Ala78 N	3,1	Ala82 N	3,1
O4	Arg112 NH1	3,0	Arg117 NH1	3,1
O1	Leu102 N	3,2	-----	-----
O3O	-----	-----	His106 ND1	2,8

Tabela 13 – Ligações de Hidrogênio entre o sítio ativo de *Xf*EPSPS e GLP e entre o sítio ativo *Ec*EPSPS (molde) e GLP.

GLP	<i>Xf</i> EPSPS	Distância (Å)	<i>Ec</i> EPSPS	Distância (Å)
O1	Lys22 NZ	2,9	Lys22 NZ	2,9
O2	Gln170 NE2	2,9	Gln171 NE2	2,9
O2	Arg123 NH1	2,8	Arg124 NH1	2,8
O3	Gly95 N	2,9	Gly96 N	2,8
O3	Arg123 NH2	2,8	Arg124 NH2	2,9
O3	Asn93 O	3,2	Lys411 NZ	2,8



O4	Lys22	NZ	2,9	Lys22	NZ	2,9
O4	Arg388	NH2	3,1	Arg386	NH1	3,1
O4	Arg388	NE	3,3	-----	-----	-----
O5	Arg388	NE	2,8	Arg386	NE	2,6
O5	Arg347	NH2	3,0	Arg344	NH1	3,0
O5	Ile315	O	2,6	Arg344	NH2	2,8
O5	Arg347	NH1	3,0	Asp313	OD2	3,2
O5	Asp316	OD2	3,2	Asp313	OD1	3,2
O5	Asp316	OD1	3,2	-----	-----	-----
N1	Glu344	OE2	2,9	Glu341	OE2	2,9
N1	-----	-----	-----	Glu341	OE1	3,3



7. CONCLUSÃO

A enzima *Xf*EPSPS apresenta uma ligação de hidrogênio a mais e um contato de área de superfície maior que o dos seu molde. Já a enzima *Xf*DHQP apesar de ter o mesmo número de ligações de hidrogênio, apresenta uma área de contato de superfície maior que o seu molde. Desta forma podemos afirmar que estes modelos estruturais estão fortemente ligados às enzimas. Todos os complexos analisados mostraram uma boa forma e complementaridade de carga e também podem ser possíveis inibidores contra *X. fastidiosa*. Estes estudos podem ser confirmados futuramente através de métodos experimentais.

8. REFERÊNCIAS

- Adobe Systems Inc.; **PostScript language Reference Manual**. Reading, MA: ADDISON-WESLEY.
- ALLEN, F. H.; BELLARD, S.; BRICE, M. D.; *et al.* The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. **Acta Cryst.** v.B35, p.2331-2339, 1979.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; *et al.* Basic local alignment search tool. **J. Mol. Biol.** v.215, p.403-410, 1990.
- ALTSCHUL S. F.; BOGUSKI M. S.; GISH W.; *et al.* Issues in searching molecular sequence databases. **Nature Genet.** v.6, p.119-129, 1994.
- ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFER, A. A.; *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. **Nucleic Acids Res.** v.25, p.3389-3402, 1997.
- ALVES, E. ***Xylella fastidiosa* – Adesão e colonização em vasos do xilema de laranjeira doce, cafeeiro, ameixeira, fumo e espécies de cigarrinhas vetoras e formação de biofilme sobre película de poliestireno.** Piracicaba, 2003. 139p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade São Paulo.
- ANDERSON, K. S.; SIKORSKI, J. A.; JOHNSON, K. A.; Evaluation of 5-enolpyruvylshikimate-3-phosphate synthase substrate and inhibitor binding by stopped-flow and equilibrium fluorescence measurements. **Biochemistry** v.27, p.1604-1610, 1988.
- APOSTOLICO, A.; GIANCARLO, R.; Sequence alignment in molecular biology. **J Comput Biol** v.5, p.173-196, 1998.
- ARCURI, H. A.; CANDURI, F.; PEREIRA, J. H.; *et al.* Molecular Models for Shikimate Pathway Enzymes of *Xylella fastidiosa*. **Biochem. Biophys. Res. Commun.** v.320, n.3, p.979-991, 2004.
- BACON, D. J.; ANDERSON, W. F.; A fast algorithm for rendering space-filling molecule pictures. **J. Mol. Graphics** v.6, p.219-220, 1988.
- Bairoch, A.; Boeckmann, B.; Ferro, S.; *et al.*; Swiss-Prot: Juggling between evolution and stability. **Brief. Bioinform.** v.5, p.39-55, 2004.
- BAKER, D; SALI, A; Protein structure prediction and structural genomics, **Science** v.294, p.93-96, 2001.
- BALASUBRAMANIAN, S.; ABELL, C.; COGGINS, J. R; Observation of an isotope effect in the chorismate synthase reaction. **J. Am. Chem. Soc.** v.112, p.8581-8583, 1990.



- BARTON G. J.; & STERNBERG M. J.; A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. **J Mol Biol** v.198, p.327-337, 1987.
- BEMSTEIN, F. C.; KOETZLE, T. F.; WILLIAMS, G. J. B.; *et al.* The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures. **J Mol Biol** v.112, p.535-542, 1977.
- BENACH, J.; LEE, I.; EDSTROM, W.; *et al.* The 2.3Å crystal of the shikimate 5-Deydrogenase orthologue YdiB from *E.coli* suggests a novel catalytic enviroment for an NAD-dependent dehydrogenase. **J. Biol. Chem** v.278, p.19176-19182, 2003.
- BENTLEY, R.; The shikimate pathway - A metabolic tree with many branches. **Critical Reviews in Biochemistry and Molecular Biology** v.25, p.307-384, 1990.
- BERENDSEN, H. J. C.; VAN DER SPOEL, D.; VAN DRUNEN, R.; GROMACS: a message-passing parallel molecular dynamics implementation. **Comp. Phys. Commun.** v.91, p.43-56, 1995.
- BERETTA, M. J. G.; HAKAWA, R.; CHAGAS, C. M.; *et al.* First report of *Xylella fastidiosa* in coffee. **Plant Dis.** v.80, p.821, 1996.
- BERMAN, H. M.; WESTBROOK, J; FENG, Z.; *et al.* The Protein Data Bank. **Nucleic. Acids Res.** v.28, p.235-242, 2000.
- BERNSTEIN, F. C.; KOETZLE, T. F.; WILLIAMS, G. J. B.; *et al.* The protein data bank: a computer-based archival file for macromolecular structures. **J. Mol. Biol.** v.112, p.535-542, 1977.
- BERNSTEIN, F. C.; POSTMA, J. P. M.; VAN GUNSTEREN, W. F.; *et al.*, Molecular dynamics with coupling to an external bath. **J. Chem. Phys.** v.81, 3684-3690, 1984.
- BEVAN, B.; The Bugs from Brazil. **Nature** v.406, p.140-141, 2000.
- BLUNDELL, T.L.; SIBANDA, B. L.; STERNBERG, M. J.; *et al.* Knowledge-based prediction of protein structures and the design of novel molecules. **Nature** v.326, p.347-352, 1987.
- BOISSEL J. P.; LEE W. R.; PRESNELL S. R.; *et al.* Erythropoietin structure-function relationships. Mutant proteins that test a model of tertiary structure. **J Biol Chem** v.268, p.15983-15993, 1993.
- BONNEAU, R.; BAKER, D.; Ab initio protein structure prediction: progress and prospects. **Annu Rev Biophys Biomol Struct** v.30, p.173-189, 2001.
- BOOCOCK, M. R.; COGGINS, J. R.; Kinetics of 5-enolpyruvylshikimate-3-phosphate synthase inhibition by glyphosate. **FEBS Letters** v.154, p.127-133, 1983.



- BOWIE, J. U.; LUTHY, R.; EISENBERG, D.; A method to identify protein sequences that fold into a known three- dimensional structure. **Science** v.253, p.164-170, 1991.
- BRAUN, W.; GO, N.; Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. **J Mol Biol** v.186, p.611-626, 1985.
- BRLANSKY, R. H.; TIMMER, L. W.; FRENCH, W. J.; *et al.* Colonization of the sharpshooter vectors, *Oncometopia nigricans* and *Homalodisca coagulata*, by xylem-limited bacteria. **Phytopathology** v.73, p.530-535, 1983.
- BROOKS III, C. L.; BRUCCOLERI R. E.; OLAFSON, B. D.; *et al.* CHARMM: A program for macromolecular energy minimization and dynamics calculations. **J Comp Chem** v.4, p.187-217, 1983.
- BROWNE, W. J.; NORTH, A. C. T.; PHILLIPS, D. C.; *et al.* A possible three-dimensional structure of bovine lactalbumin based on that of hen's egg-white lysozyme. **J Mol Biol** v.42, p.65-86, 1969.
- BRUCCOLERI, R. E.; KARPLUS, M.; Conformational Sampling Using High-Temperature Molecular Dynamics. **Biopolymers** v.29, p.1847-1862, 1990.
- BRÜNGER, A. T.; X-PLOR Version 3.1: A System for Crystallography and NMR, Yale University Press, New Haven, 1992.
- CAMPBELL, M. M.; SAINSBURY, M.; SEARLE, P. A.; The biosynthesis and synthesis of shikimic acid, chorismic acid, and related compounds. **Synthesis-Stuttgart** v.2, p.179-193, 1993.
- CAMPBELL, S. A.; RICHARDS, T. A.; MUI, E. J.; *et al.*, A complete shikimate pathway in *Toxoplasma gondii* an accident eukaryotic innovation. **International Journal for Parasitology** v.34, p.5-13, 2004.
- CARPENTER, E. P.; HAWKINS, A. R.; FROST, J. W.; *et al.* Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. **Nature** v.394, n.6690, p.299-302, 1998.
- CHANG, C. J.; GARNIER, M.; ZREIK, L.; *et al.* Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. **Curr. Microbiol.** v.27, p.137-142, 1993.
- CHOTHIA, C.; LESK, A. M.; The relation between the divergence of sequence and structure in proteins. **EMBO J** v.5, 823-826, 1986.
- CLAESSENS, M.; VAN CUTSEM, E.; LASTERS, I.; *et al.* Modelling the polypeptide backbone with 'spare parts' from known protein structures. **Protein Eng** v.2, p.335-345, 1989.
- COGGINS, J. R.; ABELL, C.; EVANS, L. B.; *et al.* Experiences with the shikimate-pathway enzymes as targets for rational drug design. **Bioch. Soc. Trans.** v.31, p.548-552, 2003.



- Collaborative Computational Project N° 4., The CCP4 suite: programs for proteins crystallography. **Acta Crystallogr.** v.D50, p.760-763, 1994.
- CORPET, F.; Multiple sequence alignment with hierarchical clustering. **Nucl. Acids Res.** v.16, n.22, p.10881-10890, 1988.
- DA SILVA, F. R.; VETTORE, A. L.; KEMPER, A. L.; Fastidian gum: the *Xylella fastidiosa* exopolysaccharide possibly involved in bacterial pathogenicity. **FEMS Microbiol. Lett.** v.203, p.165-171, 2001.
- DA SILVEIRA, N. J.; UCHOA, H. B.; CANDURI, F. ; *et al.* Structural bioinformatics study of PNP from *Schistosoma mansoni*. **Biochem. Biophys. Res. Commun.** v.322, p.100-104, 2004.
- DAVIS, B. D.; Aromatic biosynthesis. I. The role of shikimic acid. **Journal of Biological Chemistry** v.191, p.315-325, 1950.
- DAVIES, G. M.; BARRETT-BEE, K. J.; JUDE, D. A.; *et al.* (6S)-6-fluoroshikimic acid, an antibacterial agent acting on the aromatic biosynthetic pathway. **Antimicrob. Agents Chemother.** V.38, p.403-406, 1994.
- DAYHOFF, M. O; SCWARTZ, R. M.; ORCUTT, B. C.; A model of evolutionary change in proteins. in "Atlas of protein sequence and structure", vol.5, suppl. 3 pp.345-352. Natl. Biomed. Res. Found., Washington, DC, 1978.
- DE FEYTER, R. C.; PITTARD, J.; Purification and properties of shikimate kinase II from *Escherichia coli* K-12. **J. Bacteriol.** v.165, p.331-333, 1986.
- DE NEGRI, J. D.; GARCIA JUNIOR, A.; Sugestões para o manejo de pomares com clorose variegada dos citros. **Laranja** v.14, n.1, p.255-268, 1994.
- DE OLIVEIRA, J. S. A chiquimato quinase (E. C. 2.7.1.71) e a 5-enolpiruvilchiquimato-3-fosfato sintase (E. C. 2.5.1.19) de *Mycobacterium tuberculosis*: alvos potenciais para o desenvolvimento de novas drogas anti-tuberculose., em 07/05/2001.
- DOMINGUES, F.S.; LACKNER, P.; ANDREEVA, A.; *et al.* Structure-based evaluation of sequence comparison and fold recognition alignment accuracy. **J Mol Biol** v.297, p.1003-1013, 2000.
- DOSSELAERE, F.; VANDERLEYDEN, J.; A Metabolic Node in Action: Chorismate-Utilizing Enzymes in Microorganisms. **Crit. Rev. Microb.** v.27, n.2, p.75-131, 2001.
- DUNBRACK JR., R. L.; KARPLUS, M.; Backbone-Dependent Rotamer Library for Proteins. Application to Side-Chain Prediction. **J. Mol.Biol.** v.230, p.543-574, 1993.
- DUNCAN, K.; LEWENDON, A.; COGGINS, J. R.; The complete amino acid sequence of *Escherichia coli* 5-enolpyruvylshikimate 3-phosphate synthase. **FEBS Lett.** v.170, p.59-63, 1984.



- EISENBERG, D.; WESSON, M.; YAMASHITA, M.; Interpretation of protein folding and binding with atomic solvation parameters. **Chem Scr** v.29A, p.217-221, 1989.
- ENGH, R. A.; HUBER, R.; Accurate bond and angle parameters for X-ray protein structure refinement. **Acta Cryst.** v.A47, p.392-400, 1991.
- ESCHENBURG, S; HEALY, M. L.; PRIESTMAN, M. A.; *et al.* How the mutation glycine96 to alanine confers glyphosate insensitivity to 5-enolpyruvyl shikimate-3-phosphate synthase from *Escherichia coli*. **Planta** v.216, p.129-135, 2002.
- ESWAR, N.; JOHN, B.; MIRKOVIC, N.; *et al.* Tools for comparative protein structure modeling and analysis. **Nucleic Acids Res.** v.31, p.3375-380, 2003.
- FILHO, O. A. S.; ALENCASTRO, R. B.; Modelagem de Proteínas por Homologia. **Quim. Nova** v.26, n.2, p.253-259, 2003.
- FISER, A.; DO, R. K.; SALI, A.; Modeling of loops in protein structures, **Protein Sci.** v.9, p.1753-1773, 2000.
- FISER, A.; FEIG, M.; BROOKS III, C. L.; *et al.* Evolution and Physics in Comparative Protein Structure Modeling. **Acc Chem Res** v.35, p.413-421, 2002.
- GANEM, B.; From glucose to aromats: recent developments in natural products of the shikimic acid pathway. **Tetrahedron** v.34, p.3353-3383, 1978.
- GASPAR, R. T.; **Modelagem Molecular da Quinase dependente de ciclina 5 complexada com roscovitina**. São José do Rio Preto, 2003. 67p. Tese (Mestrado) – Departamento de Física do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista Julio de Mesquita Filho (UNESP).
- GERSTEIN, M.; LEVITT, M.; Using iterative dynamic programming to obtain accurate pairwise and multiple alignments of protein structures. **Proc. Int. Conf. Intell. Syst. Mol. Biol.** v.4, p.59-67, 1996.
- GIBAS, C.; JAMBECK, P.; **Desenvolvendo Bioinformática**, Rio de Janeiro, Ed. Campus, p.179-180, 2001.
- GIBSON, F.; The elusive branch-point compound of aromatic amino acid biosynthesis. **Trends in Biochemical Sciences** v.24, p.36-38, 1999.
- GILES, N. H.; CASE, M. E.; BAUM, J.; *et al.* Gene organization and regulation in the qa (Quinic acid) gene cluster of *Neurospora crassa*. **Microbiol. Rev.** v.49, p.338-358, 1985.
- GODZIK, A.; Knowledge-based potentials for protein folding: what can we learn from known protein structures? **Structure** v.4, p.363-366, 1996.
- GOLDSBROUGH, P. B.; HATCH, E. M.; HUANG, B.; *et al.* Gene amplification in glyphosate tolerant tobacco cells. **Plant Sci.** v.72, p.53-62, 1990.



