

Bioinformática Estrutural Aplicada ao Estudo de Enzimas da Via Metabólica do Ácido Chiquímico

Helen Andrade Arcuri

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

Orientador: Prof. Dr. Mário Sérgio Palma

Co – Orientador: Prof. Dr. Walter Filgueira de Azevedo Júnior

São José do Rio Preto
Estado de São Paulo – Brasil
Abril – 2008

Arcuri, Helen Andrade.

Bioinformática estrutural aplicada ao estudo de enzimas da via metabólica do ácido chiquímico/ Helen Andrade Arcuri – São José do Rio Preto : [s.n.], 2008

133 f. : il. ; 30 cm.

Orientador: Prof. Dr. Mário Sérgio Palma

Co-Orientador: Prof. Dr. Walter Filgueira de Azevedo Junior

Tese (doutorado) – 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

A pedra

O distraído nela tropeçou...
O bruto a usou como projétil...
O empreendedor, usando-a, construiu...
O camponês, cansado da vida, dela fez assento...
Para meninos, foi brinquedo...
Drummond a poetizou...
Já, David matou Golias...
Michelangelo extraiu-lhe a mais bela escultura...

O mais interessante em todos esses casos, a diferença não estava na pedra, mas sim no homem!

Não existe “pedra” no seu caminho que você não possa aproveitá-la para o seu próprio **crescimento**.

A questão é o que fazemos com as pedras que aparecem todos os dias em nossas vidas.

Autor Desconhecido

Agradecimentos

Não encontrei e jamais encontrarei palavras que descrevam suficientemente o quão grata me sinto por aqueles que colaboraram, ao longo desta etapa da minha vida, para alcançar meus objetivos. O trabalho descrito aqui envolveu muitas pessoas que acreditaram na minha capacidade e me apoiaram nos momentos mais difíceis desta caminhada.

A lista abaixo é uma relação de algumas dessas pessoas queridas em ordem alfabética, que tiveram participação fundamental nessa jornada. Algumas são de fato co-autores – oficiais ou não – desta Tese de Doutorado, por terem ensinado, pipetado, cedido espaço nos seus laboratórios, criticado e dado sugestões úteis para o desenvolvimento deste trabalho.

Há, ainda, aquelas que insistiram para que eu não desistisse nos momentos mais difíceis, naqueles momentos em que nada dá certo, o experimento não funciona, o algoritmo implementado não roda, o cluster dá pane, surgem problemas de ordem prática, de ordem pessoal, saudades, desânimo, dúvidas, stress... e elas estavam ali, repetindo para eu não desistir, não jogar fora o que eu já havia produzido que meu trabalho estava dando frutos, que essas fases difíceis fazem parte do aprendizado e passam.

Há também ainda, aquelas pessoas que assumiram responsabilidades no meio do caminho deste trabalho e que sem suas contribuições valiosas eu não conseguiria resolver os problemas que foram surgindo, apoio que foi fundamental para eu conseguir terminar este trabalho com confiança e tranquilidade. A todas elas, um obrigado verdadeiro, com muito carinho e a certeza de que serão sempre lembradas.

Aos meus familiares, especialmente aos meus pais, minha avó Maria e meu avô Toninho, meu tio Tato e sua esposa Janete, minha irmã Michele e seu namorado Daniel, minha tia Antônia, meus padrinhos Vilma e Yezo e minha segunda mãe Dona Maria, um obrigado com muito amor e carinho, pelo incentivo, apoio, paciência e companheirismo, sem vocês mais esta vitória em minha vida não seria possível.

A Deus por sempre me iluminar e acompanhar o meu caminho, colocando as pessoas certas nos momentos difíceis.

Aos amigos Angélica, Alessandra e família (seu marido Ronaldo e seu filho Miguel David), Alessandra (Rio Claro), Barbosa, Bibiana, Bona, Brianna, Dona Odete, Eduardo, Evandro, Cristiane, Daiana, Dayane, Fernanda, Geraldo, Ilva, Isa, José Carlos e sua esposa Marilene, José Renato, Júlio César, Leka, Lílian, Lisandra, Livia, Luci, Luciana, Mariane, Mário, Nelson José, Magnon, Michele, Paola, Paula, Paulinho, Rodolfo, Rose, Rosemar, Silvia.

Colegas do Laboratório de Sistemas Biomoleculares e do departamento de física da UNESP

Funcionários do departamento de física da UNESP.

Fundação de Amparo e Pesquisa do Estado de São Paulo – FAPESP pela bolsa de doutorado.

Profa. Dra. Claudia Regina Bonini Rodrigues

Prof. Dr. Elso Drigo Filho

Profa. Dra. Fátima Pereira de Souza

Prof. Dr. Gustavo Bonila

Prof. Dr. Jorge Chahine

Prof. Dr. José Marcio Machado

Prof. Dr. José Ramon Beltan Abrego

Prof. Dr. Marcelo Andres Fossey

Prof. Dr. Mario Sérgio Palma

Profa. Dra. Paula Rahal

Prof. Dr. Walter Filgueira de Azevedo Junior

Programa de Pós-graduação em Biofísica Molecular da UNESP.

Sumário

Agradecimentos	ii
Sumário	iv
Lista de Figuras	vi
Lista de Tabelas	ix
Lista de Abreviações e Símbolos	x
Resumo.....	xiii
Abstract	xiv
1. Introdução	- 1 -
1.1. A tuberculose.....	- 3 -
1.2. Via metabólica do ácido chiquímico	- 6 -
1.2.1. Descrição detalhada do mecanismo de reação das enzimas pertencentes a via	- 9 -
1.3. Aplicações da bionformática estrutural	- 17 -
2. Objetivos	- 22 -
3. Materias e Métodos.....	- 23 -
3.1. Modelagem Molecular	- 23 -
3.2. Modelagem molecular comparativa ou por homologia.....	- 23 -
3.2.1. Procura e seleção de <i>moldes</i>	- 25 -
3.2.2. Alinhamento molde/alvo	- 27 -
3.2.3. Construção do modelo	- 27 -
3.2.4. Avaliação dos modelos	- 28 -
3.3. Modelagem em larga escala das enzimas da via do ácido chiquímico	- 29 -
3.4. <i>Docking</i> Molecular	- 30 -
3.5. Clonagem, expressão, purificação das enzimas e obtenção dos ligantes	- 31 -
3.5.1. Preparação das amostras e dos ligantes	- 32 -
3.6. Cristalização de proteínas.....	- 34 -
3.6.1. Cristalização da enzima chiquimato quinase	- 36 -
3.6.2. Cristalização da corismato sintase.....	- 37 -
3.7. Coleta e processamento de dados de difração por raios X	- 37 -
3.7.1. Coletas de dados da enzima corismato sintase	- 39 -
3.8. Substituição molecular e Refinamento cristalográfico.....	- 39 -
3.9. Espectroscopia de Dicroísmo circular (CD).....	- 41 -
3.9.1. CD para estimar as frações de estrutura secundária das enzimas.....	- 42 -
3.9.2. Estudos de estabilidade térmica das enzimas por CD	- 43 -
3.10. Espectroscopia de fluorescência.....	- 44 -
3.11. Espalhamento de raios X a baixo ângulo	- 47 -
4. Resultados e Discussão	- 52 -
4.1. Desenvolvimento da ferramenta para modelagem molecular comparativa e análise em larga escala.	- 52 -
4.1.2. Implementação e formatação dos dados que compõe o banco de dados SKPDB	- 52 -
4.2. SKPDB	- 56 -
4.2.1. Implementação e descrição do banco de dados SKPDB	- 56 -
4.2.2 Tabelas e Relacionamentos	- 59 -
4.2.3. Acesso e interface do banco de dados	- 61 -
4.3. Cristalização, difração de raios X e resolução da estrutura.....	- 69 -
4.3.1. Cristalização da enzima chiquimato quinase.....	- 69 -
4.3.2. Cristalização da enzima corismato sintase	- 71 -
4.3.3. Coleta de dados e difração de raios X da enzima corismato sintase	- 72 -
4.3.4. Resolução e refinamento do modelo da estrutura 3D da enzima corismato sintase..	- 73 -

4.3.5. Análise da estrutura 3D	- 75 -
4.4. Experimentos de CD para estimar as frações de estrutura secundária das enzimas.....	- 78 -
4.4.1. Chiquimato quinase	- 79 -
4.4.2. Corismato sintase.....	- 81 -
4.5. Experimentos de CD para verificar a estabilidade das enzimas.....	- 82 -
4.5.1. Chiquimato quinase	- 83 -
4.5.2. Corismato sintase.....	- 89 -
4.6. Experimento de Fluorescência da corismato sintase.....	- 90 -
4.7. Experimento de SAXS da corismato sintase.....	- 93 -
4.8. <i>Docking</i> molecular	- 96 -
4.8.1. Chiquimato sintase	- 97 -
4.8.2. Corismato sintase	- 100 -
5. Conclusões	- 104 -
6. Referências bibliográficas.....	- 106 -

Lista de Figuras

- Figura 1. Distribuição geográfica das taxas de incidência de tuberculose estimadas em 2003 (Nunn *et al.*, 2005). No leste da Europa e África, estão aumentando as mortes após quase 40 anos de declínio. Outro fato importante é a presença da bactéria em níveis alarmantes nos países em desenvolvimento. - 4 -
- Figura 2. Via do ácido chiquímico na seqüência de sete passos metabólicos, iniciando no PEP e E4P que são convertidos em corismato. A descrição detalhada do mecanismo de reação de cada enzima é descrita durante o texto, a figura foi adaptada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>. - 7 -
- Figura 3. Estrutura molecular do glifosato ($C_3H_8NO_5P$). Este composto exemplifica o grande potencial de desenvolvimento de novos inibidores baseados nas enzimas da via metabólica do ácido chiquímico. As coordenadas atômicas do glifosato foram obtidas no PDB (código PDB:1G6S) (SCHÖNBRUNN *et al.*, 2001). - 9 -
- Figura 4. Reação catalisada pela enzima DAHPS, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>. - 10 -
- Figura 5. Reação catalisada pela enzima DHQS, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>. - 11 -
- Figura 6. Reação catalisada pela enzima DHQD, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>. - 12 -
- Figura 7. Reação catalisada pela enzima SD, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>. - 13 -
- Figura 8. Reação catalisada pela enzima SK, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>. - 14 -
- Figura 9. Reação catalisada pela enzima EPSPS, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>. - 15 -
- Figura 10. Reação catalisada pela enzima CS, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>. - 16 -
- Figura 11. Interação das áreas que têm contribuído para a formação e o desenvolvimento da bioinformática (BAYAT, 2002). - 18 -
- Figura 12. Passos da modelagem molecular comparativa adaptado do artigo de Sanchez & Šali, 1997. - 25 -
- Figura 13. Estrutura molecular dos ligantes ADP, Olomoucine e Roscovitine. As coordenadas atômicas dos ligantes foram obtidas no PDB. - 33 -
- Figura 14. Estrutura molecular dos ligantes FMN e AC. As coordenadas atômicas dos ligantes foram obtidas no PDB. - 34 -
- Figura 15. Ilustração esquemática do método da difusão de vapor pela gota suspensa. Figura retirada do texto Nangem, 2003. - 35 -
- Figura 16. Procedimento de crio-cristalização de um cristal de proteína, para coleta de raios X. Figura retirada do texto Nangem, 2003. - 38 -
- Figura 17. A enzima mtCS na presença dos ligantes FMN (co-fator) e EPSP (substrato) do lado direito e a enzima mtCS na forma apo do lado direito. Ambas as estruturas com os aminoácidos aromáticos em destaque na forma de *ball stick*. As coordenadas atômicas da mtCS e dos ligantes foram obtidas no PDB (código de acesso: 2G85, 2O12 e 2QHF). - 46 -
- Figura 18. Espectros de titulação. (A) sal em molar, (B) ácido corísmico em proporção de vezes, (C) FMN em proporção de vezes. - 47 -
- Figura 19. Fluxograma da formatação dos dados e construção do banco de dados SKPDB. - 56 -
- Figura 20. Diagrama esquemático de como a programação CGI interage com o banco de dados SKPDB, organizando e filtrando os dados devolvidos ao usuário. - 58 -
- Figura 21. Diagrama de relacionamento existente entre as tabelas que compõem o banco de dados

SKPDB.....	- 59 -
Figura 22. Página inicial do banco de dados SKPDB.....	- 62 -
Figura 23. Página sobre as enzimas e a via.....	- 63 -
Figura 24. Página com os resultados sobre busca efetuada com a enzima tendo sua estrutura 3D determinada por modelagem molecular comparativa.	- 64 -
Figura 25. Página com os resultados sobre busca efetuada com a enzima tendo sua estrutura 3D determinada experimentalmente.	- 64 -
Figura 26. Página com os resultados sobre busca efetuada com a enzima não tendo sua estrutura 3D determinada experimentalmente ou por modelagem molecular comparativa.	- 65 -
Figura 27. A interface acima representa a página com a lista de resultados sobre busca efetuada através do nome do organismo na base de dados.....	- 66 -
Figura 28. Página da aba “Model Info”, que mostra os dados sobre o modelo gerado.	- 67 -
Figura 29. Página da aba “Template Info” que mostra os dados sobre o molde utilizado na construção da estrutura 3D do modelo.....	- 68 -
Figura 30. Pagina da aba “Model Analysis”, que mostra os resultados sobre a qualidade do modelo. -	68 -
Figura 31. Cristais da enzima mtCQ. (A) Forma dos mini-cristais do pH 7,0 à 7,9 e (B) Forma dos mini-cristais do pH 8,0 à 8,6.	- 69 -
Figura 32. Cristais da enzima mtCS. (A) Na forma apo na condição (5), (B) Na forma apo na condição (4) e (C) Na presença do produto AC na condição (4).	- 72 -
Figura 33. Imagens de difração por raios X do cristal mtCS-AC na condição (5) e mtCS-FMN-AC na condição (4).	- 73 -
Figura 34. Mapa de densidade eletrônica na região de ligação dos ligantes, mostrando a ausência parcial ou total da densidade eletrônica para os referidos ligantes.	- 75 -
Figura 35. Estrutura da mtCS com as moléculas de água (em vermelho), gerada pelo programa Pymol. Colorida de acordo com a estrutura secundária: α -hélices (verde) e fitas- β (roxa) voltas e alças (rosa).	- 76 -
Figura 36. Mapa de densidade eletrônica para uma das estruturas de mtCS.	- 77 -
Figura 37. Gráfico de Ramachandran com a posição de cada resíduo de aminoácido para um dos modelos finais da estrutura 3D da mtCS.	- 78 -
Figura 38. Sobreposição dos espectros de dicroísmo circular da mtCQ sem e na presença de ligantes	- 80 -
Figura 39. Sobreposição dos espectros de dicroísmo circular da mtCS sem e na presença de ligantes.	- 82 -
Figura 40. Espectros de CD da mtCQ na forma apo em diferentes temperaturas	- 84 -
Figura 41. Espectros de CD da mtCQ em complexo com <i>olomoucine</i> em diferentes temperaturas	84 -
Figura 42. Espectros de CD da mtCQ em complexo com <i>olomoucine</i> em diferentes temperaturas	85 -
Figura 43. Transição da desnaturação térmica da mtCQ na forma apo	- 86 -
Figura 44. Transição da desnaturação térmica da mtCQ na presença do ligante <i>olomoucine</i>	- 86 -
Figura 45. Transição da desnaturação térmica da mtCQ na presença do ligante <i>roscovitine</i>	- 87 -
Figura 46. Transição da desnaturação térmica da mtCQ na presença da coenzima ADP	- 88 -
Figura 47. Transição da desnaturação térmica de mtCS na forma apo	- 89 -
Figura 48. Transição de desnaturação térmica da mtCS na presença dos ligantes FMN e AC	- 90 -
Figura 49. Espectros de fluorescência da mtCS sem e na presença dos ligantes FMN e CA.....	- 92 -
Figura 50. Gráfico de $\log I(q)$ versus q^2	- 94 -
Figura 51. Curva de espalhamento da mtCS corrigida a feixe pontual e extrapolada a concentração e a intensidade zero.....	- 95 -
Figura 52. Função de distribuição de distância $p(r)$, com a máxima dimensão de 10nm.	- 95 -

- Figura 53. Modelo *ab initio* e sua sobreposição com a estrutura cristalográfica da mtCS na forma apo. (A) Modelo *ab initio* obtido por SAXS, (B) Modelo tetramérico das coordenadas cristalográficas geradas pelo CCP4 e (C) Sobreposição dos modelos *ab initio* com as coordenadas cristalográficas..... - 96 -
- Figura 54. Diagrama das hidrofobicidades dos ligantes com mtCQ..... - 98 -
- Figura 55. Diagrama das superfícies eletrostáticas dos ligantes com mtCQ. - 99 -
- Figura 56. Diagrama da hidrofobicidade da região de interação da molécula de mtCS com o FMN. As cores tanto da molécula de mtCS quanto da molécula de FMN indicam: verde: região hidrofóbica, laranja: região hidrofílica e branco: região: neutra. Percebe-se que cada cor do substrato está na região da molécula de mesma cor..... - 100 -
- Figura 57. Figura da superfície eletrostática da região em que o substrato FMN se encaixa na mtCS. Azul: região positiva e vermelho: região negativa. (A) Zoom da superfície eletrostática da predição do sitio ativo do FMN. (B) Superfície eletrostática da interação do substrato FMN com a mtCS..... - 101 -
- Figura 58. Mapa da hidrofobicidade da região de interação da molécula de mtCS com o EPSP, a molécula de EPSP está encaixada na região hidrofóbica. As cores tanto da molécula de mtCS quanto da molécula de EPSP indicam: verde: região hidrofóbica, laranja: região hidrofílica e branco: região neutra..... - 102 -
- Figura 59. Superfície eletrostática do substrato EPSP no possível sitio ativo da mtCS..... - 102 -

Lista de Tabelas

Tabela 1. Parâmetros utilizados na coleta de dados de difração de raios X.	- 39 -
Tabela 2. Parâmetros utilizados na coleta dos espectros de dados de CD para as enzimas mtCQ e mtCS, para estimar as frações de estrutura secundária.	- 42 -
Tabela 3. Parâmetros utilizados na coleta dos espectros de dados de CD para a enzima mtCQ, para avaliar a estabilidade térmica.	- 43 -
Tabela 4. Parâmetros utilizados na coleta dos espectros de dados de CD para as enzimas mtCQ e mtCS, para avaliar a estabilidade térmica.	- 44 -
Tabela 5. Resultado do alinhamento feito para as enzimas encontradas no Swiss-Prot contra o PDB, para encontrar moldes para as enzimas.	- 53 -
Tabela 6. Resultados obtidos na substituição molecular.	- 74 -
Tabela 7. Estatísticas de refinamento para mtCS.	- 74 -
Tabela 8. Porcentagem dos elementos de estrutura secundária calculado para os dados experimentais de CD da mtCQ.	- 79 -
Tabela 9. Porcentagem dos elementos de estrutura secundária calculado para os dados experimentais de CD da mtCS.	- 81 -