T. 1719/08



- GOURLEY, D. G.; COGGINS, J. R.; ISAACS, N. W.; Crystallization of a type II dehydroquinase from *Mycobacterium tuberculosis*. **Journal of Molecular Biology** v.241, p.488-491, 1994.
- GREER, J.; Comparative model-building of the mammalian serine proteases. **J Mol Biol** v.153, p.1027-1042, 1981.
- GREER, J.; Comparative modeling methods: application to the family of the mammalian serine proteases. **Proteins** v.7, p.317-334, 1990.
- GRIBSKOV, M.; VERETNIK, S.; Identification of sequence pattern with profile analysis. **Methods Enzymol** v.266, p.198-212, 1996.
- GU, Y.; RESHETNIKOVA, L.; LI, Y.; *et al.* Cristal Structure of Shikimate Kinase from *Micobacterium tuberculosis* Reveals the Dynamic Role of the LID Domain in Catalisis. **J. Mol. Biol.** v.319, p.779-789, 2002.
- GUENTHER, B.; ONRUST, R.; SALI, A.; *et al.* Crystal structure of the delta' subunit of the clamp-loader complex of E. coli DNA polymerase III. **Cell** v.91, p.335-345, 1997.
- HASAN, N.; NESTER, E. W.; Dehydroquinase synthase in *Bacillus subtilis*, an enzyme associated with chorismate synthase and flavin reductase. **J. Biol. Chem.** v.253, p.4999-5004, 1978a.
- HASAN N.; NESTER E. W.; Purification and properties of chorismate synthase from *Bacillus subtilis*. **J. Biol. Chem.** v.253, p.4993-4998, 1978b.
- HAVEL, T. F.; SNOW, M. E.; A New Method for Building Protein Conformations From Sequence Alignments With Homologues of Known Structure. **J. Mol. Biol.** v.217, p.1-7, 1991.
- HAWKES, T. R.; LEWIS, T.; COGGLNS, J. R.; *et al.* Chorismate synthase, pre-steady-state kinetics of phosphate release from 5-enolpyruvylshikimate 3-phosphate. **Biochem. J.** v.265, p.899-902, 1990.
- HENDSON M.; PURCELL A. H.; CHEN D.; *et al.* Genetic Diversity of Pierce's Disease Strains and Other Pathotypes of *Xylella fastidiosa*. **Applied and Environmental Microbiology** v.67, p.895-903, 2001.
- HENIKOFF, S.; HENIKOFF, J.G.; Amino acid substitution matrices from proteins blocks. **PNAS** v.89, p.10915-10919, 1992.
- HENIKOFF, J. G.; PIETROKOVSKI, S.; MCCALLUM, C. M.; *et al.* Blocks-based methods for detecting protein homology. **Electrophoresis** v.21, p.1700-1706, 2000.
- HENIKOFF, S.; HENIKOFF, J.G.; Amino acid substitution matrices. **Adv Protein Chem** v.54, p.73-97, 2000.



- HERRMANN, K. M.; The shikimate pathway: early steps in the biosynthesis of aromatic compounds. **The Plant Cell** v.7, p.907-919, 1995a.
- HERRMANN, K. M.; The shikimate Pathway as an Entry to Aromatic Secondary Metabolism. **Plant Physiol.** v.107, p.7-12, 1995b.
- HERRMANN, K. M.; WEAVER, L. M.; The Shikimate Pathway. **Annu. Rev. Plant Mol. Biol.** v.50, n.3, p.473-503, 1999.
- HIGGINS, D. G.; BLEASBY, A. J.; FUCHS, R.; CLUSTAL V: improved software for multiple sequence alignment. **Comput Appl Biosci** v.8, p.189-191, 1992.
- HILL, B. L.; PURCELL, A. H.; Multiplication and movement of *Xylella fastidiosa* within grape and four other plants. **Phytopathology** v.85, n.11, p.1368-1372, 1995.
- HOOFT, R. W.; VRIEND, G.; SANDER, C.; *et al.* Errors in protein structures. **Nature** v.381, p.272, 1996.
- HOPKINS, D. L.; *Xylella fastidiosa*: Xylem-limited bacterial pathogen of plants. **Annual Review of Phytopathology** v.27, p.271-290, 1989.
- HOWELL, P. L.; ALMO, S. C.; PARSONS, M. R.; *et al.* Structure determination of turkey egg-white lysozyme using Laue diffraction data. **Acta Crystallogr** v.B48, n.2, p.200-207, 1992.
- JACOBSON, M.; SALI, A.; Comparative Protein Structure Modeling and its Applications to Drug Discovery. **Annual Reports In Medicinal Chemistry** v.39, p.259-276, 2004.
- JAROSZEWSKI, L.; RYCHLEWSKI, L.; ZHANG, B.; *et al.* Fold prediction by a hierarchy of sequence, threading, and modeling methods. **Protein Sci** v.7, p.1431-1440, 1998.
- JAROSZEWSKI, L.; RYCHLEWSKI, L.; GODZIK, A.; Improving the quality of twilight-zone alignments. **Protein Sci** v.9, p.1487-1496, 2000.
- JOHNSON, M. S.; OVERINGTON, J. P.; A structural basis for sequence comparisons. An evaluation of scoring methodologies. **J Mol Biol** v.233, p.716-738, 1993.
- JOHNSON, M. S.; SRINIVASAN, N.; SOWDHAMINI, R.; *et al.* Knowledge-based protein modelling. **CRC Crit Rev Biochem Mol Biol** v.219, p.1-68, 1994.
- JONES, T. A.; THIRUP, S.; Using Known Substructures in Protein Model Building and Crystallography. **EMBO J.** v.5, p.819-822, 1986.
- JONES, D. T.; TAYLOR, W. R.; THORNTON, J. M.; A new approach to protein fold recognition. **Nature** v.358, p.86-89, 1992.
- KABSCH, W.; A solution for the best rotation to relate two sets of vectors. **Acta Crystallogr** v.A32, p.922-923, 1976.



- KABSCH, W.; A discussion of the solution for the best rotation to relate two sets of vectors. **Acta Crystallogr** v.A34, p.827-828, 1978.
- KABSCH, W.; SANDER, C.; Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers** v.22, n.12, p.2577-2637, 1983.
- KEARSLEY, S. K.; On the orthogonal transformation used for structural comparisons. **Acta Crystallogr.** v.45A, p.208-210, 1989.
- KENN, N. T.; DUMENYO, C. K.; YANG, C.; *et al.* From rags to riches, insights from the first genomic sequence of plant pathogenic bacterium. **Genome Biology** v.1, n.3, p.1019.1-1019.4, 2000.
- KISHORE, G. M.; SHAH, D. M.; Amino acid biosynthesis inhibitors as herbicides. **Annu Rev Biochem** v.57, p.627-663, 1988.
- KNAGGS, A. R.; The biosynthesis of shikimate metabolites. **Nat. Prod. Rep.** v.20, p.119-136, 2003.
- KOLINSKI, A; BETANCOURT, M. R.; KIHARA, D.; *et al.* Generalized Comparative Modeling (GENECOMP): A Combination of Sequence Comparison, Threading, and Lattice Modeling for Protein Structure Prediction and Refinement. **Proteins** v.44, p.133-149, 2001.
- KORADI, R.; BILLETER, M.; WUTHRICH, K.; MOLMOL: a program for display and analysis of macromolecular structures. **J. Mol. Graphics** v.14: 51-55, 1996.
- KRAULIS, P. J.; MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. **J. Appl. Crystallogr.** v.24, p.946-950, 1991.
- KRELL, T.; COGGINS, J. R.; LAPTHORN, A J.; The three-dimensional structure of shikimate kinase. **Journal of Molecular Biology** v.278, p.983-997, 1998.
- KRELL, T.; MACLEAN, J.; BOAM, D. J.; *et al.* Biochemical and X-ray crystallographic studies on shikimate kinase: The important structural role of the P-loop lysine. **Protein Sci.** v.10, p.1137-1149, 2001.
- KROGH, A.; BROWN, M.; MIAN, I. S.; *et al.* Hidden Markov models in computational biology. Applications to protein modeling. **J Mol Biol** v.235, p.1501-1531, 1994.
- LAMBAIS, M. R.; GOLDMAN, M. H. S.; CAMARGO, L. E. A.; *et al.* A genomic approach to the understanding of *Xylella fastidiosa* pathogenicity. **Current Opinion in Microbiology** v.3, n.5, p.459-462, 2000.
- LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; *et al.*, PROCHECK - A program to check the stereochemical quality of protein structures, **J. Appl. Cryst.** v.26, p.283-291, 1993.
- LARANJEIRA, F. F. Ten years of citrus variegated chlorosis: What do we know? **Laranja** v.18, p.123-1141, 1997.



- LARANJEIRA, F. F.; BERGAMIN FILHO, A.; AMORIN, L.; *et al.* Aspectos práticos da epidemiologia da clorose variegada dos citros. **Revista Laranja** v.19, n.1, p.79-90, 1998.
- LEE, R. F.; BERETTA, M. J. G.; HARTUNG, J. H.; *et al.* Citrus variegated chlorosis: confirmation of a *Xylella fastidiosa* as the causal agent. **Summa Phytopathol.** v.19, p.123-125, 1993.
- LEITE, B.; ISHIDA, M. L.; ALVES, E.; *et al.* Genomics and X-ray microanalysis indicate that Ca^{2+} and thiols mediate the aggregation and adhesion of *Xylella fastidiosa*. **Braz. J. Med. Biol. Res.** v.35, n.6, p.645-650, 2002.
- LESK, A. M.; Introduction to Protein Architecture. Oxford University Press, Oxford, Great Britain, 2001.
- LESK, A. M.; NAD-binding domains of dehydrogenases. **Curr. Opin. Struct. Biol.** v.5, p.775-783, 1995.
- LEVITT, M.; Accurate modeling of protein conformation by automatic segment matching. **J Mol Biol** v.226, p.507-533, 1992.
- LEVITT, M.; GERSTEIN, M.; A unified statistical framework for sequence comparison and structure comparison. **Proc. Natl. Acad. Sci. USA** v.95, p.5913-5920, 1998.
- LINDAHL, E.; HESS, B.; VAN DER SPOEL, D.; GROMACS 3.0: a package for molecular simulation and trajectory analysis. **J. Mol. Mod.** v.7, p.306-317, 2001.
- LIPMAN, D. J.; PEARSON, W. R.; Rapid and sensitive protein similarity searches. **Science** v.227, p.1435-1441, 1985.
- LÜTHY, R.; MCLACHLAN, A. D.; EISENBERG, D.; Secondary structure-based profiles: Use of structure-conserving scoring tables in searching protein sequence databases for structural similarities. **Proteins** v.10, p.229-239, 1991.
- LUTHY, R.; BOWIE, J. U.; EISENBERG, D.; Assessment of protein models with three-dimensional profiles. **Nature** v.356, p.383-385, 1992.
- MACKERELL JR., A. D.; BASHFORD, D.; BELLITT, M.; *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. **J. Phys. Chem. B** v.102, p.3586-3616, 1998.
- MACLEAN, J.; ALI, S.; The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction. **Structure** v.11, p.1499-1511, 2003.
- MAJUMDER, K.; SELVAPANDIYAN, A.; FATTAH, F. A.; *et al.* 5-Enolpyruvylshikimate-3-phosphate synthase of *Bacillus subtilis* is an allosteric enzyme. **European Journal of Biochemistry** v.229, p.99-106, 1995.



- MACHEROUX, P.; SCHMID, J.; AMRHEIN, N.; *et al.* A unique reaction in a common pathway: mechanism and function of chorismate synthase in the shikimate pathway. **Planta** v.207, p.325-334, 1999.
- MARTI-RENOM, M. A.; STUART, A. C.; FISER, A.; *et al.* Comparative protein structure modeling of genes and genomes. **Annu. Rev. Biophys. Biomol. Struct.** v.29, p.291-325, 2000.
- MARTIN, A. C.; THORNTON, J. M.; Structural Families in Loops of Homologous Proteins: Automatic Classification, Modelling and Application to Antibodies. **J. Mol. Biol.** v.263, p.800-815, 1996.
- MARUCCI, R. C.; **Eficiência de transmissão de *Xylella fastidiosa* por cigarrinhas (Hemiptera, Cicadellidae) em *Citrus sinensis* (L.) Osbeck e *Coffea arabica* L.** Piracicaba, 2003. 139p. Tese (Doutorado) - Escola Superior de Agricultura Luiz de Queiroz, Universidade São Paulo.
- MARZABADY, M. R.; GRUYS, K. J.; PANSAGREAU, P. D.; *et al.* An EPSP synthase inhibitor joining shikimate 3-phosphate with glyphosate: synthesis and ligand binding studies. **Biochemistry** v.35, p.4199-4210, 1996.
- MATSUMOTO, R.; SALI, A.; GHILDYAL, N.; *et al.* Packaging of proteases and proteoglycans in the granules of mast cells and other hematopoietic cells. A cluster of histidines on mouse mast cell protease 7 regulates its binding to heparin serglycin proteoglycans. **J Biol Chem** v.270, p.19524-19531, 1995.
- MACKERELL JR., A. D.; BASHFORD, D.; BELLITT, M.; *et al.* All-Atom Empirical Potential for Molecular Modelling and Dynamics Studies of Proteins. **J. Phys. Chem.** v.B102, p.3586-3616, 1998.
- McREE, D. E.; XtalView/Xfit – A versatile Program for Manipulating Atomic Coordinates and Electron Density. **Journal Structural Biology** v.125, p.156-165, 1999.
- MEIDANIS, J.; BRAGA, M. D. V.; VERJOVSKI-ALMEIDA, S.; *et al.* Whole-genome analysis of transporters in the plant pathogen *Xylella fastidiosa*. **Microbiology and Molecular Biology Reviews** v.66, n.2, p.272-299, 2002.
- MERABET, E. K.; WALKER, M. C.; YUEN, H. K.; *et al.* Differential scanning calorimetric study of 5-enolpyruvylshikimate-3-phosphate synthase and its complexes with shikimate-3-phosphate and glyphosate: irreversible thermal transitions. **Biochimica Biophysica Acta** v.1161, p.272-278, 1993.
- MERRITT, E. A.; Raster3D Version 2.0. A program for photorealistic molecular graphics. **Acta Crystallogr.** v.50, p.869-873, 1994.
- MICHEL, G.; ROSZAK, A. W.; SAUVE V.; *et al.* Structures of Shikimate Dehydrogenase AroE and Its Paralog YdiB. **J. Biol. Chem.** v.278, n.21, p.19463-19472, 2003.