Lista de Abreviações e Símbolos

Símbolo dos aminoácidos

Código de 3 letras	Código de 1 letra	Nome do aminoácido
Ala	A	Alanina
Arg	R	Arginina
Asn	N	Asparagina
Asp	D	Aspartato
Cys	C	Cisteína
Gln	Q	Glutamina
Glu	E	Glutamato
Gly	G	Glicina
His	H	Histidina
Iso	I	Isoleucina
Leu	L	Leucina
Lys	K	Lisina
Met	M	Metionina
Phe	F	Fenilalanina
Pro	P	Prolina
Ser	S	Serina
Thr	T	Treonina
Trp	W	Triptofano
Tyr	Y	Tirosina
Val	V	Valina

Abreviaturas

3D – tridimensional

AC – ácido corísmico

ADP – adenosina difosfato

ATP – adenosina trifosfato

CD – dicroísmo circular

CS – Corismato Sintase

DAHPS – 3-deoxi-D-arabino-heptuloseonato 7-fosfato Sintase

DHQS – 3-Desidroquinato Sintase

DHQD – 3-Desidroquinato Desidratase

DMSO – Dimetilsulfóxido

E4P – eritrose 4-fosfato

EC – *Enzyme Classification*

EPSPS – 5-Enoilpiruvilchiquimato 3-fosfato Sintase

FMN – flavina mononucleotídeo

MS – Ministério da Saúde

MVD – *Molegro Virtual Docker*

NAD – nicotinamida adenina dinucleotídeo

OMS – Organização Mundial de Saúde

PDB – *Protein Data Bank*

PEP – fosfoenolpiruvato

RMN – ressonância magnética nuclear

RMSD – desvio médio padrão quadrático

RPM – rotações por minuto

S3P – chiquimato-3-fosfato

SAXS – espalhamento de raios X a baixos ângulos

SD – Chiquimato-5-Desidrogenase

SK – Chiquimato Quinase

SKPDB – *ShiKimate Pathway Database*

Resumo

Esta tese trata da caracterização espectroscópica por dicroísmo circular e fluorescência, caracterização estrutural utilizando as técnicas de cristalografia de raios X e espalhamento de raios X a baixos ângulos (SAXS) de duas (chiquimato quinase e corismato sintase) das sete enzimas da via metabólica do ácido chiquímico na forma apo e em complexo com seus respectivos ligantes naturais e o desenvolvimento de abordagens computacionais para implementação e integração de ferramentas de modelagem molecular comparativa e análise das estruturas tridimensionais geradas em larga escala das enzimas da via do ácido chiquímico de microorganismos e de plantas. A motivação deste trabalho é o fato de que a tuberculose e outras doenças também negligenciadas causadas por microorganismos são a causa de morte de milhões de pessoas no mundo, assim a caracterização estrutural de proteínas alvo para propor novas drogas tornou-se essencial. O principal interesse no estudo da via do ácido chiquímico é o fato que a via não está presente em humanos, o que a torna alvo seletivo para desenho de drogas, diminuindo o impacto das drogas em humanos. Os resultados obtidos a partir das técnicas experimentais utilizadas neste trabalho possibilitaram um melhor entendimento do mecanismo catalítico das enzimas chiquimato quinase e corismato sintase. A partir dos dados de modelagem molecular em larga escala, foi possível montar um banco de dados relacional e curado, o SKPDB, que contém anotações detalhadas sobre a função e estrutura das enzimas, podendo ser acessado em <http://lsbzix.rc.unesp.br/skpdb/index.html>. Este trabalho aumenta a certeza de que a modelagem molecular comparativa em conjunto com técnicas experimentais é uma ferramenta útil e valiosa na anotação de seqüências e no estudo estrutural e funcional de proteínas.

Abstract

This thesis deals with the spectroscopic characterization by fluorescence and circular dichroism, structural characterization using the techniques of X-ray crystallography and X-ray scattering at low angles (SAXS) of two (chorismate synthase and shikimate kinase) of the seven enzymes of the metabolic pathway of chiquimic acid in apo form and in complex with their natural ligands and the development of computational approaches to implementation and integration of tools for molecular modeling and comparative analysis of three-dimensional structures generated in a large scale means of the enzyme chiquimic acid of microorganisms and plants. The motivation of this work is the fact that tuberculosis and other neglected diseases also caused by microorganisms are the cause of death of millions of people around the world, so the structural characterization of proteins offer new targets for drugs has become essential. The main interest in studying the path of chiquimic acid is the fact that the route is not present in humans, which makes selective target for design of drugs, reducing the impact of drugs in humans. The results obtained from the experimental techniques used in this study allowed a better understanding of the catalytic mechanism of shikimate kinase enzymes and chorismate synthase. From the data of molecular modeling in large scale, it can build a relational database and cured, the SKPDB which contains detailed notes on the structure and function of enzymes, which can be accessed in <http://lsbzix.rc.unesp.br/skpdb/index.html>. This work increases the certainty that the comparative molecular modeling together with experimental techniques is a useful and valuable tool in the annotation of sequences and the study of structural and functional proteins.

1. Introdução

A biologia é uma ciência que tem ganhando muito destaque nos dias de hoje, em especial a biologia molecular. O crescimento dos diversos segmentos relacionados à biologia molecular fez com que houvesse a necessidade de se incorporar técnicas computacionais para a resolução de problemas biológicos. Dessa forma surge um novo termo, denominado bioinformática. A bioinformática é um subgrupo do vasto campo da biologia computacional e pode ser sucintamente descrita como a utilização de técnicas computacionais para realizar a aquisição, análise e armazenamento de informações biológicas (BOURNE & WEISSIG, 2003).

Um dos grandes desafios da bioinformática nos dias de hoje é a análise e organização dos dados de genômica e proteômica. A grande quantidade de dados armazenados em bancos de dados públicos ou privados, pode ser utilizada no cruzamento de dados, obtendo inúmeras informações relevantes ao avanço em pesquisas biológicas e tecnológicas. Com isso inúmeros bancos de dados e ferramentas computacionais de acesso público ou privado têm sido criados na tentativa de organizar e permitir acesso eficiente e rápido a estas informações pela *internet* (COLE *et al.*, 1998; BRENNER, 2001; LIU & ROST, 2002; TAYLOR, 2002).

Atualmente, as estruturas 3D das proteínas são obtidas por meio das técnicas de difração de raios X (cristalografia de proteínas) e ressonância magnética nuclear (RMN). Nesse contexto, a área de cristalografia de proteínas é responsável por cerca de 85% de todos os modelos estruturais de proteínas encontrado no *Protein Data Bank* (PDB; BERMAN *et al.*, 2000), mas nem sempre é possível obter as estruturas tridimensionais das proteínas utilizando as técnicas de cristalografia de raios X e RMN devido a suas limitações (BRENNER, 2001).

Para contornar este fator limitante os biólogos estruturais usam a bioinformática para manusear o vasto e complexo conjunto de dados de cristalografia de raios X e RMN para prever modelos de estruturas tridimensionais (3D) de proteínas pelo método de modelagem molecular comparativa, possibilitando a análise em larga escala das estruturas 3D obtidas através do método

(BURLEY *et al.*, 1999).

Espera-se que as estruturas 3D das proteínas, determinadas experimentalmente e os modelos computacionais, produzam avanços no entendimento da função molecular e no mecanismo de ação de milhares de proteínas para podermos identificar alvos moleculares para desenho racional de drogas baseado em estrutura, utilizando como seleção de alvos, vias metabólicas ausentes em humanos, mas presente em microorganismos e plantas. Esses fatos têm permitido um grande avanço da área tanto em instituições de pesquisa científica como em empresas farmacêuticas (BRENNER, 2001; LIU & ROST, 2002; TAYLOR, 2002).

Sob este aspecto, as enzimas da via metabólica do ácido chiquímico representam bons exemplos da utilidade de tal abordagem aplicada aos constituintes enzimáticos de uma rota biossintética essencial para microorganismos e plantas, por ser inexistente no organismo humano (BENTLEY, 1990). Recentemente, esta via metabólica foi identificada também em parasitas do filo apicomplexa (*Plasmodium falciparum*, *Toxoplasma gondii* e *Cryptosporidium parvum*) (ROBERTS *et al.*, 1998). Tal fato, teoricamente, minimizaria o efeito tóxico das drogas antimicrobianas (DAVIES *et al.*, 1994) e antiparasitas (ROBERTS *et al.*, 1998) para a espécie humana.

A motivação deste trabalho foi demonstrar a importância da via metabólica do ácido chiquímico em microorganismos para ser utilizada como alvo seletivo para o desenho racional de drogas baseado em estrutura tornando estas drogas específicas e eficazes contra enfermidades causadas por microorganismos tanto em humanos como em plantas. Para isso o estudo estrutural das enzimas que compõem a via se faz necessário, pois permitirá que ocorram avanços no entendimento da função molecular e no mecanismo de ação das enzimas, para tal estudo foram utilizadas as técnicas biofísicas de caracterização espectroscópica e estrutural.

A partir dos resultados obtidos neste trabalho foi desenvolvido um banco de dados intitulado como SKPDB (*ShiKimate Pathway Database*) que é um banco de dados relacional e curado, criado para evidenciar a importância do método de modelagem molecular comparativa aplicado a

caracterização estrutural de proteínas.

O SKPDB contém anotações detalhadas sobre a função molecular, mecanismo catalítico, estrutura e sua qualidade, informações biológicas gerais de cada enzima, e a maioria destas anotações estão disponíveis para *download*. Existe também *links* externos de acesso a outros bancos de dados internacionais como, por exemplo: Swiss-Prot/TrEMBL, PDB, KEGG, entre outros que fornecem mais informações sobre estas enzimas e a via. O banco de dados SKPDB pode ser acessado pelo endereço eletrônico: <http://lsbzix.rc.unesp.br/skpdb/index.html>.

1.1. A tuberculose

As doenças infecciosas ainda são responsáveis pela mortalidade de centenas de milhões de pessoas e 90% das mortes causadas por esse tipo de enfermidade ocorrem em áreas tropicais e subtropicais, regiões nas quais concentram-se 80% de toda população mundial (Figura 1). Entre estas doenças destacam-se a AIDS/HIV, infecções respiratórias, malária e tuberculose. (TROUILLER *et al.*, 2001).

A tuberculose, através do seu agente etiológico, o *Mycobacterium tuberculosis*, é responsável pelo maior número de mortes no mundo. Este microorganismo é quase exclusivamente um parasita humano, sua transmissão ocorre via aérea e causa tuberculose pulmonar, embora seja capaz de infectar outros tecidos ou órgãos do corpo (PELCZAR *et al.*, 1996).

Apesar da tuberculose ser considerada por muitos como uma doença do século XIX, na verdade ela nunca foi erradicada. Nos últimos anos tem-se verificado um aumento em sua incidência: entre 1997 e 2000, o número de novos casos de tuberculose aumentou 1,8% ao ano (CORBETT *et al.*, 2003). Estima-se que entre 2002 e 2020, se não houver um esforço para seu controle, aproximadamente 1 bilhão de pessoas serão infectadas, mais de 150 milhões ficarão doentes, e a tuberculose levará à morte mais de 38 milhões de pessoas (WHO, 2002).

É estimado que cem milhões de pessoas são infectadas anualmente. De cem milhões,

aproximadamente dez milhões desenvolvem a doença, sendo que cinco milhões dessas progridem para o estado infeccioso e, finalmente, três milhões vão a óbito (MORGAN *et al.*, 2003). Segundo o relatório anual da Organização Mundial de Saúde (OMS) de 2005, foram estimados 8,8 milhões de novos casos em 2003 e que 1,7 milhões de pessoas morreram (NUNN *et al.*, 2005). No Brasil, estima-se que cerca de 50 milhões de pessoas estejam infectadas. Conforme dados do Ministério da Saúde (MS) de 2005, foram cerca de 78.000 novos casos registrados em 2002, mantendo uma média de aproximadamente 80.000 novos casos registrados por ano.

Dada a gravidade da situação global da doença, a OMS declarou a tuberculose um sério problema de saúde pública mundial, como forma de alertar os setores de saúde pública e os governos para a necessidade de contenção desta doença.

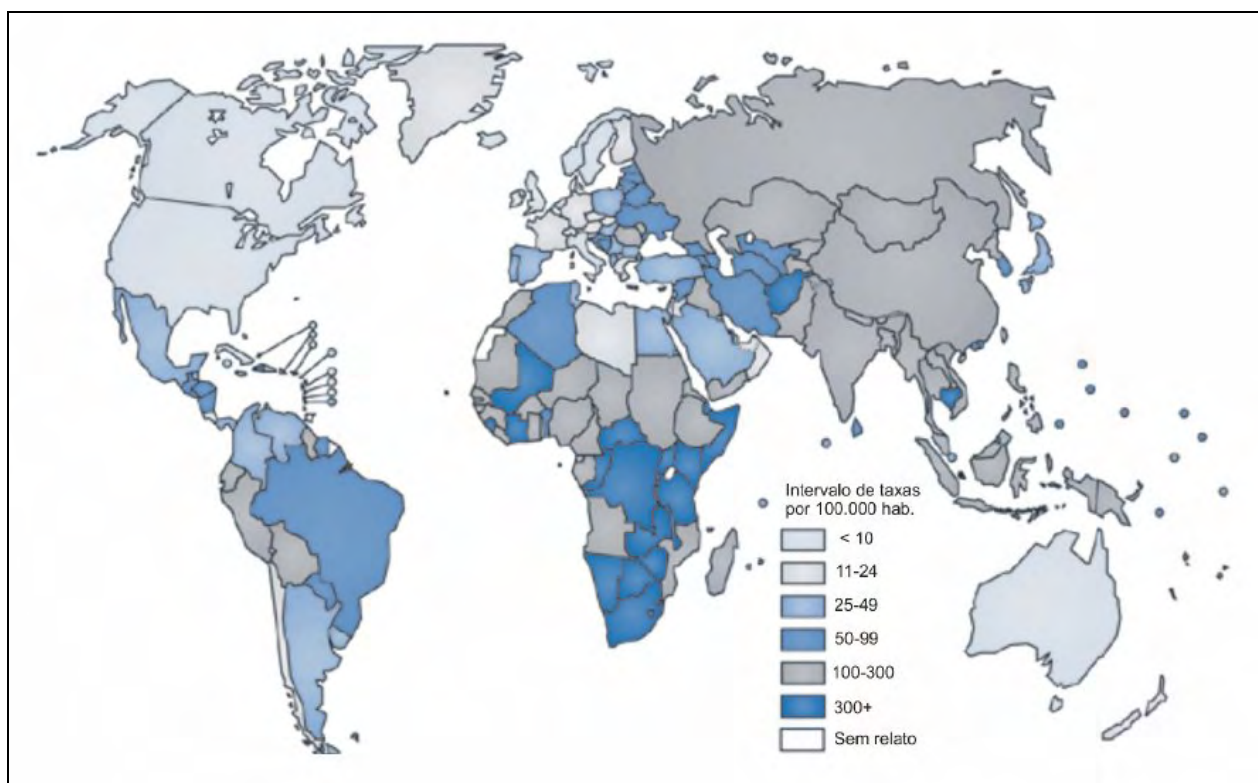


Figura 1. Distribuição geográfica das taxas de incidência de tuberculose estimadas em 2003 (Nunn *et al.*, 2005). No leste da Europa e África, estão aumentando as mortes após quase 40 anos de declínio. Outro fato importante é a presença da bactéria em níveis alarmantes nos países em desenvolvimento.

Alguns fatores vêm contribuindo para o aumento da tuberculose como: pacientes infectados

com o vírus da AIDS/HIV, diminuição do poder sócio-econômico da população e aglomeração em grandes cidades. Outro fator muito importante é o abandono do tratamento ou a prescrição de procedimentos inapropriados para o tratamento da tuberculose que acaba resultando na seleção de cepas resistentes aos fármacos de primeira linha utilizados no seu tratamento (BASSO *et al.*, 1998b).

A prescrição mais comum para o tratamento da tuberculose utiliza a combinação das drogas isoniazida, rifampicina, pirazinamida e etambutol ou estreptomicina nos primeiros dois meses, com o intuito de exercer uma ação bactericida contra o bacilo ativo, e, posteriormente, isoniazida e rifampicina para os últimos quatro meses. Estas têm a função de esterilização contra bacilos dormentes ou semidormentes. Este tratamento, apesar de longo, é importante para o controle da doença (BASSO & BLANCHARD, 1998a; BASSO *et al.*, 1998b).

O desenvolvimento de novas drogas para substituírem aquelas comprometidas pela resistência é premente para o desenvolvimento de um tratamento eficaz. O principal objetivo no desenvolvimento de novas drogas é atacar alvos peculiares aos microorganismos como, por exemplo, vias metabólicas ausentes no organismo humano. Tal fato, teoricamente, minimizaria o efeito tóxico destas drogas antimicrobianas e antiparasitas para a espécie humana (BASSO *et al.*, 1998b).

Sob este aspecto, as enzimas da via metabólica do ácido chiquímico (Figura 2) representam bons exemplos da utilidade de tal abordagem aplicada aos constituintes enzimáticos de uma rota biossintética presente em microorganismos e plantas, mas inexistente no organismo humano (BENTLEY, 1990).

Desta forma, a tuberculose ou qualquer outra doença causada por microorganismos poderá, em princípio, ser tratada por inibidores das enzimas da rota do ácido chiquímico que impossibilitarão a produção do ácido corísmico (RATLEDGE, 1982).

1.2. Via metabólica do ácido chiquímico

Nas células as reações enzimáticas não ocorrem isoladamente, mas são organizadas em seqüências de múltiplas etapas denominadas rotas ou vias, nas quais o produto de uma reação serve como substrato da reação subsequente. Por sua vez, diferentes vias se inter-relacionam, formando uma rede integrada e objetiva de reações químicas, denominada metabolismo. Cada via é composta de seqüências multi-enzimáticas e, cada enzima, por sua vez, pode exibir importantes características catalíticas ou regulatórias (RATLEDGE, 1982).

A via metabólica do ácido chiquímico (Figura 2) é uma das rotas biossintéticas que leva à produção de alguns dos metabólitos primários e secundários importantes na natureza (GANEM, 1978). Esta rota biossintética faz a conexão entre o metabolismo de carboidratos e a síntese de compostos aromáticos através de sete passos metabólicos, onde o fosfoenolpiruvato (PEP) e eritrose 4-fosfato (E4P) são convertidos em ácido corísmico (PITTARD, 1987; HASLAM, 1993).

O ácido corísmico é um precursor chave para a biossíntese de PABA (ácido *p*-aminobenzóico, precursor do tetraidrofolato), ácido *p*-hidroxibenzóico (precursor da coenzima Q ou *ubiquinona*), micobactinas, folato e dos aminoácidos aromáticos que são essenciais para microorganismos e plantas (PITTARD, 1987).