- MILLER, M. J.; BRACCOLINO, D. S.; CLEARY, D. G.; *et al.* EPSP synthase inhibitor design IV. New aromatic substrate analogs and symmetrical inhibitors containing novel 3-phosphate mimics. **Bioorg. Med. Chem. Lett.** v.21, p.2605-2608, 1994.
- MILLER, M. J.; CLEARY, D. G.; REAM, J. E.; *et al.* New EPSP synthase inhibitors: synthesis and evaluation of an aromatic tetrahedral intermediate mimic containing a 3-malonate ether as a 3-phosphate surrogate. **Bioorg. Med. Chem.** v.3, p.1685-1692, 1995.
- MILNER-WHITE, J. E.; COGGINS, J. R.; ANTON, I. A.; Evidence for an ancestral core structure in nucleotide-binding proteins with the type A-motif. **J. Mol.Biol.** v.221, p.751-754,1991.
- MIWA, J. M.; IBANEZ-TALLON, I.; CRABTREE, G. W.; *et al.* lynx1, an endogenous toxin-like modulator of nicotinic acetylcholine receptors in the mammalian CNS. **Neuron** v.23, p.105-114, 1999.
- MODI, S.; PAINE, M. J.; SUTCLIFFE, M. J.; *et al.* A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. **Biochemistry** v.35, p.4540-4550, 1996.
- MOLLENHAUER, H. H.; HOPKINS, D. L.; Xylem morphology of Pierce's disease infected grapevines with different levels of tolerance. **Physiological Plant Pathology**, v.9, p.95-100, 1976.
- MONTEIRO, P. B.; TEIXEIRA, D. C.; PALMA, R. R.; *et al.* Stable transformation of the *Xylella fastidiosa* citrus variegated chlorosis Strain with oriC plasmids. **Applied and Environmental Microbiology** v.67, n.5, p.2263-2269, 2001.
- MORRIS, A. L.; MACARTHUR, M. W.; HUTCHINSON, E. G.; *et al.* Stereochemical quality of protein structure coordinates. **Proteins** v.12, p.345-364, 1992.
- MOULT, J.; JAMES, M. N., An Algorithm for Determining the Conformation of Polypeptide Segments in Proteins by Systematic Search. **Proteins** v.1, p.146-163, 1986.
- MOUSEDALE, D. A.; COGGINS, J. R.; Purification and properties of 5-enolpyruvylshikimate-3-phosphate synthase from seedlings of *Pisum sativum* L. **Planta** v.160, p.78-83, 1984.
- MUNIZ, J. R. C. **Aplicação da bioinformática no estudo dos genes e enzimas envolvidos na síntese da goma fastidiana produzida pela *Xylella fastidiosa*.** São Carlos, 2003. 135 p. Tese (Mestrado) – Instituto de Física de São Carlos, Universidade de São Paulo.
- NEEDLEMAN, S.; WUNSCH, C.; A general method applicable to the search for similarities in the amino acid sequence of two proteins. **J. Mol. Biol.** v.48, p.443-453, 1970.
- NICHOLLS, A.; SHARP, K. A.; HONIG, B.; Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. **Proteins** v.11, n.4, p.281-296, 1991.



- ORENGO, C. A.; BRAY, J. E.; BUCHAN, D. W.; *et al.* The CATH protein family database: A resource for structural and functional annotation of genomes. *Proteomics* v.2, p.11-21, 2002.
- PADGETTE, S. R.; RE, D. B.; GASSER, C. S.; *et al.* Site-directed mutagenesis of a conserved region of the 5-enolpyruvylshikimate-3-phosphate synthase active site. *The Journal of Biological Chemistry* v.266, p.22364-22369, 1991.
- PEARSON, W. R.; Comparison of methods for searching protein sequence databases. *Protein Sci* v.4, p.1145-1160, 1995.
- PEARSON, W. R.; Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol* v.132, p.185-219, 2000.
- PEREIRA, E. F. **Estudos de fatores sazonais relacionados à transmissão de *Xylella fastidiosa* em pomares de citros.** Piracicaba, 2000. 82p. Tese (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade São Paulo.
- PEREIRA, J. H.; CANDURI, F.; DE OLIVEIRA, J. S.; *et al.* Structural bioinformatics study of EPSP synthase from *Mycobacterium tuberculosis*. *Biochem. Biophys. Res. Commun.* v.312, p.608-614, 2003.
- PEREIRA, J. H.; DE OLIVEIRA, J. S.; CANDURI, F.; *et al.* Structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimic acid. *Acta Cryst.* v.D60, p.2310-2319, 2004a.
- PEREIRA, J. H.; DE OLIVEIRA, J. S.; CANDURI, F.; *et al.* Interaction of shikimic acid with shikimate kinase. *Biochem. Biophys. Res. Commun.* v.325, p.10-17, 2004b.
- PRADO, S. S. **Eficiência de infecção e multiplicação se estirpes de *Xylella fastidiosa* de citros [*Citrus sinensis* (L.) Osbeck] e de cafeeiro (*Coffea arabica* L.) em inoculação cruzadas.** Piracicaba, 2003. 87p. Tese (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade São Paulo.
- PURCELL, A. H.; FINLAY, A. H. Evidence for noncirculative transmission of Pierce's disease bacterium by sharpshooter. *Phytopathology* v.69, p.393-395, 1979.
- PURCELL, A. H.; HOPKINS, D. L. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* v.34, p.131-151, 1996.
- RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V.; Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7:95-9, 1963.
- RAMACHANDRAN, G. N.; SASISKHARAN, V.; Conformation of polypeptides and proteins. *Advan. Prot. Chem.* v.23, p.283-437, 1968.
- RAMAKRISHNAN, C.; RAMACHANDRAN, G. N.; Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* v.5, p.909-933, 1965.



- RAMJEE, M. N.; BALASUBRAMANIAN, S.; ABEL, C.; *et al.* Reaction of (6R)-6-F-EPSP with recombinant *Escherichia coli* chorismate synthase generates a stable flavin mononucleotide semi-quinone radical. **J. Am. Chem. Soc.** v.114, p.3151-3153, 1992.
- RICHMOND, T. J.; RICHARDS, F. M.; Packing of alpha-helices: geometrical constraints and contact areas. **J Mol Biol** v.119, p.537-555, 1978.
- RING, C. S.; SUN, E.; MCKERROW, J. H.; *et al.* Structure-based inhibitor design by using protein models for the development of antiparasitic agents. **Proc Natl Acad Sci USA** v.90, p.3583-3587, 1993.
- ROBERTO, S. R.; COUTINHO, A.; LIMA, J. E. O. D.; *et al.* Transmission of *Xylella fastidiosa* by the sharpshooters *Dilobopterus costalimai*, *Acrogonia terminalis* and *Oncometopia facialis* in citrus. **Fitopatologia Brasileira** v.21, n.4, p.517-518, 1996.
- ROBERTS, C. W.; ROBERTS, F.; LYONS, R. E.; *et al.* The shikimate pathway and its branches in apicomplexan parasites. **J. Infect. Dis.** v.185, n.1, p.S25-36, 2002.
- ROBERTS, F.; ROBERTS, C. W.; JOHNSON, J. J.; *et al.* Evidence for the shikimate pathway in apicomplexan parasites. **Nature** v.393, n.6687, p.801-805, 1998.
- ROSSETTI, V.; TEIXEIRA, A. R. R.; CHAGAS, C. M.; *et al.* Chorose variegada dos citros (CVC) uma doenca transmissivel. **Arq. Inst. Biol.** v.57, p.14, 1990.
- ROSSMANN, M. G.; MORAS, D.; OLSEN, K. W.; Chemical and biological evolution of a nucleotidebinding protein. **Nature** v.250, p.194-199, 1974.
- ROSZAK, A.W.; ROBINSON, D. A.; KRELL, T.; *et al.* The structure and mechanism of the type II dehydroquinase from *Streptomyces coelicolor*. **Structure (Camb)** v.10, n.4, p.493-503, 2002.
- RUFINO, S. D.; DONATE, L. E.; CANARD, L. H.; *et al.* Predicting the Conformational Class of Short and Medium Size Loops Connecting Regular Secondary Structures: Application to Comparative Modelling. **J. Mol. Biol.** v.267, p.352-367, 1997.
- SALI, A; BLUNDELL, T. L.; Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. **J. Mol. Biol.** v.212, p.403-428, 1990.
- SALI, A; BLUNDELL, T. L.; Comparative protein modelling by satisfaction of spatial restraints. **J. Mol. Biol.** v.234, p.779-815, 1993.
- SALI, A.; OVERINGTON, J. P.; Derivation of Rules for Comparative Protein Modeling From a Database of Protein Structure Alignments. **Protein Sci.** v.3, p.1582-1596, 1994.
- SALI, A.; POTTERTON, L.; YUAN, F.; *et al.* Evaluation of comparative protein modeling by MODELLER. **Proteins** v.23, p.318-326, 1995.



SAMMONS, R. D.; GRUYS, K. J.; ANDERSON, K. S.; *et al.* Reevaluating glyphosate as a transition-state inhibitor of EPSP synthase: identification of an EPSP synthase-EPSP-glyphosate ternary complex. **Biochemistry** v.34, p.6433-6440, 1995.

SANCHEZ, R.; SALI, A.; Advances in comparative protein-structure modelling. **Curr Opin Struct Biol** v.7, p.206-214, 1997.

SANCHEZ, R.; SALI, A.; Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. **Proc. Natl Acad. Sci.** v.95, p.13597-13602, 1998.

SANCHEZ, R.; SALI, A.; Comparative protein structure modeling. Introduction and practical examples with modeller. **Methods Mol Biol** v.143, p.97-129, 2000.

SAUDER, J. M.; ARTHUR, J. W.; DUNBRACK JR., R. L.; Large-scale comparison of protein sequence alignment algorithms with structure alignments. **Proteins** v.40, p.6-22, 2000.

SCARPARI, L. M. **Modulação da expressão de genes de patogenicidade putativos em *Xylella fastidiosa* sob condições de alta e baixa densidade celular.** Piracicaba, 2001. 72p. Tese (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade São Paulo.

SCHALLER, A.; WINDHOFER, V.; AMRHEIN, N.; Purification of chorismate synthase from a cell culture of the higher plant *Corydalis sempervirens* **Pers. Arch. Biochem. Biophys.** v.282, p.437-442, 1990.

SCHALLER, A.; SCHMID, J.; LEIBINGER, U.; *et al.* Molecular cloning and analysis of a cDNA coding for chorismate synthase from the higher plant *Corydalis sempervirens* **Pers. J. Biol. Chem.** v.266, p.21434-21438, 1991a.

SCHALLER, A.; VAN AFFERDEN, M.; WINDHOFER, V.; *et al.* Purification and characterization of chorismate synthase from *Euglena gracilis*. **Plant Physiol.** v.97, p.1271 - 1279, 1991b.

SCHÖNBRUNN, E.; ESCHENBURG, S.; SHUTTLEWORTH, W. A.; *et al.* Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. **Proceedings of the National Academy of Sciences of the United States of America** v.98, p.1376-1380, 2001.

SCHWIETERS, C. D.; KUSZEWSKI, J. J.; TJANDRA N.; *et al.* The Xplor-NIH NMR Molecular Structure Determination Package. **J. Magn. Res.** v.160, p.66-74, 2003.

SETO, C. T.; BARTLETT, P.A.; (Z)-9-Fluoro-EPSP is not a substrate for EPSP synthase: implications for the enzyme mechanism. **J. Org. Chem.** v.59, p.7130-7132, 1994.

SHAH, D. M.; HORSCH, R. B.; KLEE, H. J.; *et al.* Engineering herbicide tolerance in transgenic plants. **Science** v.233, p.478481, 1986.



- SHAH, A.; FONT, J. L.; MILLER, M. J.; *et al.* New aromatic inhibitors of EPSP synthase incorporating hydroxymalonates as novel 3-phosphate replacements. **Bioorg. Med. Chem.** v.5, p.323-334, 1997.
- SHENG, Y.; SALI, A.; HERZOG, H.; *et al.* Site-directed mutagenesis of recombinant human beta 2-glycoprotein I identifies a cluster of lysine residues that are critical for phospholipid binding and anti-cardiolipin antibody activity. **J Immunol** v.157, p.3744-3751, 1996.
- SIEW, N.; ELOFSSON, A.; RYCHLEWSKI, L.; *et al.* MaxSub: an automated measure for the assessment of protein structure prediction quality. **Bioinformatics** v.6, n.9, p.776-785, 2000.
- SIKORSKI, J. A.; GRUYS, K. J.; Understanding glyphosate's molecular mode of action with EPSP synthase: evidence favoring an allosteric inhibitor model. **Acc. Chem. Res.** v.30, p.2-8, 1997.
- SIMPSON, A. J. G.; REINACH, F.C.; ARRUDA, P.; *et al.* The genome sequence of the plant pathogen *Xylella fastidiosa*. **Nature** v. 406, p.151-157, 2000.
- SMITH, T. F.; WATERMAN, M. S.; Identification of common molecular subsequences. **J. Mol. Biol.** v.147, p.195-197, 1981.
- SIPPL, M. J.; Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. **J. Mol. Biol.** v.213, p.859-883, 1990.
- SIPPL, M. J.; Knowledge-based potentials for proteins. **Curr Opin Struct Biol** v.5, p.229-235, 1995.
- SKARZYNSKI, T.; MISTRY, A.; WONACOTT, A.; *et al.* Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin. **Structure** v.4, p.1465-1474, 1996.
- SMART, C. C.; JOHILNUNG, D.; MILLER, G.; *et al.* Selective overproduction of 5-enolpyruvylshikimate 3-phosphate synthase in a plant cell culture which tolerates high doses of the herbicide glyphosate. **J. Biol. Chem.** v.260, p.16338-16346, 1985.
- SOUZA, L. C. A. **Produção e caracterização de mutantes do operon gum de *Xylella fastidiosa***. Piracicaba, 2002. 79p. Tese (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade São Paulo.
- SRINIVASAN, N.; BLUNDELL, T. L.; An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. **Protein Eng** v.6, p.501-512, 1993.
- STALLINGS, W. C.; ABDEL-MEGUID, S. S.; LIM, L. W.; *et al.* Structure and topological symmetry of the glyphosate target 5-enolpyruvylshikimate-3-phosphate synthase: A



- distinctive protein fold. **Proceedings of the National Academy of Sciences of the United States of America** v.88, p.5046-5050, 1991.
- STEINRÜCKEN, H. C.; AMRHELN, N.; The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid 3-phosphate synthase. **Biochem. Biophys. Res. Commun.** v.94, p.1207-1212, 1980.
- STEPHENS, C. M.; BAUERLE R.; Essential cysteines in 3-deoxy-uarabino-heptulosonate 7-phosphate synthase from *Escherichia coli*. **J Biol Chem** v.267, p.5762-5767, 1992.
- STUDELSKA, D. R.; KLUG, C. A.; BEUSEN, D. D.; *et al.* Longrange distance measurements of protein binding sites by rotational-echo doubleresonance NMR. **J. Am. Chem. Soc.** v.118, p.5476-5477, 1996.
- TAYLOR W. R.; FLORES T. P.; ORENGO C. A.; Multiple protein structure alignment. **Protein Sci** v.3, p.1858-1870, 1994.
- TOTROV, M.; ABAGYAN, R.; Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy. **Nat Struct Biol** v.1, p.259-263, 1994.
- TRAMONTANO, A.; LEPLAE, R.; MOREA, V; Analysis and assessment of comparative modeling predictions in CASP4. **Proteins** v.45, p.22-38, 2001.
- THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J.; CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. **Nucleic Acids Research** v.22, p.4673-4680, 1994.
- UCHÔA, H. B.; JORGE, G. E.; DA SILVEIRA, N. J. F.; *et al.* Parmodel: a web server for automated comparative modeling of proteins. **Biochem. Biophys. Res. Commun.** v.325, p.1481-1486, 2004.
- VAKSER, I. A.; Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. **Proteins Suppl** v.1, p.226-230, 1997.
- VAN GUNSTEREN, W. F.; BILLETER, S. R.; EISING, A.; *et al.* **Biomolecular Simulations: The GROMOS96 Manual and User Guide**. VdF Hochschulverlag ETHZ, Zurich, 1996.
- VAN SLUYS M. A.; DE OLIVEIRA M. C.; MONTEIRO-VITORELLO C. B.; *et al.* Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. **J Bacteriol.** v.185, n.3, p.1018-26, 2003.
- VAN VLIJMEN, H. W.; KARPLUS, M.; PDB-Based Protein Loop Prediction: Parameters for Selection and Methods for Optimization. **J. Mol. Biol.** v.267, p.975-1001, 1997.
- VOGAN, E.; Shikimate dehydrogenase structure reveals novel fold. **Structure** v.11, p.902-903, 2003.