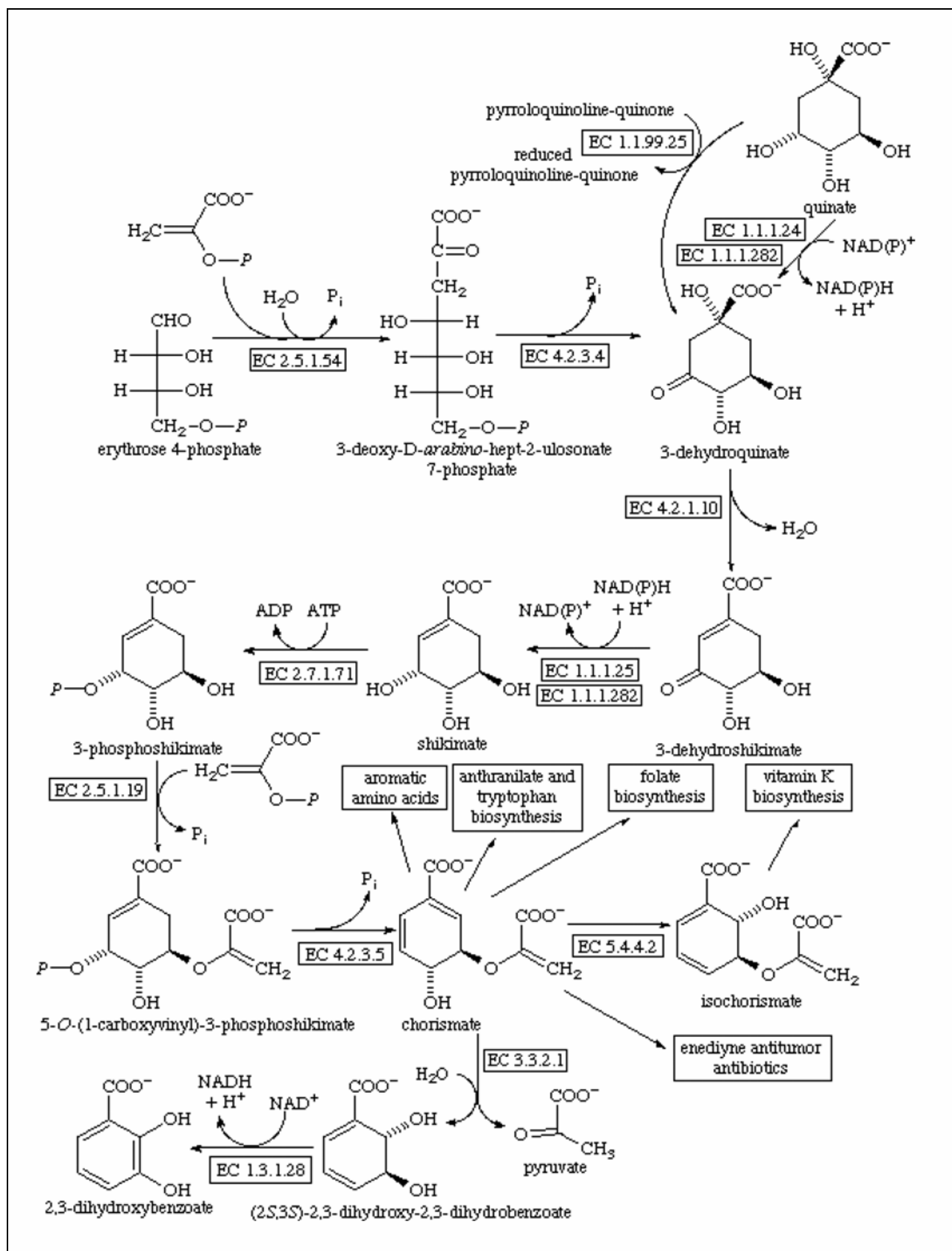


Figura 2. Via do ácido chiquímico na seqüência de sete passos metabólicos, iniciando no PEP e E4P que são convertidos em corismato. A descrição detalhada do mecanismo de reação de cada enzima é descrita durante o texto, a figura foi adaptada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>.

O ácido chiquímico (SK – 5° enzima – figura 3) 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).

A via metabólica do ácido chiquímico foi descoberta através dos estudos de Bernhard Davis e David Sprinson (DAVIS, 1950; GIBSON, 1999). A organização molecular das enzimas da via metabólica do ácido chiquímico altera-se de acordo com os grupos taxonômicos (COGGINS *et al.*, 1987).

As enzimas de bactérias apresentam-se na forma de sete cadeias polipeptídicas individuais, que são codificadas por genes separados. As plantas possuem um arranjo similar às bactérias (CAMPBELL *et al.*, 2004; HERRMANN & WEAVER, 1999), com a exceção da desidroquinase (DHQD – 3° enzima – figura 3) e a chiquimato desidrogenase (SD – 4° enzima – figura 3) que estão presentes como domínios separados em uma cadeia polipeptídica bifuncional (MOUSDALE *et al.*, 1987).

Em fungos e parasitas do filo Apicomplexa, a via do ácido chiquímico apresenta a 3-deoxi-D-arabino-heptulosonato 7- fosfato sintase (DAHPS – 1° enzima – figura 3) e a corismato sintase (CS – última enzima – figura 3) como enzimas monofuncionais e um polipeptídeo pentafuncional chamado AROM, que é responsável pela cinco reações restantes da via do ácido chiquímico (DUNCAN *et al.*, 1984).

A utilização das enzimas da via biossintética do ácido chiquímico como alvos para o desenvolvimento de inibidores pode ser plenamente justificada e validada pela inibição da EPSP sintase através do glifosato (*N*-(fosfometil) glicina) (KISHORE & SHAH, 1998).

O Glifosato (figura 3) é um importante composto guia para a construção de novos inibidores

de EPSPS. Este composto atualmente é utilizado como herbicida, e está presente nos defensivos agrícola *Round Up* e *TouchDown*, usados no controle de ervas daninhas. Roberts e seus colaboradores demonstraram também que o glifosato é capaz de inibir *in vitro* o crescimento dos parasitas *Plasmodium falciparum*, *Toxoplasma gondii* e *Cryptosporidium parvum* (ROBERTS *et al.*, 1998).

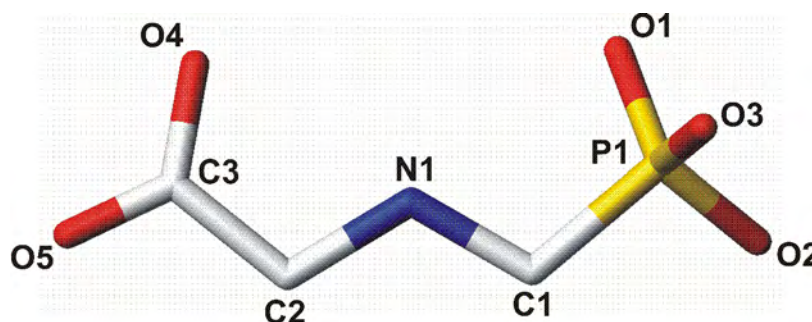


Figura 3. Estrutura molecular do glifosato (C₃H₈NO₅P). Este composto exemplifica o grande potencial de desenvolvimento de novos inibidores baseados nas enzimas da via metabólica do ácido chiquímico. As coordenadas atômicas do glifosato foram obtidas no PDB (código PDB:1G6S) (SCHÖNBRUNN *et al.*, 2001).

1.2.1. Descrição detalhada do mecanismo de reação das enzimas pertencentes a via

3-desoxi D-arabino-heptulosonato-7-fostato sintase (DAHPS)

A reação inicial (Figura 4) é feita pela enzima DAHPS (EC 2.5.1.54) que é uma metaloenzima que catalisa a condensação do PEP com o quarto carbono de açúcar de E4P para produzir DAHP e fosfato inorgânico (P_i) (HERRMANN & WEAVER, 1999; HERRMANN, 1995a).

O íon de metal tem função na catálise e pode também afetar a conformação da enzima (STEPHENS & BAUERLE, 1992; HERRMANN, 1995b). A enzima DAHPS mais investigada foi a de *Escherichia coli*, 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).

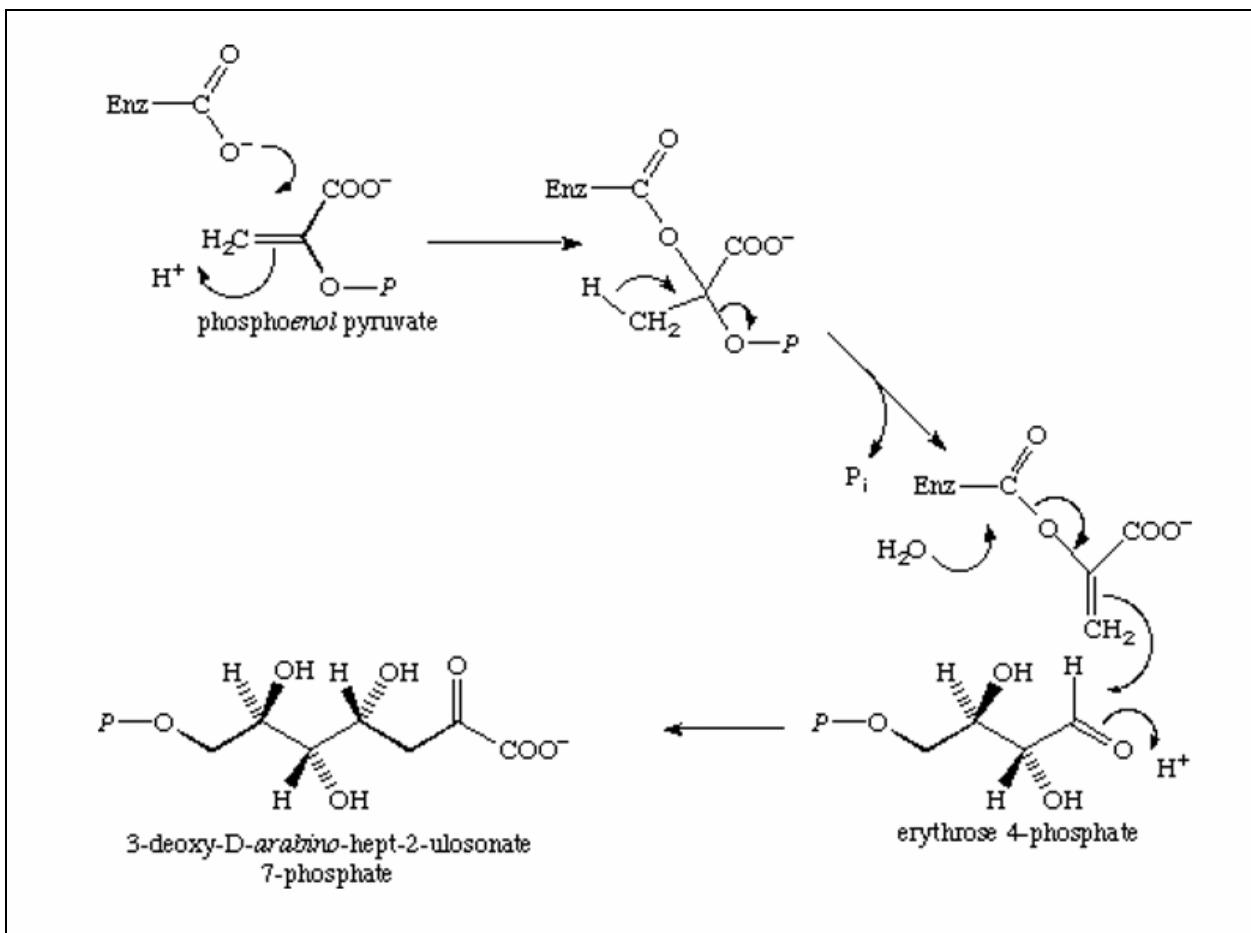


Figura 4. Reação catalisada pela enzima DAHPS, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>.

3-desidroquinato sintase (DHQS)

A segunda reação (Figura 5) na via metabólica do ácido chiquímico, é catalisada pela enzima DHQS (EC 4.2.3.4) que é ativada pelo íon 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. Em bactérias o Co²⁺ e Zn²⁺ são os íons metálicos mais ativos (HERRMANN & WEAVER, 1999; HERRMANN, 1995a).

Na segunda reação da via metabólica do ácido chiquímico há a eliminação de um fosfato do DAHP e a ciclização para 3-desidroquinato (DHQ), através da enzima DHQS usando o NAD⁺ como coenzima (CARPENTER *et al.*, 1998).

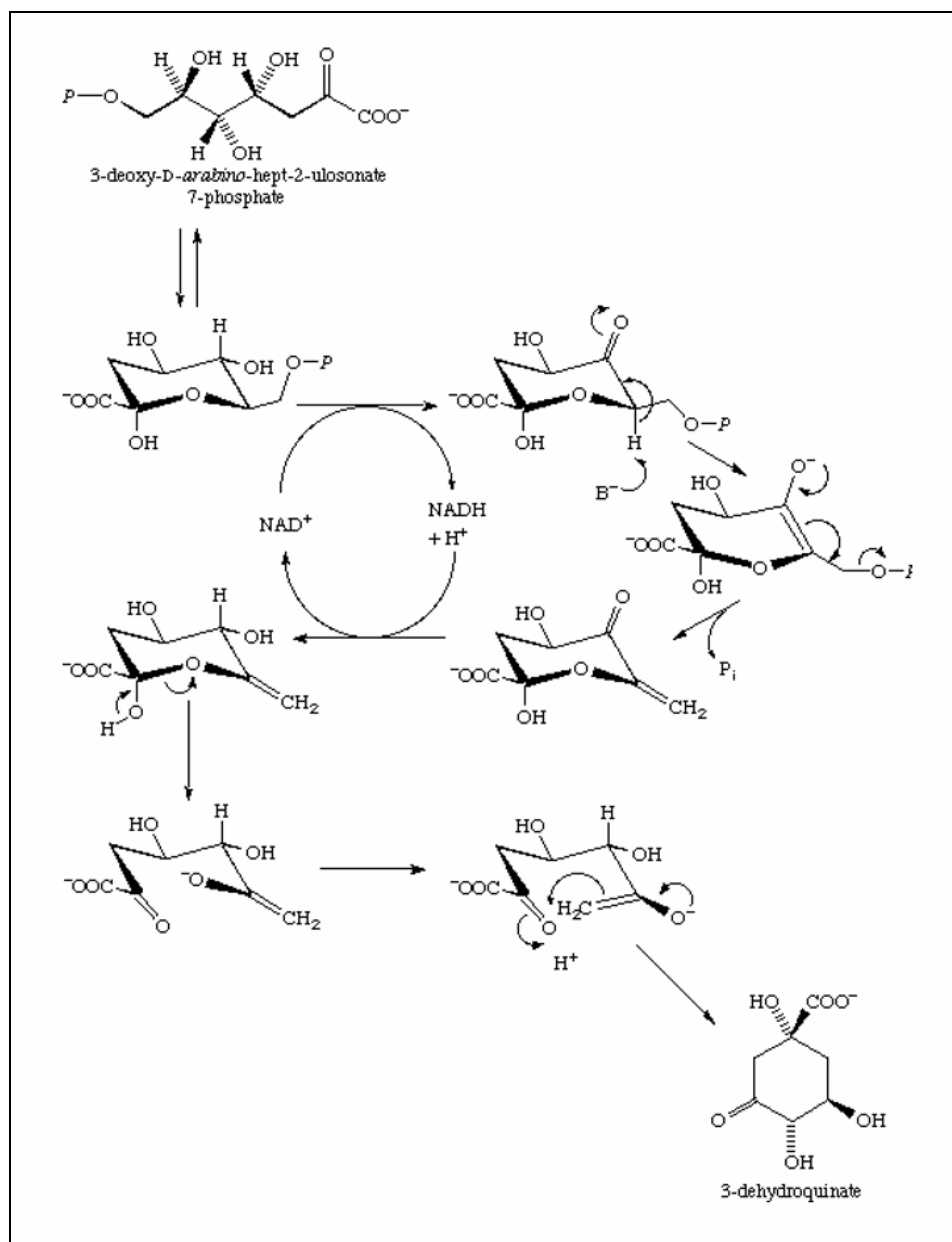


Figura 5. Reação catalisada pela enzima DHQS, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>

A ciclização acontece por uma troca intramolecular do 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 intra-molecular do aldol (HERRMANN & WEAVER, 1999; HERRMANN, 1995a).

3-desidroquinato desidratase (DHQD)

A desidratação do substrato DHQ fornece então 3-desidrochiquimato (DHS) pela enzima DHQD (EC 4.2.1.10), esta é a terceira reação da via (Figura 6). Esta enzima existe em dois tipos diferentes (tipo I e II) (HERRMANN & WEAVER, 1999). Esta reação de desidratação do DHQ resultando na síntese DHS é comum também para a via catabólica quinato (COGGINS *et al.*, 2003).

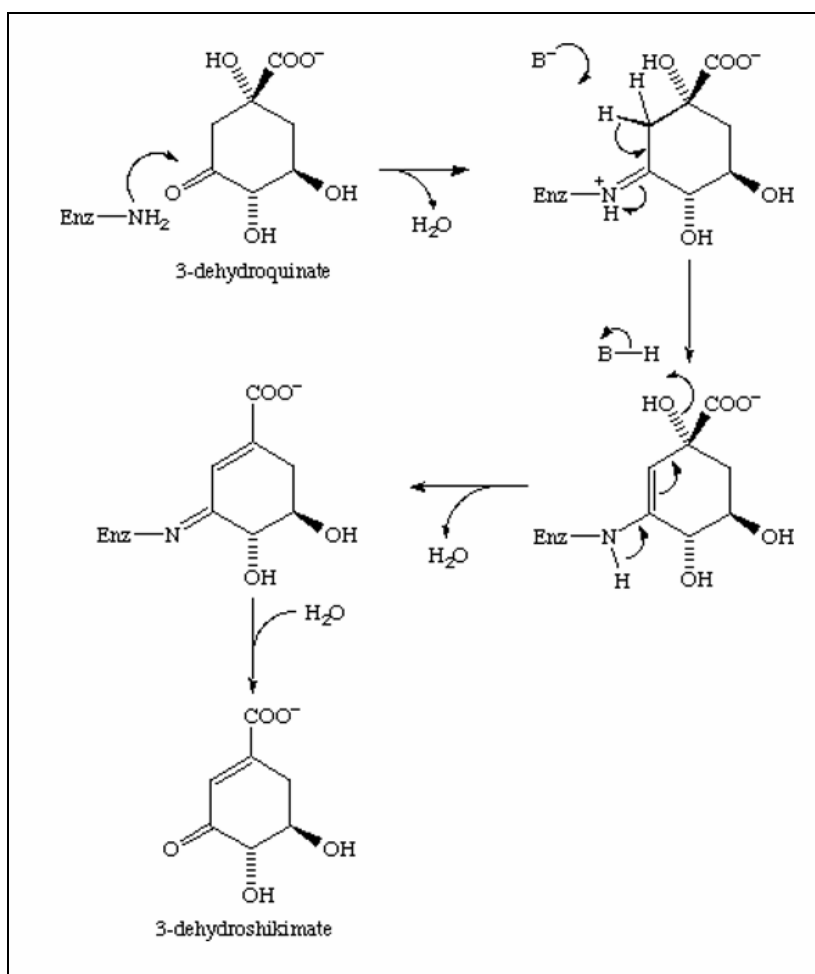


Figura 6. Reação catalisada pela enzima DHQD, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>

A enzima DHQD tipo I atua na via metabólica do ácido chiquímico e o DHQD tipo II tem as duas funções: de biossíntese e catálise. Os dois tipos de enzima DHQD têm arquiteturas completamente diferentes um da outra (COGGINS *et al.*, 2003). 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).