- VONRHEIN, C.; SCHLAUDERER, G. J.; SCHULZ, G. E.; Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. **Structure** v.3, n.5, p.483-490, 1995.
- VRIEND, G.; WHAT IF: A molecular modeling and drug design program. **J. Mol. Graph.** v.8, p.52-56, 1990.
- WAGNER, T.; SHUMILIN, I. A.; BAUERLE, R.; *et al.* Structure of 3-Deoxy-D-arabino-heptulosonate-7-phosphate Synthase from *Escherichia coli*: Comparison of the Mn^{2+} -2-Phosphoglycolate and the Pb^{2+} -2-phosphoenolpyruvate Complexes and Implications for Catalysis. **J. Mol. Biol.** v.301, p.389-399, 2000.
- WALKER G. E.; DUNBAR B.; HUNTER I.S.; *et al.* Evidence for a novel class of microbial 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase in *Streptomyces coelicolor* A3(2), *Streptomyces rimosus* and *Neurospora crassa*. **Microbiology** v.142, p.1973-1982, 1996.
- WALKER, J. E.; SARASTE, M.; RUNSWICK, M. J.; *et al.* Distantly related sequences in the α - and β - subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. **EMBO Journal** v.1, p.945-951, 1982.
- WALSH, C. T.; LIU, J.; RUSNAK, F.; *et al.* Molecular studies on enzymes in chorismate metabolism and the enterbactin biosynthetic pathway. **Chemical Reviews** v.90, p.1105-1129, 1990.
- WELCH, G. R.; COLE, K. W.; GAERTNER, F. H.; Chorismate synthase of *Neurospora crassa*: A flavoprotein. **Arch. Biochem. Biophys.** v.165, p.505-518, 1974.
- WELL, J. M.; RAJU, B. C.; HUNG, H. Y.; *et al.* *Xylella fastidiosa* gen. nov. sp. nov.: Gram-negative, xylem-limited fastidious plant bacteria related to *Xanthomonas* spp. **International Journal of Systematic Bacteriology** v.37, p.136-143, 1987.
- WESTBROOK, J.; FENG, Z.; JAIN, S.; *et al.* The Protein Data Bank: unifying the archive. **Nucleic Acids Res** v.30, p.245-248, 2002.
- WHITE, P. J.; MILIAR, G.; COGGINS, J. R.; The overexpression, purification and complete amino acid sequence of chorismate synthase from *Escherichia coli* K12 and its comparison with the enzyme from *Neurospora crassa*. **Biochem. J.** v.251, p.313-322, 1988.
- WILBUR, W. J.; LIPMAN, D. J.; Rapid similarity of nucleic acid and protein data banks. **Proc. Natl. Acad. Sci. USA** v.80, p.726-730, 1983.
- WOLF, E.; VASSILEV, A.; MAKINO, Y.; *et al.* Crystal structure of a GCN5-related N-acetyltransferase: *Serratia marcescens* aminoglycoside 3-N-acetyltransferase. **Cell** v.94, p.439-449, 1998.
- WU, G.; FISER, A.; TER KUILE, B.; *et al.* Convergent evolution of *Trichomonas vaginalis* lactate dehydrogenase from malate dehydrogenase. **Proc Natl Acad Sci USA** v.96, p.6285-6290, 1999.



- XIANG, Z.; HONIG, B.; Extending the Accuracy Limits of Prediction for Side-Chain Conformations. **J. Mol. Biol.** v.311, p.421-430, 2001.
- XU, L. Z.; SANCHEZ, R.; SALI, A.; *et al.* Ligand specificity of brain lipidbinding protein. **J Biol Chem** v.271, p.24711-24719, 1996.
- YAN, H.; YSAI, M.-D.; Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. **Advan. Enzymol. Relat. Areas Mol. Biol.** v.73, p.103-134, 1999.
- YORINORI, M. A.; RIBAS, A. F.; UENO, B.; *et al.* Detection of *Xylella fastidiosa* in coffee germplasm. **Fitopatol. bras.** v. 28, p.427-430, 2003.
- ZEMLA, A.; VENCLOVAS, C.; MOULT, J.; *et al.* Processing and analysis of CASP3 protein structure predictions. **Proteins** v.Suppl3,p.22-29, 1999.
- ZEMLA, A.; LGA: A method for finding 3D similarities in protein structures. **Nucleic Acids Res** v.31, n.13, p.3370-3374, 2003.
- ZHANG, Y.; SKOLNICK, J.; Scoring function for automated assessment of protein structure template quality. **Proteins** (em impressão), 2004.



APÊNDICE V – ARTIGO

Available online at www.sciencedirect.com



BBRC

Biochemical and Biophysical Research Communications 320 (2004) 979–981

www.elsevier.com/locate/ybbrc

Molecular models for shikimate pathway enzymes of *Xylella fastidiosa*

Helen Andrade Arcuri,^{a,d} Fernanda Canderi,^{a,d,1} José Henrique Pereira,^a
Nelson José Freitas da Silveira,^a João Carlos Câmara Jr.,^a Jaime Simões de Oliveira,^b
Luiz Augusto Basso,^b Mário Sérgio Palma,^{a,d} Diógenes Santiago Santos,^{c,*}
and Walter Filgueira de Azevedo Jr.^{a,d,*}

^a Department of Physics IBILCE/UNESP, São José do Rio Preto, SP 13066-700, Brazil

^b Instituto de Pesquisas em Tuberculose, Department of Molecular Biology and Biotechnology, UNBIO, Porto Alegre, RS 91501-970, Brazil

^c Laboratory of Structural Biology and Zoonosis, Department of Biology, Institute of Electronics, UNESP, Rio Claro, SP 13506-900, Brazil

^d Center for Applied Technology, Institute Biotecn, São Paulo, SP 05389-000, Brazil

^e Center for Research and Development in Molecular, Structural and Functional Molecular Biology, PUCRS 91501-900, Porto Alegre, RS, Brazil

Received 30 May 2004

Available online 25 June 2004

Abstract

The *Xylella fastidiosa* is a bacterium that is the cause of citrus variegated chlorosis (CVC). The shikimate pathway is of great importance for production of a plethora of aromatic compounds in plants, bacteria, and fungi. Putative structural differences in the enzymes from the shikimate pathway, between the proteins of bacteriophages and those of plants, could be used for the development of a drug for the control of CVC. However, inhibitors for shikimate pathway enzymes should have high specificity for *X. fastidiosa* enzymes, since they are also present in plants. In order to pave the way for structural and functional efforts towards antimicrobial agent development, here we describe the molecular modeling of seven enzymes of the shikimate pathway of *X. fastidiosa*. The structural models of shikimate pathway enzymes, complexed with inhibitors, strongly indicate that the previously identified inhibitors may also inhibit the *X. fastidiosa* enzymes.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Shikimate pathway; Structural Bioinformatics; Drug design; *Xylella fastidiosa*

Xylella fastidiosa (Xf) is a fastidious and gram-negative, xylem-limited bacterium which causes the citrus variegated chlorosis (CVC), a serious disease of citrus fruits [1]. The disease spread rapidly by graft propagation with infected barwood and by sharpshooter vectors, and became widely distributed in all citrus growing regions of Brazil. CVC is one of the major concerns to the Brazilian citrus industry. Certain strains of Xf also affect other commercially important products such as coffee, apple, and other fruits [2].

In plants and microorganisms, all the key aromatic compounds involved in primary metabolism, including the three aromatic amino acids found in proteins, are

produced via shikimate pathway. The shikimate pathway is essential in algae, higher plants, bacteria, and fungi, but absent from mammals that depend on these compounds for their diet [3].

Xylella fastidiosa appears to be capable of synthesizing an extensive variety of enzyme cofactors and prosthetic groups, including biotin, folic acid, pantothenate and coenzyme A, ubiquinone, glutathione, thiorubrin, glutaredoxin, riboflavin, flavin mononucleotide (FMN), haem adenine dinucleotide (FAD), pyrimidine nucleotides, porphyrin, thiamin, pyridoxal 5-phosphate, and lipotein [4].

The shikimate pathway consists of seven enzymes that catalyze the sequential conversion of erythrose-4-phosphate and phosphoenolpyruvate to chorismate [5] (Fig. 1). All pathway intermediates can also be considered branch point compounds that may serve as substrates for other metabolic pathways [6].

* Corresponding author.

E-mail address: walter@bilce.unesp.br (W.F. de Azevedo Jr.).

H.A.A. and F.C. contributed equally to this work.



Molecular models for shikimate pathway enzymes of *Xylella fastidiosa*

Helen Andrade Arcuri,^{a,1} Fernanda Canduri,^{a,d,1} José Henrique Pereira,^a
Nelson José Freitas da Silveira,^a João Carlos Camera Jr.,^a Jaim Simões de Oliveira,^b
Luiz Augusto Basso,^b Mário Sérgio Palma,^{c,d} Diógenes Santiago Santos,^{c,*}
and Walter Filgueira de Azevedo Jr.^{a,d,*}

^a Department of Physics IBILCE/UNESP, São José do Rio Preto, SP 15054-000, Brazil

^b Rede Brasileira de Pesquisas em Tuberculose, Department of Molecular Biology and Biotechnology, UFRGS, Porto Alegre, RS 91501-970, Brazil

^c Laboratory of Structural Biology and Zoochemistry, Department of Biology, Institute of Biosciences, UNESP, Rio Claro, SP 13506-900, Brazil

^d Center for Applied Toxicology, Institute Butantan, São Paulo, SP 05503-900, Brazil

^e Center for Research and Development in Molecular, Structural and Functional Molecular Biology, PUCRS 90619-900, Porto Alegre, RS, Brazil

Received 25 May 2004

Available online 25 June 2004

Abstract

The *Xylella fastidiosa* is a bacterium that is the cause of citrus variegated chlorosis (CVC). The shikimate pathway is of pivotal importance for production of a plethora of aromatic compounds in plants, bacteria, and fungi. Putative structural differences in the enzymes from the shikimate pathway, between the proteins of bacterial origin and those of plants, could be used for the development of a drug for the control of CVC. However, inhibitors for shikimate pathway enzymes should have high specificity for *X. fastidiosa* enzymes, since they are also present in plants. In order to pave the way for structural and functional efforts towards antimicrobial agent development, here we describe the molecular modeling of seven enzymes of the shikimate pathway of *X. fastidiosa*. The structural models of shikimate pathway enzymes, complexed with inhibitors, strongly indicate that the previously identified inhibitors may also inhibit the *X. fastidiosa* enzymes.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Shikimate pathway; Structural bioinformatics; Drug design; *Xylella fastidiosa*

Xylella fastidiosa (*Xf*) is a fastidious and gram-negative, xylem-limited bacterium which causes the citrus variegated chlorosis (CVC), a serious disease of citrus fruit [1]. The disease spread rapidly by graft propagation with infected budwood and by sharpshooter vectors, and became widely distributed in all citrus growing regions of Brazil. CVC is one of the major concerns to the Brazilian citrus industry. Certain strains of *Xf* also affect other commercially important products such as coffee, nuts, and other fruits [2].

In plants and microorganisms, all the key aromatic compounds involved in primary metabolism, including the three aromatic amino acids found in proteins, are

produced via shikimate pathway. The shikimate pathway is essential in algae, higher plants, bacteria, and fungi, but absent from mammals that depend on these compounds for their diet [3].

Xylella fastidiosa appears to be capable of synthesizing an extensive variety of enzyme cofactors and prosthetic groups, including biotin, folic acid, pantothenate and coenzyme A, ubiquinone, glutathione, thioredoxin, glutaredoxin, riboflavin, flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), pyrimidine nucleotides, porphyrin, thiamin, pyridoxal 5'-phosphate, and lipoate [1].

The shikimate pathway consists of seven enzymes that catalyze the sequential conversion of erythrose-4-phosphate and phosphoenolpyruvate to chorismate [4] (Fig. 1). All pathway intermediates can also be considered branch point compounds that may serve as substrates for other metabolic pathways [5].

* Corresponding author.

E-mail address: walterfa@df.ibilce.unesp.br (W.F. de Azevedo Jr.).

¹ H.A.C. and F.C. contributed equally to this work.

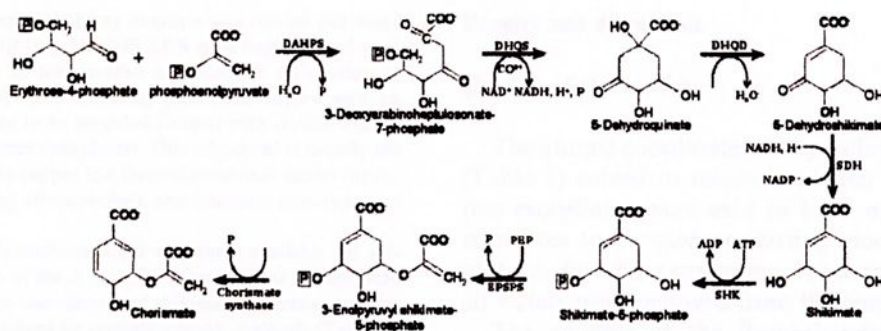


Fig. 1. The shikimate pathway in the sequence of seven metabolic steps, from phosphoenolpyruvate and erythrose 4-phosphate until the conversion to chorismate.

The molecular organization and structure of the shikimate pathway enzymes varies considerably between microorganism groups. Bacteria have seven individual polypeptides, each possessing a single enzyme activity, which are encoded by separated genes. Plants have a molecular arrangement similar to bacteria, i.e., separated enzymes encoded by separated genes, with the exception of dehydroquinase and shikimate dehydrogenase, which have been shown to be present as separated domains on a bifunctional polypeptide. Plant enzymes, although nuclear encoded, are largely active in the chloroplast and accordingly possess an N-terminal transit sequence. In contrast, all fungi examined to date have monofunctional 3-deoxy-D-arabino-heptulosonate-7-phosphate synthases and chorismate synthases and a pentafunctional polypeptide termed AROM. The AROM polypeptide has domains analogous to the bacterial enzymes: dehydroquinase synthase (DHQS), 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), shikimate kinase (SK), dehydroquinase (DHQD), and shikimate dehydrogenase (SDH) [4].

Chorismate is converted by five distinct enzymes to prephenate, anthranilate, aminodeoxychorismate, isochorismate, and *p*-hydroxybenzoate. These metabolites comprise the first committed intermediates in the biosynthesis of Phe, Tyr, Trp, folate, menaquinone, and the siderophore enterobactin, and ubiquinone, respectively. The synthesis of these precursors is in most cases highly regulated [6].