Chiquimato-5-desidrogenase (SD)

A quarta reação (Figura 7) ocorre quando o substrato DHS é reduzido pela enzima SD (EC 1.1.1.25), usando como coenzima NADH, a D-chiquimato (SHK). Esta enzima 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-ceto e desidrogenases de álcool ferro-ativado ativos (HERRMANN & WEAVER, 1999; HERRMANN, 1995a).

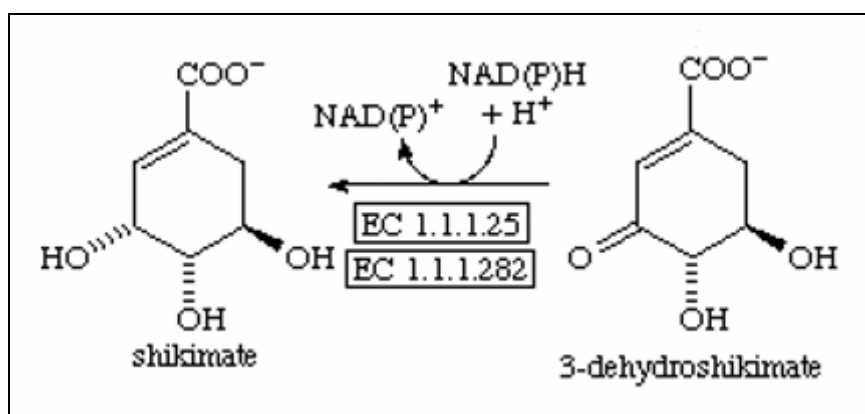


Figura 7. Reação catalisada pela enzima SD, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>

O mecanismo de reação usado pela enzima SD envolve a catálise da transferência de hidreto para a coenzima 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).

Chiquimato quinase (SK)

Na quinta reação (Figura 8) ocorre uma específica fosforização do grupo 3-hidroxil da D-

chiquimato usando como coenzima a molécula de adenosina trifosfato (ATP) através da enzima SK (EC 2.7.1.71) que fornece como produto desta reação chiquimato-3-fosfato (S3P) (KRELL *et al.*, 1998).

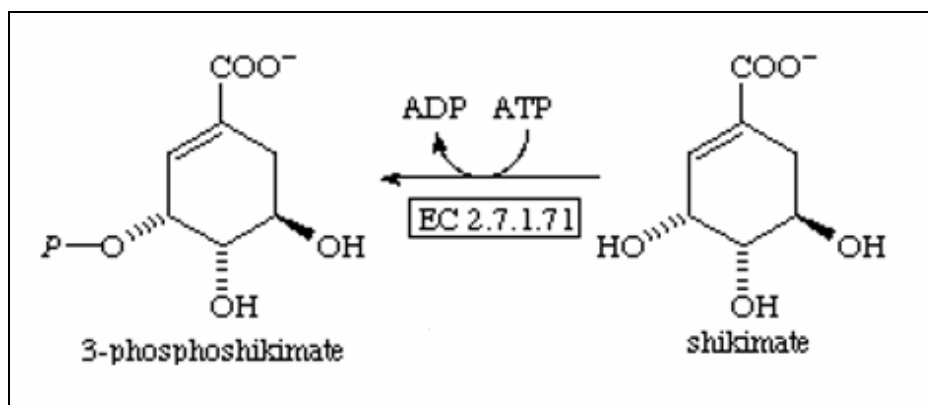


Figura 8. Reação catalisada pela enzima SK, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>.

A atividade enzimática apresenta dependência por cátions divalentes, como cofator. 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).

Em *Escherichia coli*, a reação da SK é catalisada por duas isoformas: SK I que é codificada pelo gene *aroK* (WHIPP & PITTARD, 1995) e SK II que é codificada pelo gene *aroL* (MILLAR *et al.*, 1986). A isoforma SK II apresenta um papel dominante na via do ácido chiquímico, onde a sua expressão é controlada pelo regulador *tyrR* e a sua repressão é realizada pela *Try* e *Trp* (ELY & PITTARD, 1979; DE FEYTER & PITTARD, 1986).

O papel fisiológico da SK I em *Escherichia coli* não está claro. Mutações em SK I foram associadas com a sensibilidade ao antibiótico mecilina, sugerindo que SK I deve apresentar um papel biológico alternativo e que promove a fosforilação do chiquimato somente casualmente (DE FEYTER & PITTARD, 1986).

Ao contrário de *Escherichia coli*, seqüências de genomas completos de bactérias, como

Haemophilus influenzae e *Mycobacterium tuberculosis*, demonstraram a presença de somente um gene codificador de SK. A maioria das SKs são codificadas pelo gene *aroK*, devido as seqüências de aminoácidos possuírem um maior grau de identidade com a SK I de *Escherichia coli* (GU *et al.*, 2002).

5-enoilpiruvilchiquimato 3-fosfato sintase (EPSPS)

Na penúltima reação (Figura 9) uma segunda molécula de PEP condensa-se com o substrato S3P para produzir 5-enoilpiruvilchiquimato 3-fosfato (EPSP), utilizando a enzima EPSPS (EC 2.5.1.19) (HERRMANN & WEAVER, 1999).

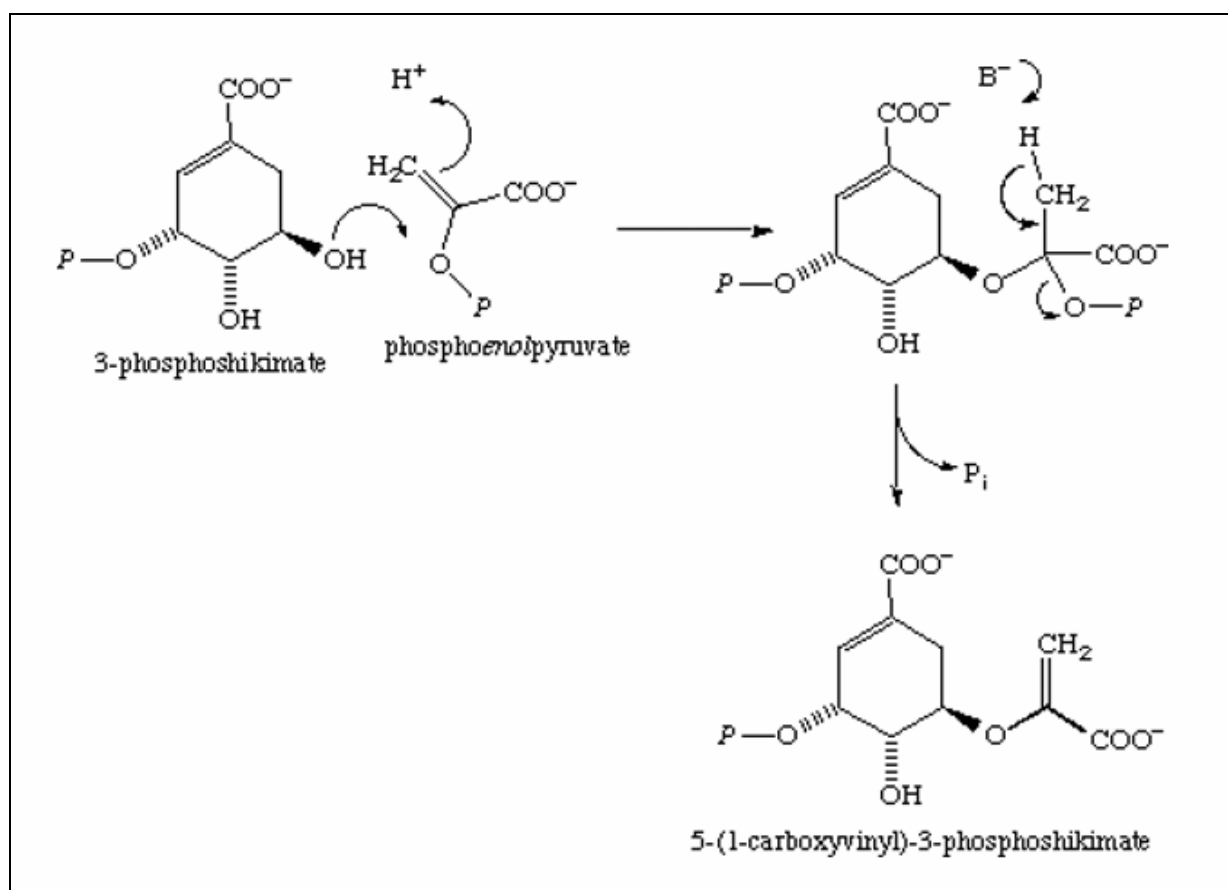


Figura 9. Reação catalisada pela enzima EPSPS, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>.

A EPSPS é inibida pelo glifosato, substância extensamente usada como herbicida de largo

espectro (STEINRÜCKEN & AMRHEIN, 1980). O glifosato 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).

Corismato sintase (CS)

Na última reação (Figura 10) o substrato EPSP perde seu grupamento fosforila, produzindo corismato pela ação da enzima CS (EC 4.2.3.5) na presença da molécula de flavina mononucleotideo (FMN) (HERRMANN & WEAVER, 1999). Esta reação, do ponto de vista bioquímico, é excepcional, pois tanto a enzima quanto a reação são únicas na natureza (BENTLEY, 1990; MACHEROUX *et al.*, 1999).

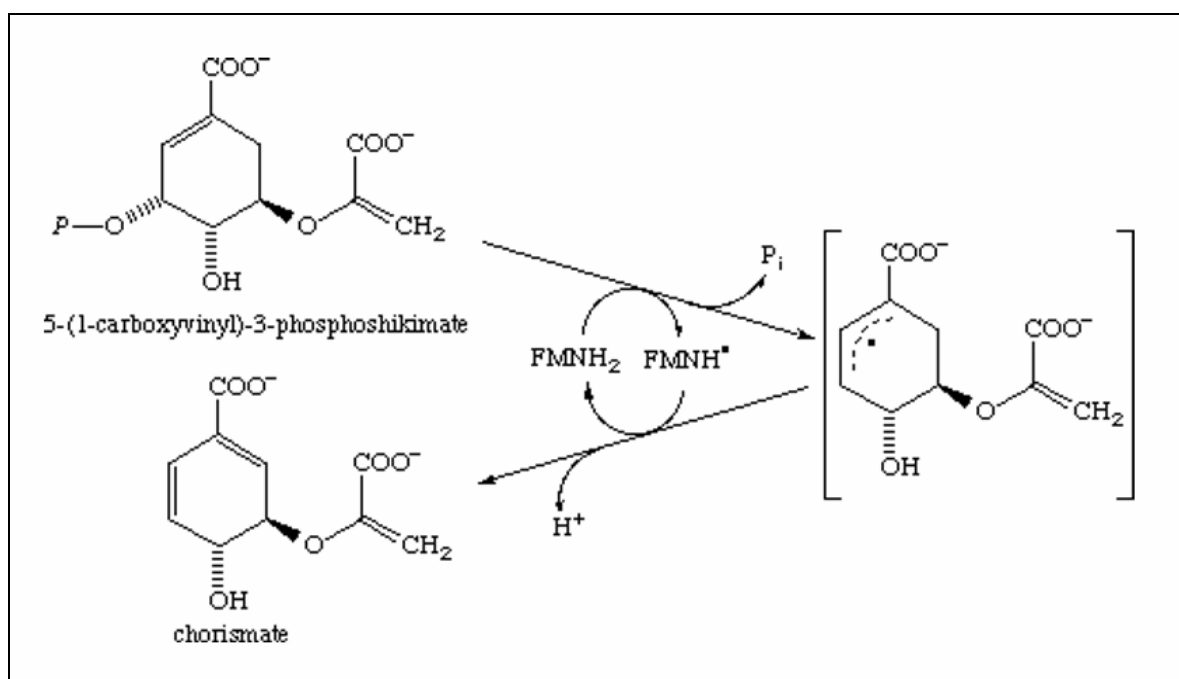


Figura 10. Reação catalisada pela enzima CS, figura retirada do endereço eletrônico: <http://www.chem.qmul.ac.uk/iubmb/enzyme/reaction/misc/shikim.html>.

A reação da CS acrescenta a segunda de três duplas ligações ao anel benzênico do substrato, e formalmente é descrita como uma reação concertada de eliminação *anti*-1,4 do grupo 3-fosfato e a perda do hidrogênio C-(6*proR*). Esta reação apresenta algumas particularidades, como o fato de não

envolver nenhuma alteração no estado redox de substrato e produto e apresentar uma estereoquímica do tipo *anti* (eliminações do tipo 1-4 apresentam preferencialmente estereoquímica do tipo *syn*) (MACHEROUX *et al.*, 1999).

Como o mecanismo de reação ainda não foi completamente elucidado, há diversas alternativas propostas, mas a mais aceita sugere o envolvimento direto do FMN na reação de eliminação, com um mecanismo radicalar com uma transferência de um elétron do FMNH₂ para o substrato, iniciando a quebra da ligação C-O do grupamento fosfato, produzindo um derivado com um radical. Como não há mudança de estado redox na reação, uma transferência reversa do elétron doado é necessária para completar o ciclo catalítico (MACHEROUX *et al.*, 1999).

Este mecanismo é corroborado pela resolução da estrutura da CS de *Streptococcus pneumoniae*, onde o FMN, o EPSP e a enzima formam um complexo ternário, onde a porção do anel isoaloxazina do FMN está a uma distância de 3,3Å do EPSP (MACLEAN & ALI, 2003). A formação deste complexo induz uma mudança conformacional na enzima (MACHEROUX *et al.*, 1998).

Existem duas explicações para a função da coenzima reduzida de FMN: (i) a FMN têm meramente uma função estrutural e não um papel funcional, (ii) a FMN 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 & NESTER 1978; MACHEROUX *et al.*, 1999).

1.3. Aplicações da bionformática estrutural

A bioinformática tem suas origens na década de 1960, quando os computadores emergiram como ferramentas importantes na Biologia Molecular. Este surgimento, segundo Hagen (2000), teria sido motivado pelo crescente número de seqüências protéicas disponíveis, que representavam, ao mesmo tempo, uma fonte de dados e um conjunto de problemáticas importantes, porém

intratáveis sem o auxílio de um computador (HAGEN, 2000).

Até o final dos anos 60, diversas técnicas computacionais (algoritmos e programas de computador) para análise da estrutura, função e evolução molecular, bem como bancos de dados rudimentares de seqüências protéicas, já haviam sido desenvolvidos (HAGEN, 2000; OUZOUNIS & VALENCIA, 2003).

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 (BAYAT, 2002). Para isso, é necessário vários especialistas de áreas como: computação, matemática, biologia, física e medicina e outros (Figura 11).

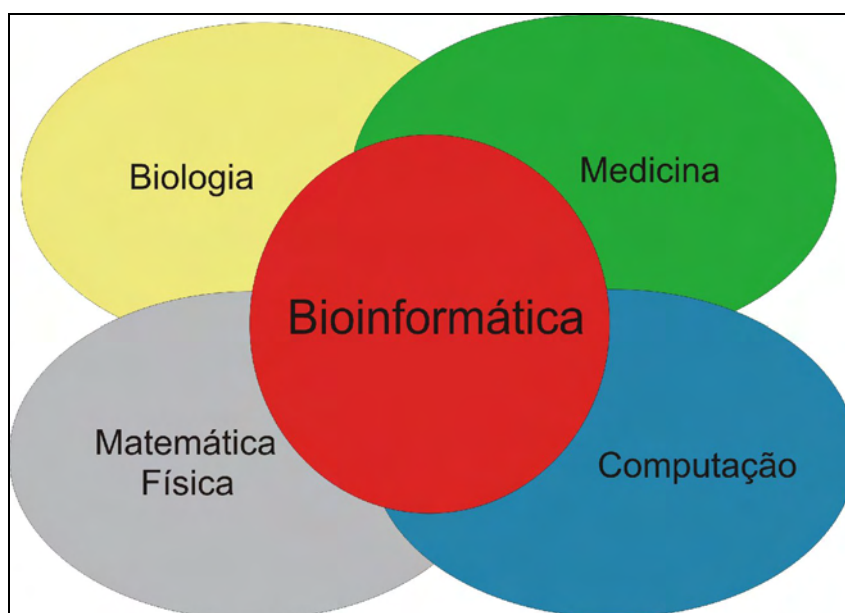


Figura 11. Interação das áreas que têm contribuído para a formação e o desenvolvimento da bioinformática (BAYAT, 2002).

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 (KITANO, 2002).

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 (HANGEN, 2000; BAYAT, 2002).

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 (SIMMONS, 1999).

Hoje em dia, já é possível replicar comportamentos biológicos através de simulação computacional. Entretanto, estas simulações possuem um alto custo computacional. As simulações computacionais dos processos biológicos são uma ferramenta importante no desenho racional de drogas. Com elas, é possível diminuir o alto custo do desenvolvimento de drogas. Outra possibilidade interessante é o tratamento personalizado, onde são coletadas informações biológicas do paciente e usando simulação desenvolve-se uma droga personalizada para o paciente (VEERAPANDIAN, 1997).

O desenho de drogas baseado em estrutura tornou-se uma tecnologia altamente desenvolvida e utilizada nas maiores empresas farmacêuticas. A modelagem molecular comparativa ou por homologia é uma característica chave de um esforço integrado no descobrimento de novas drogas, porque ela permite que estas informações genômicas sejam utilizadas no desenvolvimento de ligantes alvos ou na engenharia de especificidade de ligantes (VEERAPANDIAN, 1997).

Uma das mais importantes técnicas utilizadas em conjunto com a modelagem molecular em biologia estrutural é o *docking* de um ligante a um receptor, tal como uma proteína. Se a estrutura

do receptor é conhecida, então a aplicação é essencialmente de um desenho de droga baseado em estrutura. Estes métodos têm alguns objetivos relacionados, tais como: procurar identificar a localização do sítio ativo do ligante e talvez a geometria do ligante no sítio ativo. Uma outra meta é a classificação de uma série de ligantes relacionados em termos de sua afinidade ou avaliar a energia livre de ligação absoluta tão precisamente quanto possível (FOSTER, 2002). Uma série de trabalhos demonstra o sucesso no uso de modelos estruturais para auxiliar no desenho racional de drogas contra parasitas (RING *et al.*, 1993; QUE *et al.*, 1993).

Inúmeros bancos de dados aplicados a análise comparativa em larga escala de genomas microbianos e ferramentas computacionais de acesso público (na grande maioria) ou privado têm sido criados na tentativa de organizar e permitir acesso eficiente e rápido às informações geradas pelos projetos de larga escala (por exemplo: projeto genoma estrutural), bem como permitir a análise comparativa dessa quantidade de dados (GALPERIN *et al.*, 2005). Estes bancos de dados desenvolvidos a partir dos projetos de larga escala podem disponibilizar um enorme conjunto de informações que podem ser:

- Informação evolutiva: genes homólogos, frequências dos alelos,
- Informação genômica: localização no cromossomo, introns, regiões reguladoras,
- Informação estrutural: estruturas de proteína, tipos de dobramento, domínios estruturais,
- Informação da expressão: expressão específica em um dado tecido, fenótipos, doenças,
- Informação funcional: função molecular ou enzimática, papel em diferentes rotas,
- e outros.

A criação e manutenção de bancos de dados biológicos são por si só um desafio, devido não só a imensa quantidade de dados, mas, sobretudo devido à dificuldade de desenvolver esquemas e estruturas que represente de forma exata ou bastante aproximada a complexa relação existente entre os diversos componentes dos sistemas biológicos. Outra dificuldade é a criação de mecanismos eficientes de busca e obtenção de dados nestes bancos, que permitam a elaboração e execução de

consultas complexas e maciças, através de uma interface amigável para o usuário (GALPERIN *et al.*, 2005).

De uma forma geral, os bancos de dados que permitem análises comparativas em larga escala de genomas podem ser divididos em duas categorias principais, segundo seus propósitos: bancos de dados genômicos e bancos de dados de organismos específicos (GALPERIN *et al.*, 2005; OUZOUNIS & VALENCIA, 2003).

O SKPDB desenvolvido neste trabalho é um banco de dados relacional, específico e curado que estuda particularmente a via metabólica do ácido chiquímico em microorganismos e plantas. As informações sobre as seqüências primárias das enzimas que compõem a via foram anotadas e associadas a informações sobre função, domínios funcionais, estruturas 3D, e outras.

2. Objetivos

- Desenvolver ferramentas computacionais com o auxílio da computação de alto desempenho, integrando softwares de modelagem molecular comparativa e de análise de estruturas 3D de proteínas;
- Fazer um estudo estrutural e funcional das enzimas que compõem a via metabólica do ácido chiquímico identificada em genomas de microorganismos e plantas;
- Desenvolver um banco de dados relacional, público e curado com os resultados obtidos neste trabalho;
- Tentar cristalizar as enzimas chiquimato quinase em complexo com 2 (dois) análogos de ATP (inibidores de quinases), sendo eles *olomoucine* e *roscovitine*, e da corismato sintase em complexo com ácido corísmico (produto) e FMN (coenzima), utilizando a técnica de cristalografia de raios X, para entender melhor o modo de interação dos ligantes com estas enzimas;
- Utilizar os experimentos de CD e fluorescência para analisar a estabilidade e as possíveis mudanças conformacionais na presença dos ligantes naturais;
- Para estudar a forma nativa que a enzima corismato sintase se encontra, será utilizada a técnica experimental de SAXS.

3. Materias e Métodos

3.1. Modelagem Molecular

A modelagem molecular consiste na geração, manipulação e/ou representação realista de estruturas moleculares e cálculo das propriedades físico-químicas associadas. Ela pode ser assistida por computadores. O instrumento matemático usado é a química teórica e a computação gráfica é a ferramenta para manusear os modelos. Atualmente, os sistemas de modelagem molecular estão munidos de poderosas ferramentas para construção, visualização, análise e armazenamento de modelos de sistemas moleculares complexos que auxiliam na interpretação das relações entre a estrutura e a atividade biológica (HÖLTJE & FOLKERS, 1996).

A modelagem molecular pode ser aplicada ao planejamento de fármacos de modo direto ou indireto. Indiretamente, quando não se dispõe da estrutura do receptor, na tentativa de se obter parâmetros eletrônicos e estéricos que elucidem as relações estrutura-atividade biológica. Diretamente, quando se conhece a estrutura 3D do alvo biológico, na tentativa de compreender as interações entre o complexo ligante-receptor. Ambos os modos tentam otimizar o encaixe da molécula com o receptor. A modelagem molecular e suas representações gráficas permitem explorar aspectos tridimensionais de reconhecimento molecular e gerar hipóteses que levam ao planejamento e síntese de novos ligantes (HÖLTJE & FOLKERS, 1996).

3.2. Modelagem molecular comparativa ou por homologia

A seqüência primária de uma proteína determina sua estrutura 3D, contudo o algoritmo que permita, com precisão absoluta, determinar a estrutura 3D de uma proteína partindo-se de sua seqüência ainda está por ser determinado (LESK, 2001).

A modelagem molecular comparativa pode ser útil quando não se tem estrutura de proteínas determinadas experimentalmente por cristalografia de raios X ou RMN. A modelagem de proteínas por homologia tem sido particularmente importante, por gerar modelos confiáveis (SÁNCHEZ *et*

al., 2000). A condição necessária é que a semelhança entre a seqüência designada e as estruturas do modelo sejam detectáveis e que o alinhamento correto entre elas possa ser construído. Esta aproximação para a modelagem da estrutura é possível porque uma pequena mudança na seqüência da uma proteína normalmente resulta em uma pequena mudança em sua estrutura tridimensional (LESK, 2001).

Todas as aproximações baseadas nas restrições para modelagem molecular comparativa de proteínas extraem as distâncias e as restrições dos ângulos diedros a partir do alinhamento da seqüência alvo com as estruturas relacionadas, adicionando restrições implícitas pela topologia covalente (restrições estereoquímicas) e calculam o modelo pela minimização das violações de todas as restrições. Desta forma, as duas principais diferenças entre as várias aproximações estão na derivação e satisfação das restrições espaciais (ŠALI *et al.*, 1995).

A precisão do método está em assumir que se há semelhança detectável entre duas seqüências lineares, a semelhança estrutural pode ser assumida e a função potencial guiará o modelo no caminho dos moldes em direção à estrutura correta (SÁNCHEZ & ŠALI, 1997).

Todo método de modelagem molecular comparativa consiste basicamente de quatro passos seqüenciais: (I) Identificação e seleção de proteínas-molde; (II) Alinhamento das seqüências alvo-molde; (III) Construção das coordenadas do modelo e (IV) Validação do modelo. O fluxograma apresentado na Figura 12 mostra os passos que devem ser seguidos para a construção de um modelo por modelagem molecular comparativa.

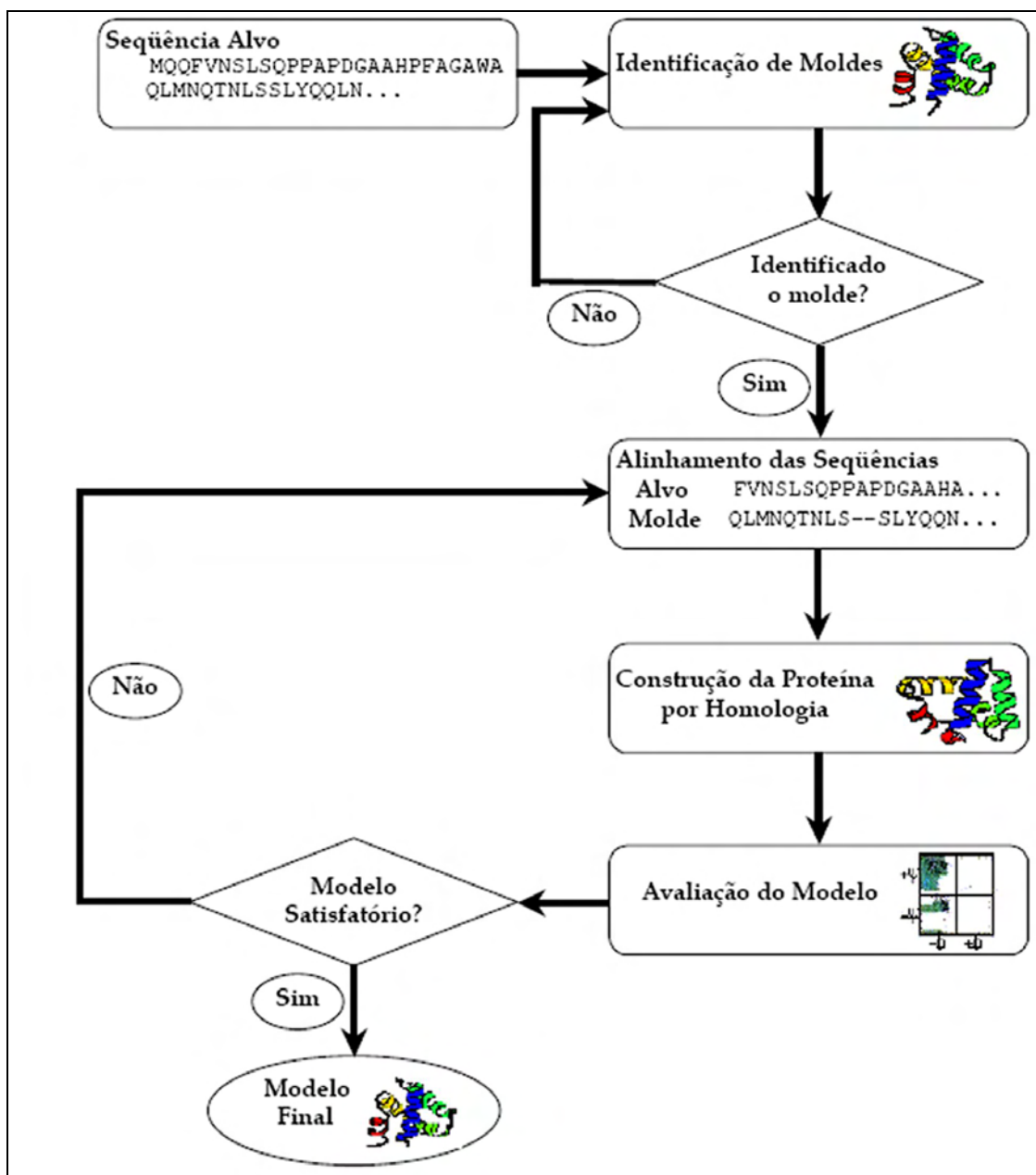


Figura 12. Passos da modelagem molecular comparativa adaptado do artigo de Sanchez & Šali, 1997.

3.2.1. Procura e seleção de *moldes*

A modelagem inicia-se pela procura de moldes em um banco de dados de estruturas de proteínas (PDB) (BERMAN, *et al.* 2000, WESTBROOK *et al.*, 2002 – <http://www.rcsb.org/pdb>), usando-se como parâmetro de entrada uma sequência de aminoácidos sem estrutura determinada experimentalmente (alvo) para que esta seja alinhada com as sequências de aminoácidos de estruturas determinadas experimentalmente (moldes) depositadas no PDB.

Neste primeiro passo de desenvolvimento do trabalho, foi realizada uma busca no banco de dados de seqüência de aminoácidos de proteínas: Swiss-Prot/TrEMBL (BOECKMANN *et al.*, 2003 – <http://us.expasy.org/sprot/>). A busca foi feita utilizando o código das enzimas, por exemplo: EC 1.1.1.25 que pertence à enzima SD (Enzyme Classification – Nomenclature Committee of the International Union of Biochemistry and Molecular Biology).

As informações sobre as seqüências de aminoácidos encontradas no Swiss-Prot/TrEMBL foram anotadas para posteriormente serem armazenadas no banco de dados, numa fase posterior as seqüências de aminoácidos repetidas e as incompletas foram retiradas, tornando o banco de dados SKPDB curado, ou seja, sem repetições ou dados incompletos.

Foi também realizada uma busca no banco de dados PDB que armazena informações sobre estruturas de proteínas resolvidas por técnicas experimentais (Cristalografia de raios X ou RMN), para verificar quais as seqüências de aminoácidos encontradas que tinha estruturas já resolvidas. As informações sobre as estruturas destas enzimas foram anotadas para também serem armazenadas no banco de dados.

Numa segunda etapa, foram feitos alinhamentos para cada uma das seqüências de aminoácidos sem estrutura 3D (alvo) contra cada uma das seqüências de aminoácidos com estruturas resolvidas adquiridas no PDB (BERMAN *et al.*, 2000; <http://www.rcsb.org/pdb/>) com a finalidade de encontrar o molde que seriam utilizados no passo de construção dos modelos destas enzimas.

Os alinhamentos foram feitos utilizando um pacote de programas para bioinformática chamado EMBOSS que é gratuito. Este programa de alinhamento utiliza o algoritmo de programação dinâmica de Needleman e Wunsch (1970) para executar os alinhamentos. Foi usado como parâmetro de restrição para a escolha do molde à porcentagem de identidade (>30%) entre a seqüência de aminoácidos do molde e seqüência de aminoácido alvo que será modelada, aliado a um parâmetro secundário de restrição quando possível que o molde estivesse na sua forma apo, ou

seja, sem ligantes.

3.2.2. Alinhamento molde/alvo

Esta segunda fase da modelagem molecular comparativa é um dos principais passos na modelagem, pois é a partir do segundo alinhamento entre a seqüência de aminoácidos do molde e do alvo que são extraídas as informações necessárias para a construção do modelo.

Para construir o arquivo de alinhamento de entrada que será utilizado pelo programa MODELLER na construção do modelo, foi utilizado o próprio alinhamento executado pelo programa MODELLER que utiliza o comando ALIGN2D, o qual é baseado também no algoritmo de programação dinâmica (NEEDLEMAN & WUNSCH, 1970).

3.2.3. Construção do modelo

Uma vez realizado o alinhamento entre a seqüência de aminoácido do alvo e do molde, o modelo é construído utilizando-se a modelagem molecular comparativa por satisfação das restrições espaciais implementadas no programa MODELLER (ŠALI & BLUNDELL, 1993) que usa distância geométrica e técnicas de otimização para satisfazer as restrições espaciais obtidas do alinhamento (SALI *et al.*, 1995).

O ponto inicial para construção do modelo, utilizando o programa MODELLER é quando o alinhamento é feito entre a seqüência de aminoácidos alvo e a(s) seqüência(s) de aminoácidos do molde (GIBAS e JAMBECK, 2001), um outro arquivo que também é essencial para construir o modelo 3D da seqüência de aminoácidos alvo é o arquivo de coordenadas atômicas 3D do molde.

O programa MODELLER gera, a partir da estrutura do molde, um conjunto de restrições que são aplicadas à seqüência de aminoácidos alvo a ser modelada. Estas restrições limitam, por exemplo, a distância entre dois resíduos no modelo que está sendo construído, sendo esta distância baseada naquela entre os dois resíduos equivalentes na estrutura molde. Esse cálculo de restrições

também é aplicado nas ligações angulares (entre três átomos) e nos ângulos diedrais (entre quatro átomos) (SALI & OVERINGTON, 1994; SALI *et al.*, 1995).

O cálculo das restrições é baseado em análises estatísticas entre estruturas comparativas. Além dessas restrições um campo de forças também controla as propriedades estereoquímicas entre os átomos, aplicando restrições químicas, adicionais na conformação molecular 3D (ŠALI & BLUNDELL, 1993). Finalmente, o modelo é obtido pela otimização da função objetiva no espaço cartesiano. Vários modelos ligeiramente diferentes podem ser calculados variando a estrutura inicial (SALI & OVERINGTON, 1994; SALI & BLUNDELL, 1993).

3.2.4. Avaliação dos modelos

A qualidade do modelo predito determina a informação que pode ser extraída dele. Assim, estimar a precisão do modelo 3D da proteína é essencial para interpretá-lo. O modelo pode ser avaliado como um todo bem como em regiões individuais, com base na similaridade entre as seqüências do molde e do alvo, observando resíduos importantes em regiões da proteína como o sítio ativo e sua conservação (SÁNCHEZ & ŠALI, 1997).

Um requerimento básico para um modelo é ter uma boa qualidade estereoquímica. Os programas utilizados para este tipo de análise foram o programa PROCHECK (LASKOWSKI *et al.*, 1993) que pertence ao pacote CCP4 (Collaborative Computational Project N° 4) e o programa WHATCHECK (HOOFT *et al.*, 1996). As características bioquímicas de um modelo que são checadas por estes programas incluem comprimento de ligação, ângulo de ligação, ligação peptídica e planaridade de anéis da cadeia lateral, quiralidade, ângulos de torção da cadeia principal e da cadeia lateral e choques entre pares de átomos não ligados.

Há também métodos para testar modelos 3D que implicitamente carregam muitas características espaciais compiladas de estruturas de proteínas a alta resolução. Estes métodos são baseados nos perfis 3D e potenciais estatísticos de força (SIPPL, 1990; LUTHY *et al.*, 1992). Um

dos programas que implementa esta aproximação é o VERIFY 3D (LUTHY *et al.*, 1992), escolhido para ser usado neste trabalho. O VERIFY 3D avalia o ambiente químico de cada resíduo em um modelo com respeito ao ambiente químico esperado como encontrado em estruturas de raios X à alta resolução.

Para analisar quanto à posição de cada átomo da estrutura modelada desvia em relação aos parâmetros observados por ENGH e HUBER, (1991) utilizamos o RMSD da geometria ideal (KEARSLEY, 1989), que é calculado como uma função de todos os átomos de uma proteína (ENGH & HUBER, 1991). 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 X-PLOR (SCHWIETERS *et al.*, 2003; BRÜNGER, 1992) foi utilizado para executar os cálculos de comprimento e ângulos de ligação.

3.3. Modelagem em larga escala das enzimas da via do ácido chiquímico

Devido ao excelente progresso na biologia, existe a necessidade de descrever e entender a função de muitas proteínas em mais detalhes. Embora funções de proteínas sejam melhores determinadas experimentalmente (OLIVER, 1996), algumas vezes podem ser preditas pela comparação da seqüência de uma proteína com proteínas de funções conhecidas (OLIVER, 1996; DUJON, 1996). Isto é possível porque seqüências de proteínas similares tendem a ter funções similares, embora exceções também ocorram (ORENGO *et al.*, 2002).

Devido à via metabólica do ácido chiquímico ser essencial para a sobrevivência de microorganismos e plantas vimos nesta via um bom exemplo para se criar uma ferramenta computacional automatizada que pudesse, com a utilização da modelagem molecular comparativa de proteínas, determinar todos os modelos possíveis para as enzimas da via metabólica do ácido chiquímico identificadas nos genomas de microorganismos e plantas.

Estes modelos poderão ser utilizados posteriormente para a interpretação da existência de dados funcionais, construção de proteínas mutantes para teste de novas hipóteses funcionais e para o desenho racional de drogas baseado em estrutura tornando as drogas propostas específicas e eficazes contra enfermidades causadas por microorganismos tanto em humanos como em plantas, como por exemplo, a malária e tuberculose; doenças que são negligenciadas por serem consideradas enfermidades existentes predominantemente nos países em desenvolvimento.

Os recursos computacionais envolvidos e utilizados para o desenvolvimento deste trabalho foi um *Cluster Beowulf* instalado no Laboratório de Sistemas Biomoleculares do Departamento de Física do IBILCE/UNESP, endereço eletrônico: (<http://www.biocristalografia.df.ibilce.unesp.br>), o qual se apresenta com uma configuração de *hardware* composta de 16 nós com AthlonXP-2100+, 1 Gb de RAM, 80 Gb HD e placas de rede 3COM de 10 Mbits conectadas a um Switch 3COM SuperStack 3300 10/100 Mbits, com sistema operacional UNIX/Linux Debian. Todos os modelos estão acessíveis no banco de dados SKPDB no endereço eletrônico: <http://lsbzix.rc.unesp.br/skpdb/index.html>.