In monogastric animals, Phe and Trp are essential amino acids that have to come with the diet and Tyr is directly derived from Phe. Since bacteria use in excess 90% of their metabolic energy for protein biosynthesis, for most prokaryotes, the three aromatic amino acids represent nearly the entire output of aromatic biosynthesis, and regulatory mechanisms for shikimate pathway activity are triggered by the intracellular concentrations of Phe, Tyr, and Trp. This is not so in higher plants, in which the aromatic amino acids are the precursors for a large variety of secondary metabolites with aromatic ring structures that often make up a substantial amount of the total dry weight of a plant. Among the many aromatic secondary

metabolites are flavonoids, many phytoalexins, indole acetate, alkaloids such as morphine, UV light protectants, and, most important, lignin [7].

In microorganisms, the shikimate pathway is regulated both by feedback inhibition and by repression of the first enzyme 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHPS). In higher plants, no physiological feedback inhibitor has been identified, suggesting that pathway regulation may occur exclusively at the genetic level. This difference between microorganisms and plants is reflected in the unusually large variation in the primary structures of the respective first enzymes [5].

Shikimate pathway enzymes are attractive targets for drug development, however any new inhibitor for shikimate pathway enzymes should have high specificity for *Xf* enzymes, since they are also present in the plants. Therefore, detailed structural information about the protein targets will help us to identify structural differences between enzymes of plant and *Xf*, which may serve as basis to design specific inhibitor against shikimate pathway enzymes of the *X. fastidiosa*. The seven enzymes modeled in the present work were 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (DAHPS), 3-dehydroquinase synthase (DHQS), 3-dehydroquinase dehydratase (DHQD), shikimate-5-dehydrogenase (SDH), shikimate kinase (SK), 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), and chorismate synthase (CS).

Knowledge of the three-dimensional structures of the enzymes will undoubtedly aid in the design of useful inhibitors that may be used as a bactericide against *X. fastidiosa*.

Methods

Molecular modeling. Homology modeling is usually the method of choice when there is a clear relationship of homology between the sequence of a target protein and at least one known structure. This computational technique is based on the assumption that tertiary structures of two proteins will be similar if their sequences are related, and it is the approach most likely to give accurate results [8]. There are two main approaches to homology modeling: (1) fragment-based comparative modeling [9,10] and (2) restrained-based modeling [11]. For modeling of the shikimate pathway we used the second approach.

Model building of shikimate pathway enzymes was carried out using the program MODELLER [11]. MODELLER is an implementation of an automated approach to comparative modeling by satisfaction of spatial restraints [12–14]. The modeling procedure begins with an alignment of the sequence to be modeled (target) with related known three-dimensional structures (templates). This alignment is usually the input to the program. The output is a three-dimensional model for the target sequence containing all main-chain and sidechain non-hydrogen atoms.

Since there are no three-dimensional structures available for shikimate pathway enzymes of the *X. fastidiosa* (targets) the models were built based on the atomic coordinates of shikimate pathway enzymes (templates), which were solved by crystallographic methods (Table 1). The percentage of primary sequence identity (Table 1) between enzymes of *Xf* and of templates indicates that the crystallographic structures of enzymes (templates) are good models to be used as templates for the enzymes of *X. fastidiosa* (targets). The alignment of enzymes of the *X. fastidiosa* (targets) and of templates is shown in Figs. 2A–G.

Next, the spatial restraints and CHARMM energy terms enforcing proper stereochemistry [15] were combined into an objective function. Finally, the model is obtained by optimizing the objective function in Cartesian space. The optimization is carried out by the use of the variable target function method [16] employing method of conjugate gradients and molecular dynamics with simulated annealing. Several slightly different models can be calculated by varying the initial structure. A total of 1000 models were generated for each enzyme and the final models were selected based on stereochemical quality. All optimization process was performed on a Beowulf cluster with 16 nodes (BioComp, AMD Athlon XP 1800+).

Analysis of the models. The overall stereochemical quality of the final models for each enzyme of the shikimate pathway of the *X. fastidiosa* was assessed by the program PROCHECK [17]. The root mean square deviation (rmsd) between C α –C α atom's distance was superposed using the program LSQKAB from CCP4 [18]. The cutoff for hydrogen bonds and salt bridges was 3.3 Å. The contact area for the complexes was calculated using AREAIMOL and RESAREA [18]. The root mean square deviation (rmsd) differences from ideal geometries for bond lengths and bond angles were calculated with X-PLOR [19,20]. The G-factor value is essentially just log-odds score based on the observed distributions of the stereochemical parameters. It is computed for the following properties: torsion angles (the analysis provided the observed distributions of ϕ – ψ , χ_1 – χ_2 , χ_1 , χ_3 , χ_4 , and ω values for each of the 20 amino acid types) and covalent geometry (for the main-chain bond lengths and bond angles). These values' average was calculated using PROCHECK [17]. The Verify-3D measures the compatibility of a protein model with its sequence, using a 3D profile [21–23].

Results and discussion

Quality of the models

The atomic coordinates of crystallographic structures (Table 1) solved to resolution better than 2.0 Å (with two exceptions) were used to build up an ensemble of structures to be used as starting models for modeling of the *X. fastidiosa* structures. The atomic coordinates of all waters were removed from the templates.

The analysis of the Ramachandran diagram ϕ – ψ plots for the templates (Table 2) was used to compare the overall stereochemical quality of the *X. fastidiosa* structures against that of templates solved by biocrystallography. They present over 89.0% of the residues in the most favorable regions.

Overall description and interactions with inhibitors

Among the seven enzymes modeled in the present work four proteins were modeled with inhibitors (DAHPS:Mn²⁺:PGL, DHQS:CRB:NAD, DHQD:FA1, and EPSPS:glyphosate:S3P), because only these templates have inhibitors in their crystallographic structure.

The structures of all molecular models were superposed with their templates (Figs. 6A–G). The rmsd for each superposition is given in Table 1. It was considered only C α in the superposition. The rmsd value of bond lengths and bond angles is presented in Table 2.

The superpositions of the complexes are shown in Figs. 5A–E and the average G-factor and Verify 3D values are shown in Table 3.

3-Deoxy-D-arabino-heptulosonate-7-phosphate synthase

The first step of the shikimate pathway is the condensation of phosphoenolpyruvate (PEP) with the four-carbon sugar D-erythrose-4-phosphate (E4P) to produce 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP)

Table 1

The crystallographic structures of the shikimate pathway enzymes used as templates, the percentage of amino acid identity, and rmsd value after superposition of templates and targets

Enzymes	PDB access code	Resolution (Å)	References	Organism	Sequence identity (%)	rmsd (Å)	
						Site	Protein
DAHPS ^a	1GG1	2.00	[24]	<i>Escherichia coli</i> (Ec)	61.1	0.086	0.665
DHQS ^a	1DQS	1.80	[26]	<i>Emmericella nidulans</i> (En)	31.8	0.055	0.039
DHQD ^a	1GU1	1.80	[30]	<i>Streptomyces coelicolor</i> (Sc)	42.5	0.059	0.055
SDH ^a	1NYT	1.50	[32]	<i>E. coli</i> (Ec)	42.3		0.166
SK ^a	1L4Y	2.00	[38]	<i>Mycobacterium tuberculosis</i> (Mt)	32.6		0.452
	1WE2	2.30	[39]	<i>M. tuberculosis</i> (Mt)	32.6		0.366
	2SHK	2.60	[40]	<i>Erwinia chrysanthemi</i> (Ech)	31.9		0.377
EPSPS ^a	1G6S	1.50	[46]	<i>E. coli</i> (Ec)	29.9	1.419	0.118
CS ^a	1QXO	2.00	[49]	<i>Streptococcus pneumoniae</i> (Sp)	30.6	0.582	0.650

^aDAHPS, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHQS, 3-dehydroquinase synthase; DHQD, 3-dehydroquinase dehydratase; SDH, shikimate-5-dehydrogenase; SK, shikimate kinase; EPSPS, 5-enolpyruvylshikimate-3-phosphate synthase; and CS, chorismate synthase.

A	1GG1	ENYQNDLRIKEIKELLPPVALLKFPATENAANTVAHARKAIHKILKGNDRLLVVGIP
	XfDAHPS	ENPFTDDLRIKIEPLTPNQLLLQLPCDEQASLTVATSRNALHDLHGENDRLAVVIGP * .*****::: * * * * :*: * *: * *: * *: * *: * *: * *: * *: *
	1GG1	CSIHDPVAAKEYATRLALREELKDELEIVERVYFEKPRRTIVGKGLINDPHMNSFQIN
	XfDAHPS	CSIHDPVAAKEYATRLALREELKDELEIVERVYFEKPRRTIVGKGLINDPHMNSFQIN *****::: * * * * :*: * *: * *: * *: * *: * *: * *: * *: *
	1GG1	DGLRIARKLLLDINDSGLPAAGEFLDMITPQYLADLMSWGAIGARTTESQVHRELASGLS
	XfDAHPS	KGLHIARNLLRDINHGLPAGVEFLDTISPQYLADLVAWGAIGARTTESQIHREHASGLS * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: *
	1GG1	CPVGFKNGTDTIKVAIDAINAAGAPHCFLSVTKUGHSATVNTSGNGDCHIILRGKKEPN
	XfDAHPS	CPVGFKNSTSGDVKIAADAIAASHPHFLSVTKTGVAAVVATQGNPDCHVILRGGRVFN * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: *
	1GG1	YSAKHVAEVKEGLNKAGLPAQVHIDFSHANSSKQFKKQNDVCADVCQIAGGEKAIIGVM
	XfDAHPS	YDAINVEAASQTLKSLPVRLLHIDASHANSGKNPNQPKVNIENIVEQLEAGERIRIVGM * . * : * . . : * * : * *: * *: * *: * *: * *: * *: * *: * *: *
	1GG1	VESHVLEGNQSLSEGEPLAYGKSITDACIGWEDTDALLRLANAVKARRG
	XfDAHPS	IESHIVGGRQLVLGNTLTGQSITDACIDWDTVAHLEQLAQAVRVROT : * *: * . * . * : * *: * *: * *: * *: * *: * *: * *: * *: *
B	1DQS	---MSNPTKISILGRESIIADFLWRNYVAKDLISDCSTTYVLVTDNIGSIYTPSFEE
	XfDHQS	HTTLPLRSVTNNTPYTTIATGGLLDPLRLAATIRGRHALILSDSEVAPRYAAQLQE : : . : : . . : * : . * : . : : : * : : . * : : *
	1DQS	AFRKRAAEITPSPRLLIYNRPPEVSKSRQTKADIEDWHLSONPCCRDTVVIALGGGVI
	XfDHQS	TLLR----ARPDHLNVFTLPAGETSKLENFSAIAQLATLG---ATRDACLALGGGVI : : : * . : * : . * . * * : : : : : . * : : * : * : *
	1DQS	GDLTGFVASTYERGVYVQVPTTLANVDSSIGGKTAIDTPLGNLIGAIWQPTKIYIDL
	XfDHQS	GDLAGFTAACWHRGIDYVQVPTTLANVDSSVGGKTAVIDPQGNHVGAFHPRAVIADT * *: * *: * : * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: *
	1DQS	EFLTLPVREFINGHAEVIKTAISSEEEFTALEENAETILKAVRREVTPGEHRFEGTEE
	XfDHQS	DTLATPLRELRLAGLSEVIKYGAIRDPVFFHULQTTREALLARDP-----A : * * *: * *: * : * *: * * . * * : . * : *
	1DQS	ILKARILASARHKATVVSADEREGLRNLLNUGHSIGHAIEAILTP-----QILHGECV
	XfDHQS	ALAQAIRASCEKADIVGRDPLEKGERVLLNLGHTFGHAIETAQGYSTPGNNLNHGEAV * * * . * . * : * . * * * * * * * : * : * : * : *
	1DQS	AIGHVKEAEALARHLGILKGVAVSRIVKCLAAAYGLPTSLKDARIRKLTAGKECSVDQLNFM
	XfDHQS	AVGVLAARLSNALSLAPADTETLKNLLDAYGLPTVLPSG-----LTPENLLER * : * * * . * . : . . : : * * * * * * . : : * : .
	1DQS	HALDKKNDGPKKIVLLSAIGTPYETRASVVANEDIRVVLP
	XfDHQS	HRLDKKNIAGRRLRLVLRGIG--HAEVSDVDEATVRQILAN * * * * * . : : * * . * * : . * * : : * : * : *

Fig. 2. (A) The sequence alignment for DAHPS from *X. fastidiosa* and *E. coli* (1GG1); (B) DHQS from *X. fastidiosa* and *E. nidulans* (1DQS); (C) DHQD from *X. fastidiosa* and *S. coelicolor*; (D) SDH from *X. fastidiosa* and *E. coli* (1NYT); (E) SK from *X. fastidiosa* and *Mycobacterium tuberculosis* (1L4Y–1WE2), *E. chrysanthemi* (2SHK); (F) EPSPS from *X. fastidiosa* and *E. coli* (1G6S); and (G) CS from *X. fastidiosa* and *Streptococcus pneumoniae* (1QXO). The alignments were performed with the program CLUSTAL W [54], the meaning of the special characters in the last line being: (*) indicates positions which have a single, fully conserved, (:) indicates that one of the following 'strong' groups is fully conserved, and (.) indicates that one of the following 'weaker' groups is fully conserved.

and inorganic phosphate [5]. This reaction is catalyzed by the enzyme DAHPS, a metal-activated enzyme, which in microorganisms is the target for negative-feedback regulation by pathway intermediates or by end products. The basic structure of DAHP was confirmed by several different chemical syntheses [7]. Typically, the comparison of the primary amino acid sequence of plant with DAHPS shows a surprisingly low 20% pairwise identity. In fact, PRETTYBOX (GCG program package), a program used to make multiple sequence alignments, identifies only 24 invariant residues for all known DAHPS [5].

The Fig. 3A shows the secondary structure of *Xf*DAHPS in the complex with manganese (II) ion (Mn^{2+}) and 2-phosphoglycolate (PGL). This model shows that the protein has 8 (β/α) barrel enhanced by several additional secondary structure elements, including a C-terminal tail that forms a three-stranded, anti-parallel β -sheet with the $\beta 6a/\beta 6b$ strands [24]. The active site of *Xf*DAHPS consists of a channel between C-terminal and N-terminal region of the 8 (β/α) barrel with bound Mn^{2+} and PGL.

We observed a total of 12 hydrogen bonds between *Xf*DAHPS: Mn^{2+} :PGL, in binary model, involving the


```
F      1G6S      MESLTLQPIARVDGTINPGSKSVSNRALLAALAHAGKTVLTNLLSDDVVRHMLNALTAL
XIEPSPS HDYWIHQGTPLHGVLSPDGKSISHRAVHF AALADGTSRIDGFLEAEDTCSTAEILARL
          :   :   : .,*.:**.*:*::***.*.:   : .*:**:*.   : *:*

1G6S      GVSYTLSADRTRCEIIGNGGPLHAEGALEFLGNAGTAMRPLAAALCLGSNDIVLTGEPR
XIEPSPS  GVRIETPLS-TQRIVEHGVDGLQASHPLDCGNAGTGHRLLAGLLVAQFPD SVLVGDAS
**       . . * : : * *           . : * ****.* ** * . * ***,*:.

1G6S      HKERPIGHLVDALRLGGAKITYLEQENYPPLRLQGFFTGGNVVDVDG-SVSSQFLTALLE
XIEPSPS  LSKRP HRRVTDPLSQHGARIDTSD-DGTPPLRIYGGQLLHGIDFISFVASAQIKSAVLLA
:,:*: : .*,* **:* : : . ****: **     .*,. . .*: : :*: :

1G6S      APLAPEDTVIRIKGDLVSKPYIDITLNLKFTGV E IENQHYYQQFVVKGGSYSQSGTYLV
XIEPSPS  GLYARNETVVREP-----HPTRDYTERELTAFGVDID-VSTGCARLRGGORLCATDITIP
. * :*:*: *         : * * .:~::~***:*:             :***    .. :

1G6S      EGDASSASYFLAAAAIKGGTVKV T GIGRNSHQGDIRFADVLEKNGATICUGDDYIS----
XIEPSPS  ADFSSAAFYLVAASVIPGSDITLRAVGLNPRR--IGLLTVLRLHGANIVESNRHEQGQGP
. :*,* **:~::~* *. : : .,* *. : * : **, ***.* . : : .

1G6S      -----CTRGE LNADHD MNHIPDAANTIAALFAKGTTTLRNINYNRVKETDRLF
XIEPSPS  VVDLRVRYAPLGQTRVPEDLVADHIDEFFPALFVAAAAGQT VVGAAELRVKESDRLAA
..       . . : : : * : : ** *:* *. : . : *****:**

1G6S      MATELRKVGAEEVEEGHDYIRITPEKLNF AEIATYNDHRMANCFSLVALSDTPVTILDPK
XIEPSPS  NVTGLRVLGVOVDETADGATINGG-PIGHGTINSRGDHRIANAFSIAGQLSVSTVRIEDV
*,* ** :*,*:~* * *        :... * :~::~***:**.*:.. ..... :

1G6S      CTAKTFPPDYFEQLARISQAA
XIEPSPS  ANVATSFPDYETLARISAGFG
... *      : * *** : .

G      1QXO      -----MRYLTAGESHGPRLTAIIEGIPAGLP LTAEDINEDLRRQGGYGRGGRNKI
XiCS      NGANTFGKLLAVTTTFGESHGPAIGCVIDGCCPPGLELAEEFAHDLQRRATGRSRHTSARR
          :   * ***** : ,*:~* ** *:~::~.**,~* * *,* :

1QXO      ENDQVFTSGVRHGKTGAPITHDVINKDHQKWLDINSAEDI EDRLKSKRKITHPRPGA
XiCS      EADEVEILSGVYEGR TTGTPIALLRNTDQRSKYATIAR-----QFRPGA
* *:* : *** .*:~::~***: : *,*:.          * : *****

1QXO      DLVGGIKYRFDDL RNSLERSSARETTH RVAUGAVAKRLLAELDHEIANHVVFVGGKEIDV
XiCS      DTTYWQKYGI RDPRGGG-RSSARETTH RVAAGVAKWKQR-----YGVIVRGFLSQL
* .   ** : * .. *****~::~*,***: * :           :*,. * : :

1QXO      PENLTVAEIKQRAAQSEVSIVNQEREQE IKDYIDQIKRDGDTIGGVVETVVGGV PVGLGS
XiCS      GEIRPEGFAU DAVEDNPFFWPQA AQVPELEYND ALRKSNGSVGARVDVVAE GVPFGWE
* . .   . : . . : : ~*:* * :~::~*,*:,*,*, *** * *,

1QXO      YVQWRDKLDARLAQAVV SINAFKGV E FGLGFEAGYRKGSQVHDEILWSKEDGYTRRTNNL
XiCS      PIYG--KLDGELAALAALMS INAVGVEIGAGFGSTVQKGTEHRDLNTP-----LGFLSNHA
:   ***.,** *:~::~****,* ** : ~*~*~* : * :           :*:

1QXO      GGFEGGTINGQPIVVRGVHKPIPTLYKPLEHSV DIETH EPHYKATVERS DPTALPAAGHVZE
XiCS      GGIIGGITTTGQPIIVSIALKPTSSLRLPGETVDVDGHPVQVITKRHDPCV GIRAPP IAE
**; ***,*,***** .:~*~* .,* * :~::~* * * * * . * * *

1QXO      AVVATVLAQEILEKFSSD-NLEELKEAVAKHRDYTKNY-
XiCS      AHVALVLMDQALRHRAQC GDVGEMSPCILENVGFNRADD
*:~*~* ** : : * : . : . : : : : : : : : : : : : : : :
```

acid (CRB) and nicotinamide-adenine-dinucleotide (NAD⁺). The structure is composed of an N-terminal α/β domain and a C-terminal α -helical domain [26].

The N-terminal domain includes a Rossmann fold, but surprisingly binds NAD⁺ in an inverted orientation to that observed in all other classic Rossmann-fold proteins. This domain consists of a seven-stranded

β -sheet, with strand order 1296534 (Fig. 3B). Strands $\beta 1$ and $\beta 2$ form a β -hairpin which is connected by helix $\alpha 1$ to a five-stranded Rossmann-fold structure [27,28]. This conformation orientates the NAD⁺ so that the active site is on the opposite side of the sheet to that seen in all other known Rossmann type NAD⁺-binding proteins [26], comprising a $\beta\alpha\beta$ unit and a $\beta\alpha\beta\alpha\beta$ unit. Strands $\beta 7$

Table 2
Analysis of the Ramachandran plot and rmsd from ideal geometry for the templates and models of *X. fastidiosa*

Enzymes	Region of the Ramachandran plot				rmsd	
	Most favorable (%)	Additional allowed (%)	Generously allowed (%)	Disallowed (%)	Bond lengths (Å)	Bond angles (°)
DAHPS	94.4 (90.5)	4.9 (8.1)	0.7 (1.4)	0.0 (0.0)	0.019 (0.012)	2.92 (1.50)
DHQS	91.2 (92.5)	7.2 (7.5)	1.3 (0.0)	0.3 (0.0)	0.020 (0.13)	3.70 (2.30)
DHQD	96.8 (92.2)	3.2 (7.0)	0.0 (0.8)	0.0 (0.0)	0.020 (0.019)	4.12 (1.80)
SDH	93.2 (93.3)	5.6 (6.7)	0.9 (0.0)	0.4 (0.0)	0.032 (0.018)	4.51 (1.80)
SK	93.8 (90.5)	5.6 (8.8)	0.6 (0.7)	0.0 (0.0)	0.020 (0.005)	3.63 (1.20)
	(93.4)	(5.9)	(0.7)	(0.0)	(0.017)	(1.90)
	(95.6)	(4.4)	(0.0)	(0.0)	(0.013)	(1.70)
EPSPS	89.0 (90.8)	7.1 (8.9)	2.5 (0.3)	1.4 (0.0)		
CS	90.7 (93.1)	7.3 (6.9)	1.7 (0.0)	0.3 (0.0)	0.022 (0.013)	3.92 (2.58)

In parentheses are values obtained for the templates used for modeling.

Table 3
Analysis of the 3D profile and G-factor for the templates and models *X. fastidiosa*

Enzymes	3D profile ^a			G-factor ^b	
	Total score	Ideal score	S_{ideal} score	Torsion angles	Covalent geometry
DAHPS	135.23 (153.33)	159.94 (154.87)	0.845S (0.990S)	0.12 (0.18)	-0.10 (0.45)
DHQS	150.27 (180.65)	169.15 (173.76)	0.888S (1.039S)	-0.05 (0.07)	-0.36 (-0.02)
DHQD	58.79 (67.47)	67.17 (67.62)	0.875S (0.998S)	-0.03 (0.04)	-0.46 (-0.17)
SDH	125.98 (131.02)	125.42 (123.58)	1.004S (1.060S)	-0.12 (-0.05)	-0.83 (0.31)
SK	69.16 (82.02)	81.82 (74.95)	0.854S (1.094S)	0.02 (0.30)	-0.21 (0.57)
	(83.33)	(74.95)	(1.112S)	(0.08)	(0.23)
	(68.25)	(71.74)	(0.953S)	(-0.01)	(-0.51)
EPSPS	175.97 (230.44)	196.36 (195.44)	0.896S (1.179S)	-0.21 (0.08)	-0.87 (0.31)
CS	145.42 (163.31)	170.07 (171.92)	0.855S (0.949S)	-0.16 (0.00)	-0.73 (-0.16)

In parentheses are values obtained for the templates used for modeling.

^a Total score: sum of the 3D–1D scores (statistical preferences) of each residue present in protein. Ideal score: $S_{ideal} = \exp(-0.83 + 1.008 \times \ln(L))$; where L is number of amino acids S_{ideal} . Score: compatibility of the sequence with their 3D structure. It is obtained total score/ideal score. S_{ideal} score above 0.45 S_{ideal} .

^b Ideally, scores should be above -0.5. Values below -1.0 may need investigation.

and $\beta 8$ form a separated β -hairpin. Both DHQS and the classic Rossmann-fold domains bind NAD(P)^+ at the C-terminal end of the sheet [27,28].

However, uniquely among known NAD(P)^+ -binding protein structures, the phosphate moieties of NAD^+ in DHQS are associated with the glycine-rich first turn of the second $\beta\alpha\beta$ unit, rather than the first ($\beta\alpha$) $\beta\alpha\beta$ unit [26–28].

We observed a total of 16 hydrogen bonds in the binary complex $X/DHQS:\text{NAD}^+:\text{CRB}$, involving the residues Asp143, Lys149, Asn159, Glu191, Lys194, Lys233, Arg247, Asn251, His254, His276, and Lys335. For the $EnDHQS:\text{NAD}^+:\text{CRB}$ 16 hydrogen bonds involving the residues Asp146, Lys152, Asn162, Glu194, Lys197, Lys250, Arg264, Asn268, His271, His287, and Lys356 were observed. The contact areas for the complexes of $EnDHQS:\text{NAD}^+:\text{CRB}$ and $X/DHQS:\text{NAD}^+:\text{CRB}$ are 582 and 566 Å², respectively.

3-Dehydroquinase dehydratase

The third step of the shikimate pathway, dehydration of DHQ to give 3-dehydroshikimate (DHS), is catalyzed by DHQD that exists in two forms: types I and II [5]. In

higher plants, the third and fourth steps of the shikimate pathway are catalyzed by a bifunctional enzyme [25]. This enzyme belongs to the superfamily of NAD(P)H -dependent oxidoreductases, which function in anabolic and catabolic enzyme pathways as well as in xenobiotic detoxification. This superfamily is usually subdivided into several families, including short chain dehydrogenases, medium chain dehydrogenases, aldo-keto reductases, iron-activated alcohol dehydrogenases, and long chain dehydrogenases [29].

The overall topology of the type I DHQD is an eight-stranded α/β barrel while the type II DHQD subunit consists of a five-stranded β -sheet core flanked by four α -helices. The type I enzymes catalyze a *cis* (*syn*)-dehydration involving loss of the 2-pro-*R* hydrogen via a covalent imine intermediate, while the type II enzymes catalyze a *trans* (*anti*)-dehydration which results in the loss of the 2-pro-*S* hydrogen, probably via an enolate intermediate [3].

Fig. 3C shows the secondary structure of $X/DHQD$ in complex with 2,3-anhydro-quinic acid (FA1). The enzyme is an α/β protein with a central five-stranded parallel β -sheet whose observed strand order is 21345, which is a flavodoxin-like fold, as defined in the SCOP

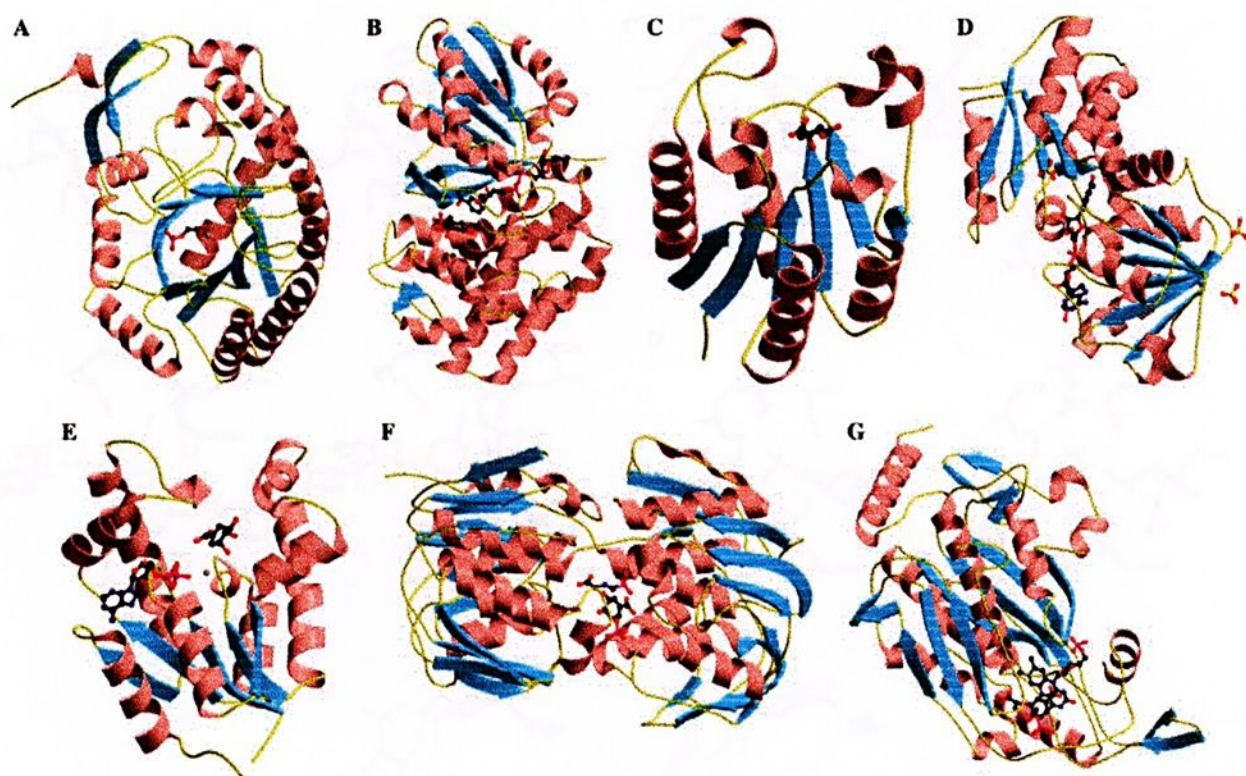


Fig. 3. Ribbon diagrams for: (A) *X*/DAHPS:PGL:Mn²⁺, (B) *X*/DHQS:CRB:NAD⁺, (C) *X*/DHQD:FA1, (D) *X*/SDH:NAD⁺, (E) *X*/SK:ATP:Mg²⁺:SHK, (F) *X*/EPSPS:glyphosate:S3P, and (G) *X*/CS:FMN:EPSP. Figures were generated by Molscript [55] and Raster3d [56].