3.4. Docking Molecular

Este tópico foi escrito utilizando o seguinte livro: *Protein – Ligand Interactions: structure and spectroscopy* (HARDING & CHOWDHRY, 2001).

Uma das questões fundamentais em biologia estrutural é o estudo da interação entre uma proteína e um ligante, um método eficaz muito usado é o *docking* molecular. Através destas interações podem ser descobertos inibidores para as proteínas. Este processo é muito importante para aplicações farmacológicas, como o desenho de drogas baseado em estrutura, pois estes ligantes podem inibir uma proteína vital para um determinado organismo.

O *docking* molecular é mais simples e conseqüentemente mais rápido, pois não envolve a introdução do campo de forças, nem a solvatação do sistema, além de considerar as moléculas como

estruturas rígidas. Estas características possibilitam a redução significativa dos graus de liberdade do sistema. Embora, na natureza o princípio da chave-fechadura seja válido para chaves e fechaduras flexíveis, os algoritmos de *docking* rígidos apresentam um êxito considerável como ferramenta de alocação do ligante no sítio ativo da enzima.

Entre as diversas metodologias existentes para fazer o *docking* molecular, optamos em utilizar uma metodologia que visa somente o tratamento da flexibilidade de ligantes. O programa escolhido para executar o *docking* molecular dentre vários foi o Molegro Virtual Docking (User manual – MVD 2007), adquirido no seguinte endereço eletrônico: <http://www.molegro.com>.

O Molegro Virtual Docker (MVD) é um programa gratuito que tem uma plataforma integrada de análise e predição de *docking* molecular entre proteína e ligante e pode ser instalado em várias plataformas como Windows, Linux, MAC, entre outros. O MVD processa todos os aspectos relativos ao *docking* molecular entre proteína-ligante, desde a preparação até a predição dos melhores modos de acoplamento de ligantes em sítios ativos potenciais da estrutura protéica em estudo.

3.5. Clonagem, expressão, purificação das enzimas e obtenção dos ligantes

As proteínas foram gentilmente cedidas pelo Prof. Dr. Diógenes Santiago Santos e pelo Prof. Dr. Luiz Augusto Basso do Grupo de Microbiologia Molecular e Funcional do Departamento de Biologia Molecular e Biotecnologia do Centro de Biotecnologia da UFRGS (Universidade Federal do Rio Grande do Sul), que realiza a clonagem e super expressão das enzimas da via metabólica do ácido chiquímico de *Mycobacterium tuberculosis*, além de trabalhar na síntese enzimática e na purificação dos substratos não disponíveis comercialmente.

Também temos a colaboração do Grupo de Biologia Estrutural e Zooquímica/CEIS-IBRC-UNESP, que realiza a purificação das enzimas super expressas na forma solúvel e caracterização da estrutura primária por métodos de espectrometria de massas, uma medida de segurança e de

qualidade das enzimas recombinantes.

As enzimas da via do ácido chiquímico de *Mycobacterium tuberculosis*, utilizadas neste projeto (chiquimato quinase e corismato sintase), já foram cristalizadas sem e com outros ligantes pelo grupo do Laboratório de Sistemas Biomoleculares da UNESP de São José do Rio Preto (DIAS *et al.*, 2004; PEREIRA *et al.*, 2004).

3.5.1. Preparação das amostras e dos ligantes

Chiquimato Quinase

A chiquimato quinase de *Mycobacterium tuberculosis* (mtCQ) veio precipitada em sulfato de amônio (80%), foi solubilizada e dialisada em Tris HCl 20mM (pH 7,4) e KCl 40mM.

Como já descrito anteriormente a mtCQ usa ATP como substrato (KRELL *et al.*, 1998), então é possível que a mtCQ possa ser inibida por inibidores análogos de ATP, por exemplo, que inibem CDKs (Quinases dependentes de Ciclinas), como os ligantes *roscovitine* e *olomoucine*. Estes inibidores foram escolhidos pelo fato de apresentarem bons resultados de inibição em CDK's (DE AZEVEDO *et al.*, 1996). Para este trabalho será utilizado o ADP no lugar do ATP para não correr o risco que ocorra a catalise durante os experimentos.

Os ligantes *roscovitine*, *olomoucine* e ADP foram adquiridos da empresa Sigma Co (Figura 13), bem como os outros reagentes químicos utilizados para realizar os experimentos com esta proteína neste trabalho. O ligante ADP foi diluído no próprio tampão da proteína mtCQ e os outros ligantes *roscovitine* e *olomoucine* foram diluídos em dimetil sulfóxido (DMSO).

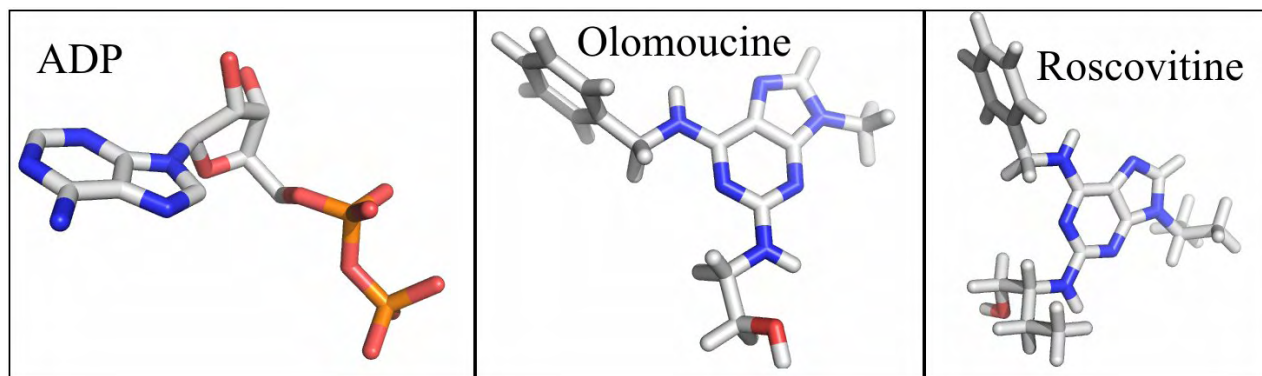


Figura 13. Estrutura molecular dos ligantes ADP, Olomoucine e Roscovitine. As coordenadas atômicas dos ligantes foram obtidas no PDB.

Para os experimentos de cristalização as amostras foram preparadas na forma apo e na forma dos seguintes complexos: *mtCQ-roscovitine-MgCl²⁺* e *mtCQ-olomoucine-MgCl²⁺*. Os ligantes foram adicionados na seguinte proporção: *roscovitine* 2:1 de proteína, *olomoucine* 2:1 de proteína e *MgCl²⁺* 10:1 de proteína, as misturas foram centrifugadas por 5 minutos a 10.000 rpm.

Já nos experimentos de CD as amostras foram preparadas na forma apo e na forma dos seguintes complexos: *mtCQ-roscovitine-MgCl²⁺*, *mtCQ-ADP-MgCl²⁺* e *mtCQ-olomoucine-MgCl²⁺*. Os ligantes foram adicionados na seguinte proporção: ADP 5:1 de proteína, *roscovitine* e *olomoucine* 2:1 de proteína e *MgCl²⁺* 10:1 de proteína, cada mistura foi centrifugada por 5 minutos a 10.000 rpm.

Corismato sintase

A proteína corismato sintase *Mycobacterium tuberculosis* (mtCS) veio precipitada em sulfato de amônio (65%), foi solubilizada e dialisada em Tris HCl 50mM (pH 7,5).

Como já discutido anteriormente, o mecanismo de reação catalítica da mtCS ainda não foi completamente elucidado, há diversas alternativas propostas, a mais aceita é a de MACHEROUX, (1999). Este mecanismo foi corroborado pela resolução da estrutura da CS de *Streptococcus pneumoniae* em complexo com o FMN e o EPSP (MACLEAN & ALI, 2003).

Uma informação mais detalhada da reação catalítica da mtCS em complexo com os ligantes

poderia elucidar a forma que ocorre a catálise e as possíveis mudanças conformacionais nos *loops* do sítio ativo, mostrando assim que o movimento dessas regiões resulta em um maior ambiente apolar para a ligação do co-fator.

Os ligantes FMN e ácido corísmico (AC) foram adquiridos da empresa Sigma Co (Figura 14), bem como os outros reagentes químicos utilizados para realizar os experimentos com esta proteína neste trabalho, os ligantes foram diluídos no próprio tampão da proteína mtCS.

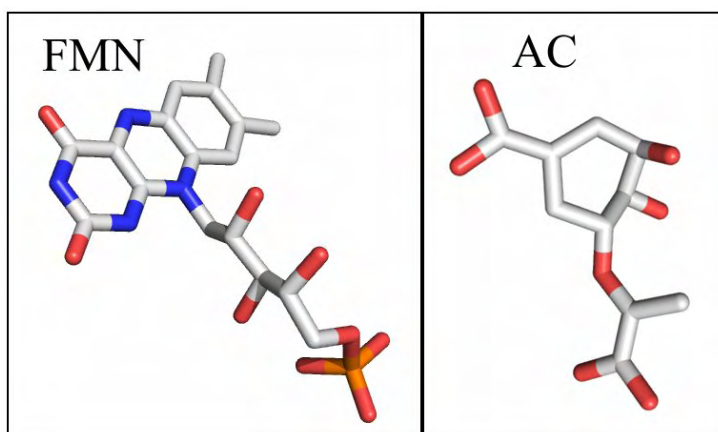


Figura 14. Estrutura molecular dos ligantes FMN e AC. As coordenadas atômicas dos ligantes foram obtidas no PDB.

Para os experimentos de cristalização, CD e fluorescência as amostras foram preparadas na forma apo e formando os complexos mtCS-FMN, mtCS-AC e mtCS-FMN-AC. Os ligantes foram adicionados na seguinte proporção: AC 5:1, 10:1 e 15:1 de proteína e FMN 5:1, 10:1 e 15:1 de proteína, as misturas foram centrifugadas por 5 minutos a 10.000 rpm.

No experimento de fluorescência também foi alterando a concentração do sal usado como precipitante em 1M, 2M e 3M, para podermos estudar se o sal poderia interferir na afinidade entre a mtCS e os ligantes (FMN e AC).

3.6. Cristalização de proteínas

Este tópico foi escrito utilizando os seguintes livros: *Fundamentals of Crystallography* (GIACOVAZZO *et al.*, 1992) e *Principles of Protein X-ray Crystallography* (DRENTH, 1999).

A cristalização de proteínas é considerada por muitos como um processo de tentativa e erro

no qual as proteínas são precipitadas, lentamente, a partir de uma solução que as contém. Os cristais de proteína devem ser crescidos bem lentamente para alcançarem um maior grau de ordem em sua estrutura. Existem vários métodos de cristalização de proteínas, o mais conhecido atualmente e que foi utilizado em todos os experimentos de cristalização descritos nesta tese foi o método da gota suspensa (Figura 15).

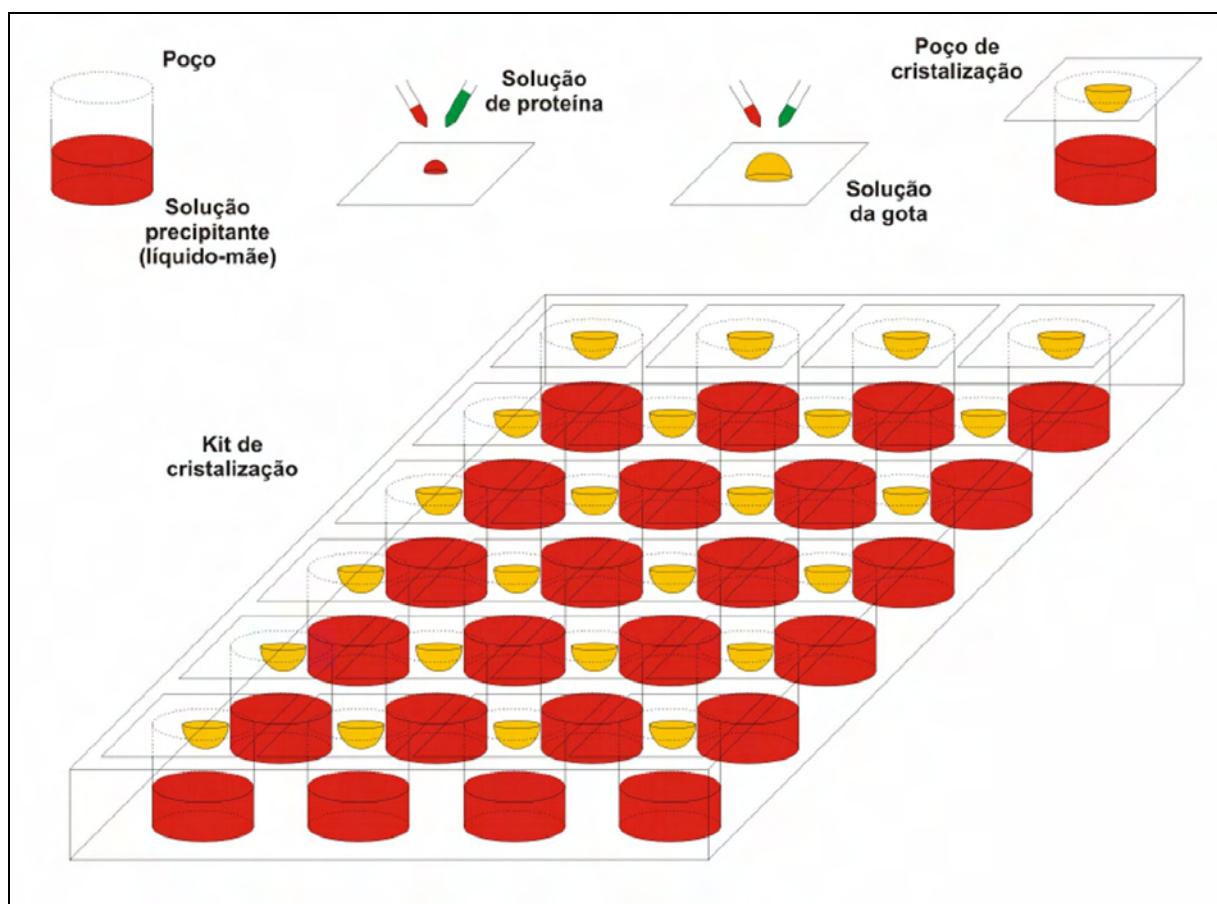


Figura 15. Ilustração esquemática do método da difusão de vapor pela gota suspensa. Figura retirada do texto Nangem, 2003.

Normalmente neste método, uma única gota, formada pela mistura de 1 a 5 μl de uma solução de proteína com a mesma quantidade de uma solução precipitante, deve ser colocada sobre uma lamínula de silicone devidamente preparada. A lamínula deve ser invertida e colocada em cima de um pequeno recipiente, vedando-o. O recipiente, conhecido como poço, deve estar preenchido com a mesma solução precipitante utilizada anteriormente, como mostra a figura 15. A difusão do

vapor do solvente entre a solução precipitante contida no poço e a gota promoverá o estado de supersaturação na gota, permitindo a formação dos primeiros núcleos de cristalização.

Dentre os parâmetros que mais afetam o processo de cristalização, pode-se citar a natureza e a concentração da proteína e do precipitante, o tipo da solução tampão e o seu pH, além da temperatura. Dessa forma, para se otimizar o processo de cristalização, muitos desses parâmetros precisam ser testados e analisados em todas as suas extensões, o que torna a cristalização de proteínas um processo que envolve inúmeras experiências sob diversas circunstâncias.

Para os experimentos de cristalização deste trabalho foram utilizadas lamínulas siliconizadas e placas da marca Linbro. E os reservatórios da placa foram vedados com graxa de silicone. Esses acessórios foram adquiridos comercialmente da empresa Hampton Research.

3.6.1. Cristalização da enzima chiquimato quinase

Para cristalização da mtCQ na forma apo e em complexo com os ligantes tentamos as seguintes condições descritas por: (1) Cheng *et al.*, 2005; (2) Gan *et al.*, 2006; (3) Gu *et al.*, (2002); (4) Krell *et al.*, 2001 e (5) Pereira *et al.*, (2004) e com algumas variações das condições citadas acima. Os ensaios de cristalização foram realizados numa sala com temperatura controlada em aproximadamente 18 – 20 °C.

As gotas foram feitas sobre uma lamínula siliconizada de 22x22mm, na qual foi utilizado 1 µl da solução do reservatório e 1,5 µl da solução de proteína a uma concentração variando de 9,3 à 17 mg.mL⁻¹. Portanto, cada gota apresentava um volume final de 2,5 µl.e foi colocado em cada um dos reservatórios um volume final de 400µl da solução precipitante.

Nos experimentos de cristalização da mtCQ em complexo com ligantes foi utilizado o método de co-cristalização, mas o cristais obtidos não tinham tamanho suficiente para ser feita a coleta dos dados, os problemas encontrados na obtenção e crescimento dos cristais de mtCQ em complexo com os ligantes vão ser discutidos mais adiante neste texto, no tópico 4. E não foi

possível utilizar o método de *soaking*, pois não foi obtido o cristal na forma apo.

3.6.2. Cristalização da corismato sintase

Para cristalização da mtCS em complexo com ligantes tentamos as seguintes condições descritas por: (1) Ahn *et al.*, 2004; (2) Dias *et al.*, 2004; (3) Maclean & Ali, 2003; (4) Quevillon-Cheruel *et al.*, 2004 e (5) condição kit cristalização com algumas variações nas condições citadas acima. Os ensaios de cristalização foram realizados numa sala com temperatura controlada em aproximadamente 18 – 20 °C.

As gotas foram feitas sobre uma lamínula siliconizada de 22x22mm, na qual foi utilizado 1 µl da solução do reservatório e 1,0 µl e 2 µl da solução de proteína a uma concentração 10 mg.ml⁻¹. Portanto, cada gota apresentava um volume final de 2,0 µl e 3,0 µl respectivamente e foi colocado em cada um dos reservatórios um volume final de 400µl da solução precipitante. As gotas de mtCQ complexada com FMN foram feitas em condições que simulavam uma câmara escura, devido a sensibilidade da molécula de FMN a luz.

Nos experimentos de cristalização da mtCS em complexo com ligantes foi utilizado o método de co-cristalização e o método de *soaking* na obtenção dos cristais. Os melhores cristais em complexo foram os obtidos através do método de co-cristalização, os ligantes estavam na proporção de 10 de ligante para 1 de proteína.

3.7. Coleta e processamento de dados de difração por raios X

Durante a coleta dos dados de difração de raios X, os cristais de proteína sofrem um desgaste devido à formação de radicais livres. Sabe-se, entretanto, que o rápido resfriamento dos cristais a uma temperatura próxima à do nitrogênio líquido pode ser usado para atenuar o processo (GARMAN & SCHNEIDER, 1997).

Para a coleta de dados de difração de raios X dos cristais obtidos foi feito o procedimento

usual de congelamento que consiste, inicialmente, em retirar o cristal da gota de cristalização e transferi-lo para uma solução crioprotetora (por exemplo, uma solução à base de glicerol ou etilenoglicol). Em seguida, o cristal deve ser levado rapidamente para o equipamento de raios X, onde um fluxo contínuo de vapor de nitrogênio (100 K) incidirá sobre o cristal durante toda a coleta dos dados (Figura 16).