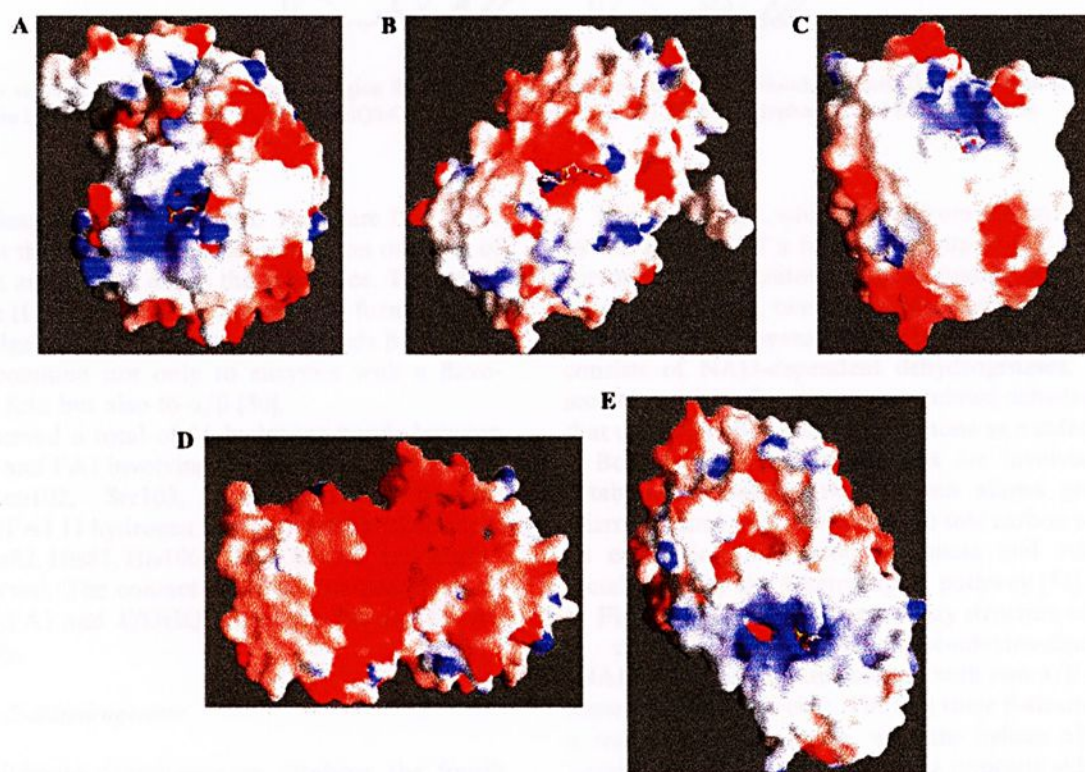


Fig. 4. Electrostatic potential surface for the complexes: (A) *X*/DAHPS:PGL, (B) *X*/DHQS:CRB, (C) *X*/DHQD:FA1, (D) *X*/EPSPS:glyphosate, and (E) *X*/CS:FMN calculated with GRASP [53].

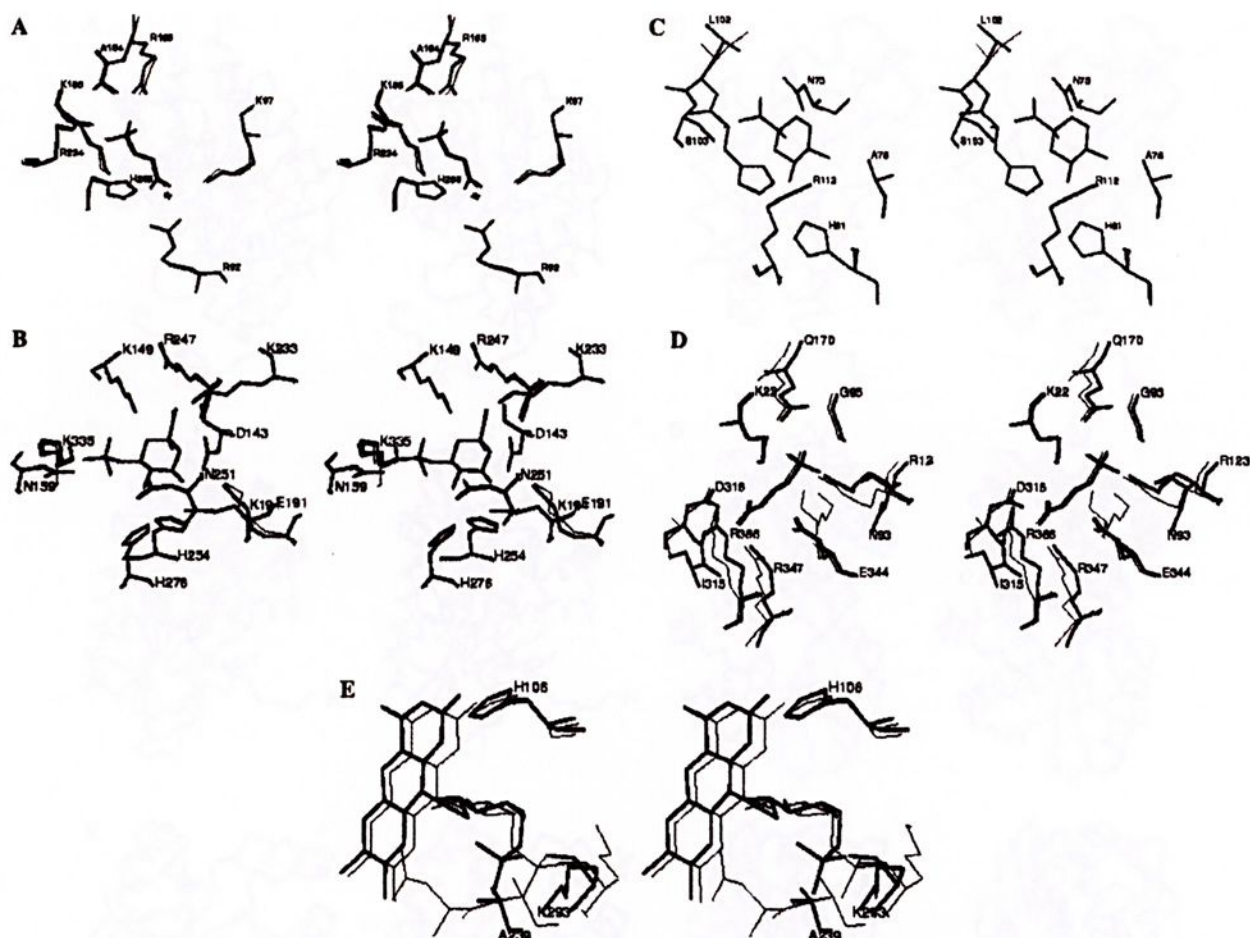


Fig. 5. Stereo views of the interdomain interface region illustrating the binding pockets of the molecular models (thick line) with their template structures (thin line). (A) *X/DAHPS:PGL*, (B) *X/DHQS:CRB*, (C) *X/DHQD:FA1*, (D) *X/EPSPS:glyphosate*, and (E) *X/CS:FMN*.

[31] classification of protein folds. There are four main α -helices in the structure, with $\alpha 1$ and $\alpha 4$ on one face of the β -sheet and $\alpha 2$ and $\alpha 3$ on the other face. The active site in type II DHQD is located in the cleft formed at the carboxy edge of the β -sheet between strands $\beta 1$ and $\beta 3$, which is common not only to enzymes with a flavodoxin-like fold but also to α/β [30].

We observed a total of 11 hydrogen bonds between *X/DHQD* and FA1 involving the residues Asn75, Ala78, His81, Leu102, Ser103, and Arg112. For the *ScDHQD:FA1* 11 hydrogen bonds involving the residues Asn79, Ala82, His85, His106, Ile107, Ser108, and Arg117 were observed. The contact areas for the complexes of *ScDHQD:FA1* and *X/DHQD:FA1* are 105 and 144 Å², respectively.

Shikimate-5-dehydrogenase

The shikimate dehydrogenase catalyzes the fourth reaction in the shikimate pathway, the NADP-dependent reduction of 3-dehydroshikimate to shikimate [5].

This compound, which differs from shikimic acid only by the addition of a hydroxyl group at C1, is the precursor to the ubiquitous plant secondary product chlorogenate. To date, two independent families of quinate/shikimate dehydrogenases have been identified. The first consists of NAD-dependent dehydrogenases, and the second consists of membrane-associated dehydrogenases that utilize pyrrolo-quinoline-quinone as a cofactor [32].

Both types of dehydrogenases are involved in the catabolic quinate pathway, which allows growth of microorganisms with quinate as the sole carbon source by its conversion into protocatechuate and subsequent metabolism by the β -ketoadipate pathway [32].

Fig. 3D shows that the secondary structure of *X/SDH* in complex with nicotinamide-adenine-dinucleotide (NAD⁺) displays an architecture with two α/β domains separated by a wide cleft. The first three β -strands follow a regular α/β succession, with the helices $\alpha 1$ and $\alpha 2$ parallel to the β -strands, flanking opposite sides of the sheet. The domain is completed by a C-terminal α -helical hairpin ($\alpha 9$ and $\alpha 10$), which packs against the

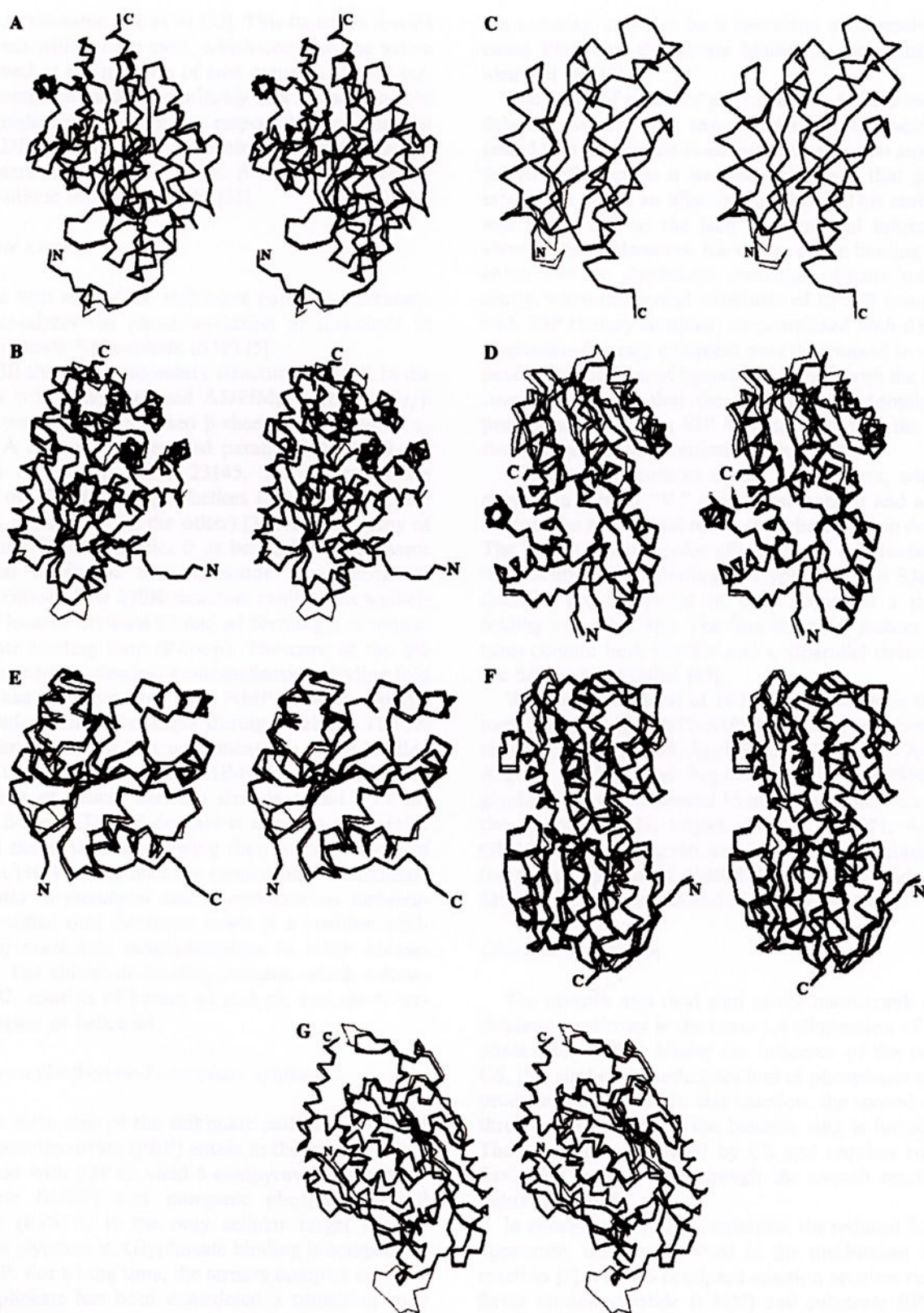


Fig. 6. Superposition of the molecular models (thick line) with their template structures (thin line). (A) X/DAHPS, (B) X/DHQS, (C) X/DHQD, (D) X/SDH, (E) X/SHK, (F) X/EPSPS, and (G) X/CS.

β -sheet on the same side as $\alpha 1$ [32]. This structure reveals an enzyme with a deep cleft, which contains the active site formed at the junction of two domains. The C-terminal domain is easily recognizable as a Rossmann-fold dinucleotide-binding domain, responsible for binding the NADP cofactor. The N-terminal substrate binding and dimerization domain, an α - β - α sandwich, represents a unique topological fold [33].

Shikimate kinase

In the fifth step of the shikimate pathway, shikimate kinase catalyzes the phosphorylation of shikimate to yield shikimate 3-phosphate (S3P) [5].

Fig. 3E shows the secondary structure of *XfSK* in the complex with shikimate and ADP/Mg²⁺. It is an α/β protein consisting of a mixed β -sheet surrounded by α -helices. A central five-stranded parallel β -sheet ($\beta 1$ - $\beta 5$) presents the strand order 23145. The β -strands are flanked on either side by α -helices ($\alpha 1$ and $\alpha 8$ on one side, $\alpha 4$, $\alpha 5$, and $\alpha 7$ on the other) [34]. The ordering of the strands 23145 classifies it as belonging to the same structural family as the nucleoside monophosphate (NMP) kinases. The *XfSK* structure exhibits the walker-A motif located between $\beta 1$ and $\alpha 1$ forming a canonical phosphate-binding loop (P-loop). The core of the SK structure forms a classical mononucleoside-binding fold [35]. It has been reported that NMP kinases undergo large conformational changes during catalysis. The regions responsible for this movement are NMP-binding site and the lid domain. The NMP-binding site is formed by a series of helices between strands 1 and 2 of the parallel β -sheet. The lid domain is a region of variable size and the structure following the fourth β -strand of the sheet [36,37]. The docking simulations of shikimate to *Erwinia chrysanthemi* and *Mycobacterium tuberculosis* identified that shikimate binds in a position analogous to nucleotide monophosphate in NMP kinases [38–40]. The shikimate-binding domain, which follows strand $\beta 2$, consists of helices $\alpha 2$ and $\alpha 3$, and the N-terminal region of helice $\alpha 4$.

5-Enolpyruvylshikimate-3-phosphate synthase

In the sixth step of the shikimate pathway, a second phosphoenolpyruvate (PEP) enters in the pathway. It is condensed with S3P to yield 5-enolpyruvylshikimate-3-phosphate (EPSP) and inorganic phosphate. EPSP synthase (EPSPS) is the only cellular target for the herbicide glyphosate. Glyphosate binding is competitive with PEP. For a long time, the ternary complex enzyme-S3P-glyphosate has been considered a transition state analog in which glyphosate takes the place of PEP [5]. Enzyme-ligand distances within the glyphosate-containing ternary complex were measured by NMR [41], and now reports from several laboratories indicate that

the complex may not be a transition state analog, because PEP and glyphosate binding is apparently not identical [42,43].

The mode of action of glyphosate on EPSPS has been debated over the past two decades. While studies suggested that glyphosate is an intermediate-state analog of the EPSPS reaction it was later proposed that glyphosate might act as an allosteric inhibitor. This ambiguity was partly due to the lack of structural information about EPSPS. However, the nature of the binding site of either S3P or glyphosate remained obscure until recently, when the crystal structures of EPSPS complexed with S3P (binary complex) or complexed with S3P and glyphosate (ternary complex) were determined to atomic detail. Comparison of ligand-free EPSPS with the binary complex suggests that the two globular domains approach each other on S3P binding, such that the active site emerges in the interdomain cleft [44].

The two flat surfaces of the hemispheres, which in projection form a "V," are almost normal and accommodate the N-terminal of the six helices in each domain. The helical macrodipolar effects create a potential well that facilitates the binding of glyphosate and S3P. The domains are composed of three copies of a $\beta\alpha\beta\alpha\beta\beta$ folding unit (Fig. 3F). The four-stranded β -sheet structures contain both parallel and antiparallel strands and the helices are parallel [45].

We observed a total of 16 hydrogen bonds in the ternary complex *XfEPSPS*:S3P:glyphosate, involving the residues Lys22, Asn93, Arg123, Gln170, Ile315, Asp316, Arg347, Glu344, and Arg388. For the *EcEPSPS*:S3P:glyphosate it was observed 15 hydrogen bonds involving the residues Lys22, Gly96, Arg124, Gln171, Asp313, Glu341, Arg344, Arg386, and Lys411. The contact areas for the complexes of *EcEPSPS*:glyphosate and *XfEPSPS*:glyphosate are 186 and 211 Å², respectively.

Chorismate synthase

The seventh and final step in the main trunk of the shikimate pathway is the *trans*-1,4 elimination of phosphate from EPSP. Under the influence of the enzyme CS, this enol ether undergoes loss of phosphoric acid to produce chorismate. In this reaction, the second of the three double bonds of the benzene ring is introduced. The reaction is catalyzed by CS and requires reduced flavin for activity even though the overall reaction is redox neutral [5].

In chorismate synthase catalysis, the reduced flavin is apparently directly involved in the mechanism of the reaction [5]. The CS catalyzed reaction requires reduced flavin mononucleotide (FMN) and substrate EPSP to produce chorismate, a phosphate molecule, and reduced FMN. It is an unusual reaction in that the redox state of the functional cofactor FMN remains unchanged. Another unusual feature of the CS catalyzed reaction is the

nonconcerted anti-1,4-elimination of the 3-phosphate and the C-(6-*pro-R*) hydrogen 4. It is expected that the three-dimensional structure of CS can significantly enhance our understanding of its enzymatic mechanism for this unusual chemistry [46,47].

Studies on CS from bacteria, fungi, and plants revealed that in these organisms the reduced FMN cofactor is made available in different ways to CS: chorismate synthases in fungi—in contrast to those in bacteria and plants—carry a second enzymatic activity which enables them to reduce FMN at the expense of NADPH [48].

The Fig. 3F shows the secondary structure of *Xf*CS in the complex with EPSP and FMN. *Xf*CS has been predicted to be an α - β barrel as a result, of secondary structure prediction efforts, is clearly very different from what was expected. It comprises a single large core domain, which is surrounded by loops and discrete stretches of secondary structure. The core consists of an internal two long α -helices, sandwiched between a pair of four-stranded antiparallel β -sheets [49].

We observed a total of four hydrogen bonds between the ternary complex *Xf*CS:EPSP:FMN, involving the residues His106, Ala239, and Lys293. For the *Sp*CS:EPSP:FMN 6 hydrogen bonds involving the residues His110, Asn251, Ala252, and Lys311 were observed. The contact areas for the complexes of *Sp*CS:EPSP:FMN and *Xf*CS:EPSP:FMN are 542 and 476 Å², respectively.

Conclusions

The specificity and affinity between enzyme and its inhibitor depend on directional hydrogen bounds and ionic interactions, as well as on shape complementarity of the contact surfaces of both partners [50–52]. The electrostatic potential surface of the all models was calculated with GRASP [53] and shown in Figs. 4A–E. Analysis of these structural models indicates that inhibitors such as glyphosate for EPSPS and FAI for DHQD are strongly bound to the enzymes. All complexes show good shape and charge complementarity and exhibit a higher number of intermolecular hydrogen bonds, which strongly indicates that these inhibitors could also inhibit *X. fastidiosa* enzymes. Especially interesting is the *Xf*EPSPS:glyphosate:S3P complex that shows 16 intermolecular hydrogen bonds, more than that observed for *Escherichia coli* enzyme (14 bonds), and a higher contact area, which indicates that glyphosate may have higher affinity for *Xf*EPSPS than that observed for *Ec*EPSPS. Further inhibitor analysis may confirm this prediction.

Acknowledgments

This work was supported by grants from FAPESP (SMOLBNet 01/07532-0, 02/04383-7), CNPq, CAPES and Instituto do Milênio

(CNPq-MCT). WFA (CNPq, 300851/98-7), and MSP (CNPq, 300337/2003-5) are researchers for the Brazilian Council for Scientific and Technological Development.

References

- [1] A.J.G. Simpson et al., The genome sequence of the plant pathogen *Xylella fastidiosa*, *Nature* 406 (2000) 151–157.
- [2] M. Bevan, The bugs from Brazil, *Nature* 406 (2000) 140–141.
- [3] J.R. Coggins, C. Abell, L.B. Evans, M. Frederickson, D.A. Robinson, A.W. Roszak, A.P. Laphorn, Experiences with the shikimate-pathway enzymes as targets for rational drug design, *Biochem. Soc. Trans.* 31 (2003) 548–552.
- [4] S.A. Campbell, T.A. Richards, E.J. Mui, B.V. Samuel, J.R. Coggins, R. McLeod, C.W. Roberts, A complete shikimate pathway in *Toxoplasma gondii* an accident eukaryotic innovation, *Int. J. Parasitol.* 34 (2004) 5–13.
- [5] K.M. Herman, L.M. Weaver, The shikimate pathway, *Annu. Rev. Plant Mol. Biol.* 50 (3) (1999) 473–503.
- [6] F. Dosselaere, J. Vanderleyden, A metabolic node in action: chorismate-utilizing enzymes in microorganisms, *Crit. Rev. Microbiol.* 27 (2) (2001) 75–131.
- [7] K.M. Herrmann, The shikimate pathway as an entry to aromatic secondary metabolism, *Plant Physiol.* 107 (1995) 7–12.
- [8] R.T. Kroemer, S.W. Doughty, A.J. Robinson, W.G. Richards, Prediction of the three-dimensional structure of human interleukin-7 by homology modeling, *Protein Eng.* 9 (6) (1996) 493–498.
- [9] T.L. Blundell, B.L. Sibanda, M.J. Sternberg, J.M. Thornton, Knowledge-based prediction of protein structures and the design of novel molecules, *Nature* 326 (6111) (1987) 347–352.
- [10] T.L. Blundell, D. Carney, S. Gardner, F. Hayes, B. Howlin, T. Hubbard, J. Overington, D.A. Singh, B.L. Sibanda, M. Sutcliffe, 18th Krebs, Hans lecture knowledge-based protein modeling and design, *Eur. J. Biochem.* 172 (3) (1988) 513–520.
- [11] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779–815.
- [12] A. Sali, J.P. Overington, Derivation of rules for comparative protein modeling from a database of protein structure alignments, *Protein Sci.* 3 (9) (1994) 1582–1596.
- [13] A. Sali, L. Potterton, F. Yuan, H. van Vlijmen, M. Karplus, Evaluation of comparative protein modeling by MODELLER, *Proteins* 23 (3) (1995) 318–326.
- [14] A. Sali, Modeling mutations and homologous proteins, *Curr. Opin. Biotechnol.* 6 (4) (1995) 437–451.
- [15] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy minimization and dynamics calculations, *J. Comp. Chem.* 4 (1983) 187–217.
- [16] W. Braun, N. Go, Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm, *J. Mol. Biol.* 186 (3) (1985) 611–626.
- [17] R.A. Laskowski, M.W. MacArthur, D.K. Smith, D.T. Jones, E.G. Hutchinson, A.L. Morris, D. Naylor, D.S. Moss, J.M. Thornton, PROCHECK v.3.0 Program to check the stereochemistry quality of Protein structures Operating instructions (1994).
- [18] Collaborative Computational Project No. 4, The CCP4 suite: programs for proteins crystallography, *Acta Crystallogr. D* 50 (1994) 760–763.
- [19] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, The Xplor-NIH NMR molecular structure determination package, *J. Magn. Res.* 160 (2003) 66–74.
- [20] A.T. Brünger, X-PLOR Version 3.1: A System for Crystallography and NMR, Yale University Press, New Haven, 1992.

- [21] J.U. Bowie, R. Luthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, *Science* 253 (1991) 164–170.
- [22] R. Luthy, J. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, *Nature* 356 (1992) 83–85.
- [23] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22 (12) (1983) 2577–2637.
- [24] T. Wagner, I.A. Shumilin, R. Bauerle, R.H. Kretsinger, Structure of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*: comparison of the Mn^{2+} -2-phosphoglycolate and the Pb^{2+} -2-phosphoenolpyruvate complexes and implications for catalysis, *J. Mol. Biol.* 301 (2000) 389–399.
- [25] K.A. Herrmann, The shikimate pathway: early steps in the biosynthesis of aromatic compounds, *Plant Cell* 7 (1995) 907–919.
- [26] E.P. Carpenter, A.R. Hawkins, J.W. Frost, K.A. Brown, Structure of dehydroquinase synthase reveals an active site capable of multistep catalysis, *Nature* 394 (6690) (1998) 299–302.
- [27] M.G. Rossmann, D. Moras, K.W. Olsen, Chemical and biological evolution of a nucleotide binding protein, *Nature* 250 (1974) 194–199.
- [28] A.M. Lesk, NAD-binding domains of dehydrogenases, *Curr. Opin. Struct. Biol.* 5 (1995) 775–783.
- [29] J. Benach, I. Lee, W. Edstrom, A.P. Kuzin, Y. Chiang, T.B. Actons, G.T. Monteliones, J.F. Hunt, The 2.3 Å crystal of the shikimate 5-dehydrogenase orthologue YdiB from *E. coli* suggests a novel catalytic environment for an NAD-dependent dehydrogenase, *J. Biol. Chem.* 278 (2003) 19176–19182.
- [30] A.W. Roszak, D.A. Robinson, T. Krell, I.S. Hunter, M. Fredrickson, C. Abell, J.R. Coggins, A.J. Laphorn, The structure and mechanism of the type II dehydroquinase from *Streptomyces coelicolor*, *Structure (Camb.)* 10 (4) (2002) 493–503.
- [31] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.* 247 (1995) 536–540.
- [32] G. Michel, A.W. Roszak, V. Sauvé, J. Maclean, A. Matte, J.R. Coggins, M. Cygler, A.J. Laphorn, Structures of Shikimate dehydrogenase AroE and its paralog YdiB: a common structural framework for different activities, *J. Biol. Chem.* 278 (21) (2003) 19463–19472.
- [33] E. Vogan, Shikimate dehydrogenase structure reveals novel fold, *Structure* 11 (2003) 902–903.
- [34] G.E. Schulz, Binding of nucleotides by proteins, *Curr. Opin. Struct. Biol.* 2 (1992) 61–67.
- [35] Y. Matsuo, K. Nishikawa, Protein structural similarities predicted by a sequence–structure compatibility method, *Protein Sci.* 3 (11) (1994) 2055–2063.
- [36] C.W. Müller, G.J. Schlauender, J. Reinstein, G.E. Schulz, Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding, *Structure* 4 (1996) 147–156.
- [37] M. Gerstein, G. Schulz, C. Chothia, Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers, *J. Mol. Biol.* 229 (2) (1993) 494–501.
- [38] Y. Gu, L. Reshetnikova, Y. Li, Y. Wu, H. Yan, S. Singh, X. Ji, Crystal structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the dynamic role of the LID domain in catalysis, *J. Mol. Biol.* 319 (3) (2002) 779–789.
- [39] J.H. Pereira, J.S. de Oliveira, F. Canduri, M.V.B. Dias, M.S. Palma, L.A. Basso, W.F. de Azevedo Jr., D.S. Santos, X-ray crystallographic studies of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimate acid, *Nat. Struct. Biol.* (submitted).
- [40] T. Krell, J.R. Coggins, A.J. Laphorn, The three-dimensional structure of shikimate kinase, *J. Mol. Biol.* 178 (5) (1998) 983–997.
- [41] L.M. McDowell, A. Schmidt, E.R. Cohen, D.R. Studelska, J. Schaefer, Structural constraints on the ternary complex of 5-enolpyruvyl shikimate 3-phosphate synthase from rotational-echo double resonance NMR, *J. Mol. Biol.* 256 (1996) 160–171.
- [42] R.D. Sammons, K.J. Gruys, K.S. Anderson, K.A. Johnson, J.A. Sikorski, Reevaluating glyphosate as a transition state inhibitor of EPSP synthase: identification of an EPSP synthase–EPSP–glyphosate ternary complex, *Biochemistry* 34 (1995) 6433–6440.
- [43] L.M. McDowell, C.A. Klug, D.D. Beusen, J. Schaefer, Ligand geometry of the ternary complex of 5-enolpyruvyl shikimate 3-phosphate synthase from rotational-echo double-resonance NMR, *Biochemistry* 35 (1995) 5395–5403.
- [44] S. Eschenburg, M.L. Healy, M.A. Priestman, G.H. Lushington, E. Schönbrunn, How the mutation glycine96 to alanine confers glyphosate insensitivity to 5-enolpyruvyl shikimate-3-phosphate synthase from *Escherichia coli*, *Planta* 216 (2002) 129–135.
- [45] J.H. Pereira, F. Canduri, J.S. de Oliveira, N.J. da Silveira, L.A. Basso, M.S. Palma, W.F. de Azevedo Jr., D.S. Santos, Structural bioinformatics study of EPSP synthase from *Mycobacterium tuberculosis*, *Biochem. Biophys. Res. Commun.* 312 (3) (2003) 608–614.
- [46] E. Schonbrunn, S. Eschenburg, W.A. Shuttleworth, J.V. Schloss, N. Amrhein, J.N. Evans, W. Kabsch, Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail, *Proc. Natl. Acad. Sci. USA* 98 (4) (2001) 1376–1380.
- [47] C.M. Viola, V. Saridakis, D. Christendat, Crystal structure of chorismate synthase from *Aufex aeolicus* reveals a novel beta alpha beta sandwich topology, *Proteins* 54 (2004) 166–169.
- [48] P. Macheroux, J. Schmid, N. Amrhein, A. Schaller, A unique reaction in a common pathway: mechanism and function of chorismate synthase in the shikimate pathway, *Planta* 207 (1999) 325–334.
- [49] J. Maclean, S. Ali, The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction, *Structure* 11 (2003) 1499–1511.
- [50] W.F. De Azevedo, F. Canduri, N.J.F. da Silveira, Structural basis for inhibition of cyclin-dependent kinase 9 by flavopiridol, *Biochem. Biophys. Res. Commun.* 293 (2002) 566–571.
- [51] W.F. De Azevedo, R.T. Gaspar, F. Canduri, J.C. Camera, N.J.F. da Silveira, Molecular model of cyclin-dependent kinase 5 complexed with roscovitine, *Biochem. Biophys. Res. Commun.* 297 (2002) 1154–1158.
- [52] W.F. De Azevedo, J.S. de Oliveira, L.A. Basso, M.S. Palma, J.H. Pereira, F. Canduri, D.S. Santos, Molecular model of shikimate kinase from *Mycobacterium tuberculosis*, *Biochem. Biophys. Res. Commun.* 295 (1) (2002) 142–148.
- [53] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins* 11 (4) (1991) 281–296.
- [54] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [55] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of proteins, *J. Appl. Crystallogr.* 24 (1991) 946–950.
- [56] E.A. Merritt, D.J. Bacon, Raster3D: photorealistic molecular graphics, *Methods Enzymol.* 277 (1997) 505–524.