Isto permite a redução dos efeitos danosos da radiação direta e aumenta o tempo de vida do cristal. Dessa forma há um melhor aproveitamento do material, pois apenas um único cristal pode fornecer dados suficientes para a resolução da estrutura e não é preciso realizar a integração com dados de cristais diferentes.

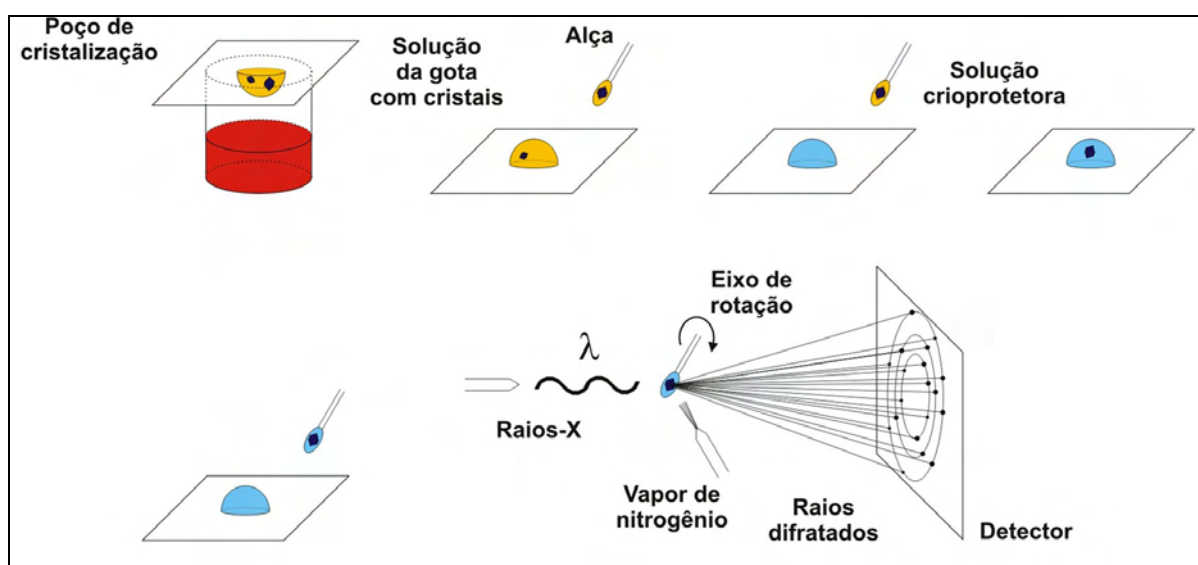


Figura 16. Procedimento de crio-cristalização de um cristal de proteína, para coleta de raios X. Figura retirada do texto Nangem, 2003.

Os conjuntos de dados foram coletados usando-se como fonte radiação síncrotron (Estação PCr, Laboratório Nacional de Luz Síncrotron, LNLS, Campinas, Brasil) (POLIKARPOV *et al.*, 1998a; POLIKARPOV *et al.*, 1998b), e um detector CCD (MARCCD 165 mm). A distância do detector ao cristal foi regulada de forma a chegar o mais próximo possível da melhor resolução, limitada pela qualidade do cristal. O protetor criogênico utilizado era composto de glicerol 20%, 80% da solução tampão da mtCS e a coleta foi mantida a uma temperatura de 104 K.

Os dados foram pré-processados utilizando o programa MOSFLM (LESLIE, 1992) que determinou uma resolução inicial para os cristais. Os dados foram escalonados com o programa SCALA, ambos programas integrantes do CCP4 (COLLABORATIVE COMPUTACIONAL PROJECT N° 4, 1994).

3.7.1. Coletas de dados da enzima corismato sintase

Para coleta foram utilizados os cristais conseguidos na condição (5) que estavam na forma apo e (4) que estavam na presença do CA, fizemos *soaking* e montamos os seguintes complexos mtCS-FMN, mtCS-CA e mtCS-FMN-CA e deixamos por 4 horas, mas nos complexos onde havia somente o FMN utilizando a condição (5) o cristal se dissolveu depois do *soaking*.

Os parâmetros utilizados na coleta de dados de difração de raios X para os complexos mtCS-AC e mtCS-FMN-AC, estão apresentados na tabela 1.

Tabela 1. Parâmetros utilizados na coleta de dados de difração de raios X.

	mtCS-AC	mtCS-FMN-AC
Número de imagens	160	120
Comprimento de Onda (graus)	1,431	1,425
Tempo de exposição	30	60
Ângulo de oscilação (graus)	1	1
Distância do detector (mm)	85	100

3.8. Substituição molecular e Refinamento cristalográfico

Muitas estruturas podem ter o problema da fase resolvido utilizando uma estrutura homóloga e o método de substituição molecular. Nesse método, a estrutura homóloga encontrada em um banco de dados de estruturas de proteínas é encaixada na cela unitária da estrutura desconhecida e as fases são utilizadas como um esboço inicial da estrutura desconhecida.

O modelo inicial (também chamado modelo de busca) a ser usado na procura deve ser o mais idêntico possível da macromolécula a ser estudada. No caso de proteínas, uma identidade igual ou superior a 50% de identidade deve ser suficiente para a solução desta etapa. O maior problema em utilizar modelos com baixa identidade está na estimativa grosseira das fases do modelo a ser estudado, o que pode causar problemas durante a etapa de refinamento da estrutura a ser resolvida. Assim, a utilização de modelos de busca de altíssima identidade com a macromolécula alvo é fortemente indicada (RUSSO *et al.*, 1996).

A estrutura cristalográfica da mtCS foi determinada pelo método de substituição molecular usando o programa AmoRe (NAVAZA, 1994), tendo sido usado como modelo de busca a estrutura da corismato sintase de *Mycobacterium tuberculosis* resolvida anteriormente (código de acesso PDB: 2G85) (DIAS *et al.*, 2004). As posições atômicas obtidas por substituição molecular foram usadas para iniciar o refinamento cristalográfico.

Para o refinamento da estrutura foi utilizado o programa REFMAC, que faz um refinamento posicional (LASKOWSKI *et al.*, 1993). O refinamento é o processo de encontrar uma melhor concordância entre o modelo proposto para a estrutura da molécula e a sua estrutura real. Esta concordância deve refletir numa igualdade entre os fatores de estrutura calculados (F_{calc}) e os fatores de estrutura observados (F_{obs}).

O acompanhamento da qualidade deste processo é feito por meio do cálculo do fator R (R_{factor}) durante o processo. Todas as etapas do refinamento foram monitoradas nos arquivos *.log* do CCP4, com ênfase para os valores da convergência dos fatores de estrutura para valores compatíveis para estruturas de qualidade estrutural ($R_{\text{factor}} < 20\%$ e R_{free} o mais próximo do valor do R_{factor}).

O modelo final da mtCS foi avaliado quanto aos vários parâmetros de qualidade estereoquímica, utilizando o programa PROCHECK (LASKOWSKI *et al.*, 1994) e WHATCHECK (HOOFT *et al.*, 1996). Modelos atômicos foram superpostos usando o programa LSQKAB, todos

os programas utilizados citados acima com exceção do programa WHATCHECK são do pacote CCP4 (COLABORATIVE COMPUTATIONAL PROJECT N° 4, 1994).

Os mapas de densidade eletrônica foram visualizados com programa XtalView/Xfit (McREE, 1999) e Coot pertencente ao pacote do CCP4 (COLABORATIVE COMPUTATIONAL PROJECT N° 4, 1994), nos quais também foram realizada a “reconstrução” manual da estrutura de alguns resíduos de aminoácidos que apresentaram densidade e que no molde não tinha estes resíduos de aminoácidos.

3.9. Espectroscopia de Dicroísmo circular (CD)

A conformação de uma molécula está intimamente relacionada à sua atividade ótica e, considerando estas características, a técnica de CD é utilizada para estimar a quantidade de estrutura secundária de proteínas e avaliar as mudanças conformacionais ocasionadas por ligantes, agentes desnaturantes (acidez, uréia, cloridrato de guanidina – Gdn-HCl, etc) e temperatura (WOODY, 1995; KELLY & PRICE, 1997).

Em proteínas, os principais grupos opticamente ativos são as ligações amida da cadeia peptídica, monitoradas na região do ultravioleta (UV) distante [190-250 nm]; as cadeias laterais aromáticas e as ligações dissulfeto, monitoradas no UV próximo [250-360 nm]. Portanto, o espectro de CD entre 190 e 250 nm (região ultravioleta distante) pode identificar diferentes tipos de estrutura secundária como hélices, folhas, voltas (*loop*) e estruturas desordenadas (*coil*) (WOODY, 1994; VENYANINOV & YANG, 1996).

Esta avaliação pode ser feita com o auxílio de programas computacionais que avaliam o espectro de CD de uma proteína. Estes programas comparam o espectro a ser analisado com um banco de dados de espectros de CD de proteínas que já possuem suas estruturas em alta resolução determinadas. Este procedimento aumenta a precisão na estimativa da quantidade de cada tipo de estrutura (GREENFIELD, 1996). A avaliação da estabilidade de uma proteína, por CD, pode

fornecer informações sobre a presença de domínios com diferentes estabilidades e sobre processos dinâmicos, como interações com ligantes (WOODY, 1995).

Os experimentos de CD foram realizados neste trabalho, foram feitos utilizando um espectropolarímetro Jasco J-710 (Departamento de Física – IBILCE/UNESP) com temperatura controlada. O programa utilizado para registro e tratamento dos dados foi o Spectra Manager (JASCO), com o auxílio do programa Origin Pro 7.0 (OriginLab Corporation).

3.9.1. CD para estimar as frações de estrutura secundária das enzimas

A concentração das enzimas foram determinadas por espectrometria, usando um coeficiente de extinção 3840 para a mtCQ, sua concentração foi de 21,7 mg.ml⁻¹ e um coeficiente de extinção 26030 para a mtCS, sua concentração foi de 4,6 mg.ml⁻¹.

Os parâmetros utilizados na coleta dos espectros de dados do CD para a mtCQ e mtCS na forma apo e nos complexos com ligantes, estão apresentados na tabela 2.

Tabela 2. Parâmetros utilizados na coleta dos espectros de dados de CD para as enzimas mtCQ e mtCS, para estimar as frações de estrutura secundária.

	mtCQ	mtCS
Concentração da proteína (mg.ml ⁻¹)	5	2,5
Caminho óptico	0,06 mm	3,0 cm
Comprimento de onda (nm)	190-260	198-260
Sensibilidade (mdeg)	50	50
Velocidade de leitura (mm/min)	50	50
Largura da banda (nm)	2,0	1,0
Tempo de resposta (seg)	2,0	1,0
Resolução (nm)	0,5	0,2
Total de acumulações*	10	8

* Foi realizada uma média das acumulações, para obtenção de um espectro médio final.

3.9.2. Estudos de estabilidade térmica das enzimas por CD

A concentração das enzimas foram determinadas por espectrometria, usando um coeficiente de extinção 3840 para a mtCQ, sua concentração foi de 37,6 mg.ml⁻¹ e um coeficiente de extinção 26030 para a mtCS, sua concentração foi de 4,6 mg.ml⁻¹.

Os parâmetros utilizados na coleta dos espectros de dados do CD para a mtCQ na forma apo e nos complexos com ligantes com o controle da variação da temperatura sendo feito manualmente, estão apresentados na tabela 3.

E os parâmetros utilizados na coleta dos espectros de dados do CD para a mtCQ na forma apo e nos complexos com ligantes com o controle da variação da temperatura sendo feito pelo aparelho, estão apresentados na tabela 4.

Tabela 3. Parâmetros utilizados na coleta dos espectros de dados de CD para a enzima mtCQ, para avaliar a estabilidade térmica.

	mtCQ
Concentração da proteína (mg.ml ⁻¹)	0,017
Caminho óptico	3,0 cm
Comprimento de onda (nm)	190-260
Sensibilidade (mdeg)	50
Velocidade de leitura (mm/min)	50
Largura da banda (nm)	2,0
Tempo de resposta (seg)	2,0
Resolução (nm)	0,5
Total de acumulações*	10

*Foi realizada uma média das acumulações, para obtenção de um espectro médio final.

Tabela 4. Parâmetros utilizados na coleta dos espectros de dados de CD para as enzimas mtCQ e mtCS, para avaliar a estabilidade térmica.

	mtCQ	mtCS
Concentração da proteína (mg.ml ⁻¹)	0,017	0,017
Caminho óptico	3,0 cm	3,0 cm
Comprimento de onda (nm)	222	222
Sensibilidade (mdeg)	50	50
Velocidade de leitura (min)	0-150	0-150
Largura da banda (nm)	2,0	1,0
Tempo de resposta (seg)	1,0	1,0
Resolução (seg)	10	10
Total de acumulações*	1	10

* Foi realizada uma média das acumulações, para obtenção de um espectro médio final.

Para os experimentos de CD com a mtCQ foi necessário diluir 10 vezes a concentração do tampão de Tris HCl 2mM (pH 7,4) e KCL 4mM para que o mesmo não causasse interferência nas medidas dos espectros.

3.10. Espectroscopia de fluorescência

A espectroscopia de emissão de fluorescência é muito utilizada para caracterizar macromoléculas biológicas e verificar mudanças conformacionais discretas no microambiente, baseando-se na emissão da radiação eletromagnética devido a transições eletrônicas entre estados eletrônicos excitados e estados de menor energia. As proteínas contêm fluorescentes naturais (fluoróforos) como Trp, Tyr e Phe (LAKOWICZ, 1983).

A fluorescência destes resíduos é altamente sensível ao ambiente em que se encontram, portanto, é possível monitorar suas alterações por parâmetros como posição do máximo de emissão

(λ_{max}) e rendimento quântico (ϕF) devido à ação de solventes, ligantes, supressores, íons ou agentes solubilizantes. Estas pequenas mudanças conformacionais induzem resultados distintos no espectro de emissão da fluorescência, no rendimento quântico ou no deslocamento do máximo de emissão (LAKOWICZ, 1983).

Por exemplo, quando o resíduo de Trp se encontra bem exposto ao solvente, ele gasta uma quantidade maior de energia no estado excitado com a reorganização das moléculas de água que o envolvem e emite fluorescência em menos energéticos (próximos a 355 nm). Já quando o resíduo de Trp se encontra em uma região no interior da proteína, geralmente hidrofóbico e com baixa acessibilidade a água, a emissão de fluorescência ocorrerá em comprimentos de onda mais energéticos (310 nm a 335 nm), pois não haverá grande perda de energia com a reorganização de moléculas de água. A mudança de exposição do Trp ao ambiente pode ser induzida por ligantes ou pela desnaturação e fornecem informações sobre as mudanças conformacionais e sobre a estabilidade da proteína (LAKOWICZ, 1983).

Os experimentos de fluorescência foram realizados utilizando um fluorímetro RF-5301 (SHIMADZU) do Departamento de Química – IBILCE/UNESP, a temperatura controlada de 25°C, em cubetas de quartzo de 1 cm de caminho óptico. O comprimento de onda de excitação foi 270 nm, portanto excitando-se somente os resíduos de tirosina e triptofano, e a emissão monitorada de 280 a 450 nm.

A mtCS contem 07 tirosinas e 03 triptofanos, a maioria destes resíduos de aminoácidos aromáticos se localiza na região do sitio de ligação do FMN (co-fator) e EPSP (substrato) (figura 17) e por isso podemos acompanhar de uma forma mais detalhada o comportamento da mtCS quando ocorre a ligação destes ligantes no sitio, pois a tendência é que esta região que é predominantemente de *coil* e portanto uma região randômica, se compacte com a entrada dos ligantes e com isso diminua e/ou mude a faixa do espectro de fluorescência pois estes aminoácidos aromáticos deixariam de estar totalmente ou parcialmente expostos ao meio.

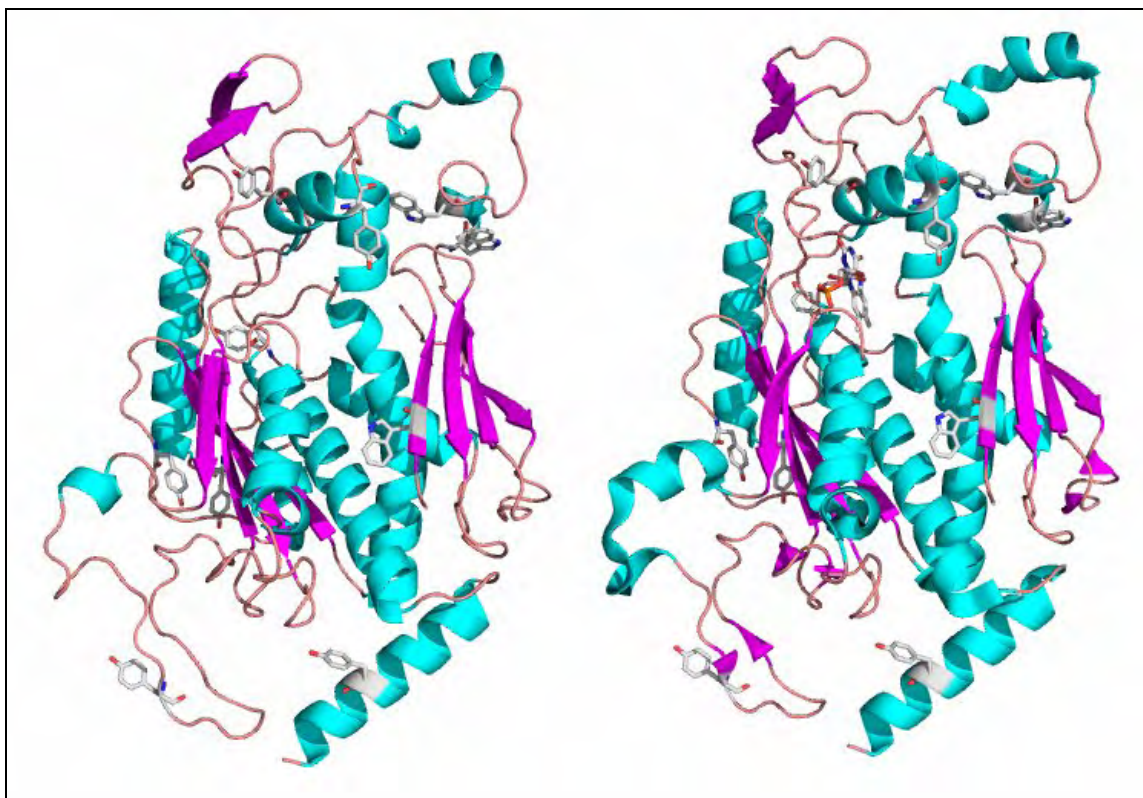


Figura 17. A enzima mtCS na presença dos ligantes FMN (co-fator) e EPSP (substrato) do lado direito e a enzima mtCS na forma apo do lado direito. Ambas as estruturas com os aminoácidos aromáticos em destaque na forma de *ball stick*. As coordenadas atômicas da mtCS e dos ligantes foram obtidas no PDB (código de acesso: 2G85, 2O12 e 2QHF).

A concentração da proteína foi determinada por espectrometria, usando um coeficiente de extinção 26030, sua concentração foi de $1,0 \text{ mg}\cdot\text{ml}^{-1}$, mas para coletar os espectros na forma apo e na presença dos ligantes foi usada a proteína mtCS na concentração de $0,02\text{mg}\cdot\text{ml}^{-1}$.

Para o experimento foram coletados os picos no espectro de fluorescência para os ligantes e para os sais, para que posteriormente estes espectros da titulação dos ligantes e do sais fossem deduzidos dos espectros da enzima para verificar mudanças conformacionais mesmo que discretas somente na enzima.

Para obter este valor titulamos os ligantes em várias concentrações, para o sal: Formato de Sódio 1M e Bis Tris 0,04M, Formato de Sódio 2M e Bis Tris 0,07M, Formato 3M e Bis Tris 0,1M e para os ligantes CA e FMN: $0,0024\text{mM}$ (5X), $0,0048\text{mM}$ (10X), $0,0072\text{mM}$ (15X), $0,0096\text{mM}$ (20X) e $0,012\text{mM}$ (25X). Foi observado um valor máximo de $\lambda=312\text{nm}$ e um outro menor de

$\lambda=356\text{nm}$ para os sais, $\lambda=315\text{nm}$ para o AC e $\lambda=306\text{nm}$ para o FMN, como mostrado na figura 18.

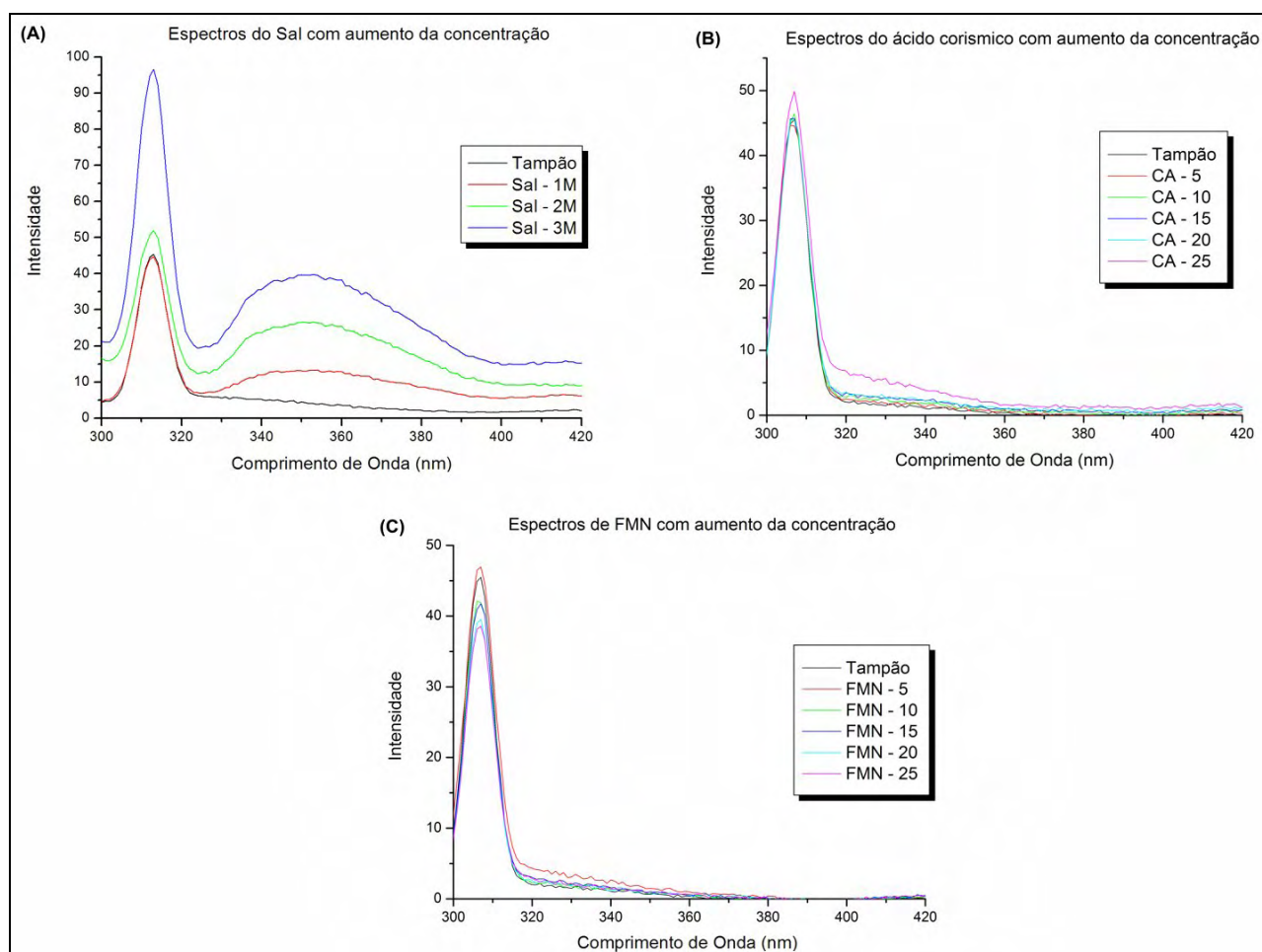


Figura 18. Espectros de titulação. (A) sal em molar, (B) ácido corísmico em proporção de vezes, (C) FMN em proporção de vezes.

3.11. Espalhamento de raios X a baixo ângulo

O espalhamento de raios X a baixo ângulo (SAXS) é uma ferramenta fundamental no estudo de macromoléculas biológicas em solução que variam entre poucos quilo daltons (KDa) a vários mega daltons (MDa) (FEIGIN & SVERGUN, 1987). A maior vantagem deste método está na sua habilidade de dar informações estruturais sobre sistemas parcialmente ou completamente desordenados (SVERGUN & KOCH, 2003).

O SAXS permite estudar a estrutura de partículas nativas em ambientes próximos ao fisiológico e analisar mudanças estruturais em resposta a variações nas condições externas

(SVERGUN & KOCH, 2003). Determinando várias características das macromoléculas, tais como raio de giro, peso molecular, volume, grau de hidratação, diâmetro máximo e forma (GLATTER & KRATKI, 1982).

A obtenção de dados no experimento consiste em direcionar um feixe de radiação para a amostra e medir a variação de intensidade espalhada em função do ângulo de espalhamento. Essa radiação deve possuir comprimento de onda bem definido na faixa de raios X, entre 0,5 e 2 Å, ter uma boa colimação e alta intensidade (LINDER & ZEMB, 1991).

O espalhamento de raios X a baixos ângulos somente é observado, quando as amostras analisadas apresentam heterogeneidade em sua densidade eletrônica e com dimensões características da ordem de 10 a 1000 Å (GLATTER & KRATKY, 1982).

A densidade eletrônica ρ é definida como a quantidade em mols de elétrons por unidade de volume. Quando partículas estão no vácuo, a amplitude de luz espalhada é proporcional à densidade eletrônica das partículas. Porém, quando partículas analisadas estão dispersas em uma matriz, a diferença entre a densidade eletrônica da partícula e da matriz é que influencia na amplitude de luz espalhada. Essa diferença é chamada de contraste de densidade eletrônica e é um dos principais parâmetros da amostra quando a técnica de SAXS é utilizada (FEIGIN & SVERGUN, 1987).

A função $\rho(r)$ é chamada de função de distribuição de distâncias (GLATTER & KRATKY, 1982). Ela possui uma definição geométrica que pode ser entendida dividindo-se uma partícula em um grande número de elementos de volumes pequenos e idênticos. A função $\rho(r)$ descreve a distribuição de distâncias r que podem ser encontradas a partir de combinação de qualquer par de elementos de volume (GLATTER & KRATKY, 1982).

Em princípio, a função de distribuição de distâncias $\rho(r)$ contém a mesma informação da intensidade de espalhamento $I(s)$, mas a representação do espaço real é mais intuitiva e a informação sobre a forma da partícula pode freqüentemente ser deduzida diretamente por uma inspeção visual da $p(r)$ (GLATTER & KRATKY, 1982).

Variações na forma, tamanho e estrutura das partículas do sistema analisado, são refletidas na função de espalhamento $I(q)$ (a intensidade de onda espalhada em função do espaço recíproco) e na função de distribuição de distâncias $\rho(r)$; portanto, a monodispersividade pode então ser determinada a partir das curvas de espalhamento obtidas. Para sistemas polidispersos, a distribuição de tamanhos pode ser analisada, assumindo-se que todas as partículas possuem a mesma forma (LINDER & ZEMB, 1991).

O raio de giro (R_g) é definido como o raio de uma fina camada esférica de mesma massa e momento de inércia (GLATTER & KRATKY, 1982). Este parâmetro é uma medida da distribuição das cargas em relação ao centro de massa eletrônico de uma partícula e pode ser obtido através da função de distribuição de distâncias $\rho(r)$ ou diretamente pela curva de espalhamento (GLATTER & KRATKY, 1982). O conhecimento do raio de giro pode ser muito útil na investigação estrutural de uma partícula em solução, pois um aumento ou diminuição do mesmo, provocado por qualquer variação nas condições da solução, indicaria uma mudança nas posições relativas entre os átomos que compõe a partícula (LINDER & ZEMB, 1991).

A visualização e manipulação de estruturas atômicas e modelos de baixa resolução, a partir de modelos de alta resolução, é obtida através de programas como, por exemplo, o CRY SOL (SVERGUN *et al.*, 1995) e o GNOM (SVERGUN, 1992), entre outros. Quando não se têm modelos de alta resolução para comparação, parte-se para uma análise mais cuidadosa, utilizando-se a função $\rho(r)$ para inferir a curva experimental dada pelos programas CREDO (PETOUKHOV *et al.*, 2002), DAMAVER (SVERGUN *et al.*, 2001), DAMIM (SVERGUN, 1999), entre outros.

Foi utilizada a técnica de SAXS para estudar mtCS na sua forma apo em solução com o intuito de obter um modelo tridimensional e assim determinar a caracterização e o comportamento desta proteína em solução. A mtCS já possui dados cristalográficos depositados no PDB com o código de acesso 1ZTB e 2G85, porém foi resolvido apenas como monômero, deixando a dúvida sobre como essa proteína se encontra em seu ambiente fisiológico.

As curvas de espalhamento foram coletadas no laboratório de biocristalografia do departamento de física da UNESP/IBILCE, utilizando de uma fonte de raios X, constituída por um ânodo rotatório de cobre da marca RIGAKU DENKY RU, com comprimento de onda característico ($\text{CuK}_\alpha - 1,54\text{Å}$) e potência de operação de 4 KW, o alinhamento do feixe de raios X foi feito com um sistema de colimação por blocos, este equipamento é dotado de uma câmara de Kratky com um porta amostra adaptado no interior da câmara e assim o espalhamento da amostra é totalmente no vácuo com isso diminuindo o espalhamento parasita e a absorção de raios X pelas moléculas de ar.

A amostra foi colocada em um capilar de vidro Lyndemann de 1 mm de diâmetro interno e encaixado no porta amostra na posição horizontal, a intensidade de espalhamento foram registrada através de um detector sensível a posição Tennelec PSD-1000 situado a 210 mm da amostra com 1024 canais, as intensidades foram amplificadas, discriminadas, analisadas e convertidas em contas por segundo através do sistema MCA modelo 916 da (EG&G ORTEC). As soluções de proteínas com o tampão (Tris HCl 50 mM – pH 7,5) foram expostas por um período de 2 horas diante do feixe de raios X.

A amostra de mtCS estava em três diferentes concentrações 15mg/ml, 10mg/ml e 5mg/ml, com o intuito de obter a curva extrapolada a concentração zero as curvas foram coletadas em temperatura ambiente 25°C com uma faixa angular de $0,02814 \text{ Å}^{-1} < q < 0,2955 \text{ Å}^{-1}$, onde q é o vetor de espalhamento definido como $q = 4\pi \sin\theta/\lambda$ e 2θ representa o ângulo de espalhamento.

Os dados de SAXS foram devidamente tratados e a correção das curvas de espalhamento por efeitos geométricas do feixe, assim como o cálculo da função de distribuição de distância $\rho(r)$ foi obtido através da aplicação da função de transformada indireta de Fourier utilizando do programa GNOM (SVERGUN, 1992). O raio de giração foi obtido através da aproximação da lei de Guinier (GUINIER & FOURNET, 1955) para valores de q tendendo a zero mediante a inclinação α da região linear do gráfico de $\log I$ versus q^2 . O volume e área da superfície externa da proteína foram calculados pelas equações $V = 2\pi^2 I(0)/Q$ e $S = \pi V \lim(I(q)q^4)/Q$ respectivamente, onde Q

representa o invariante $Q = \int_0^{\infty} I(q)q^2 dq$, a intensidade da curva extrapolada $I(0)$ foi obtida da lei de

Guinier.

A determinação da forma do envelope externo da mtCS foi gerada pelo método *ab initio* através do programa DAMIMM (SVERGUM, 1999), foram calculados 10 modelos independentes com o propósito de fazer uma media e obter o melhor modelo esse processo foi feito usando-se o programa DAMAVER (VOLKOV & SVERGUN, 2003). Finalmente a sobreposição entre o modelo dos dados cristalográficos (PDB) e as coordenadas dos átomos dummies obtido por SAXS foi feito com o programa SUPCOMB20 (PETOUKHOV *et al.*, 2002).

4. Resultados e Discussão

4.1. Desenvolvimento da ferramenta para modelagem molecular comparativa e análise em larga escala.

A ferramenta para modelagem molecular comparativa e análise em larga escala desenvolvida teve como finalidade de viabilizar a modelagem de forma recursiva e sem a intervenção manual para ajustes de parâmetros como alinhamento e seleção de moldes.

4.1.2. Implementação e formatação dos dados que compõe o banco de dados SKPDB

As seqüências primárias dos moldes depositados no PDB e as seqüências primárias dos alvos obtidas no Swiss-Prot/TrEMBL foram formatadas, filtradas e anotadas através de um *script* em PERL para posteriormente serem adicionadas a base de dados. A busca feita nestas bases de dados foi feita dia 12/04/2007.

Cada uma das seqüências foi submetida a um *script* desenvolvido em PERL, o qual utiliza as seqüências primárias de todas as estruturas depositadas no PDB e as alinham contra cada seqüência primária alvo, criando um arquivo de saída com todos os moldes que tiveram identidade superior a 30%, discriminando a identidade entre os pares de seqüências, o código de acesso PDB dos moldes, as cadeias monoméricas associadas à proteína alvo, os escores de alinhamento e as similaridades, para cada seqüência primária alvo.

Após a realização do alinhamento, um outro *script* é chamado e entra em cada arquivo texto criado, identificando o molde de maior identidade e aciona automaticamente o acesso ao *site* do PDB e executa o *download* das coordenadas atômicas do molde no PDB, e formata este arquivo deixando somente as coordenadas atômicas da cadeia escolhida durante o alinhamento e retira todo o cabeçalho, todas as águas, ligantes ou qualquer anotação que esteja no arquivo. Antes de entrar no próximo arquivo texto e executar o mesmo procedimento o *script* verifica se aquele molde já não foi baixado anteriormente, se o molde já foi baixado, o *script* pula para o arquivo texto seguinte caso contrário executa a baixa das coordenadas atômicas do molde.

Depois o programa gera os protocolos que serão utilizados pelo próximo passo da modelagem molecular comparativa que é a construção do modelo, através do programa MODELLER (ŠALI & BLUNDELL, 1993).

Um total de 3947 enzimas pertencentes a via foram encontradas no banco de dados Swiss-Prot/TrEMBL, destas seqüências um total de 45 já têm estruturas resolvidas por algum tipo de método experimental e estão depositadas no PDB. E 3438 enzimas foram possíveis de serem modeladas, por apresentarem identidade seqüencial superior a 30%, os resultados do alinhamento para cada enzima são mostrados na tabela 5.

Tabela 5. Resultado do alinhamento feito para as enzimas encontradas no Swiss-Prot contra o PDB, para encontrar moldes para as enzimas.

Enzimas	DAHPS	DHQD	DHQS	SDH	EPSPS	SK	CS
Encontrado molde	624	508	378	434	434	497	570
Não encontrado molde	20	160	62	93	78	42	2
Já resolvidas	8	10	3	7	5	6	6
Total	652	678	443	534	517	545	578

Todo o processo de busca e seleção de moldes foi executado particionando o arquivo de seqüências pelo número de nós do *cluster* (16) e disparando processos para iniciar o alinhamento de cada parte do arquivo com todas as seqüências primárias dos moldes e baixando as coordenadas atômicas do molde que será utilizado durante a modelagem.

A ferramenta desenvolvida para modelagem molecular comparativa em larga escala só foi possível com a implementação dos processos de modelagem utilizando a tecnologia de *clusters*, que integrou a capacidade de gerar dados e analisá-los com a rapidez necessária ao desenvolvimento da pesquisa. A partir deste passo da modelagem em larga escala, os processos são executados utilizando as rotinas de modelagem que rodam paralelizadas.

Estas rotinas visam agilizar todo o processo de modelagem molecular e análise de estruturas 3D de proteínas. Esta agilidade é obtida de duas formas: integrar automaticamente todas as etapas

do processo de modelagem molecular para que não haja nenhuma intervenção do usuário no decorrer deste processo e paralelizar a etapa de modelagem para que possa se obter uma diminuição significativa no seu tempo de processamento.

A paralelização da modelagem molecular comparativa utilizando o MODELLER foi realizada utilizando uma biblioteca de linguagem C, o MPI (*Message Passing Interface*), que controla a distribuição dos processos de modelagem pela divisão dos arquivos de entrada do MODELLER no *cluster Beowulf*. Isto permite paralelizar a execução do MODELLER e diminuir o tempo de processamento das rotinas de modelagem.

A arquitetura do *cluster Beowulf* utilizada no Laboratório de Sistemas Biomoleculares foi inteiramente projetada aos interesses da bioinformática estrutural, adaptando todos os programas e construindo ferramentas que tornassem o *cluster* o mais específico possível para a pesquisa estrutural de proteínas.

Após a seleção dos moldes e a criação das entradas para realizar a modelagem, é executado o alinhamento molde-alvo, do qual é obtido o alinhamento a ser utilizado no terceiro passo da modelagem como arquivo de entrada para o MODELLER. Antes de iniciar o passo que executará a construção dos modelos, um *script* verifica se há alguma modelagem sendo realizada no momento. Isto porque, caso esteja sendo realizada alguma modelagem e outra modelagem for submetida, ocorrerá uma perda no desempenho do programa. Assim, é verificada a existência da execução de alguma modelagem.

Caso haja alguma modelagem sendo executada, então os parâmetros de entrada necessários à execução do programa serão gravados em uma fila de espera e os arquivos que o usuário submeteu permanecerão no diretório que foi criado. Ao término de cada modelagem, o programa verificará a existência de alguma modelagem nesta fila. Caso não haja modelagens sendo feitas, o *script* executará um programa em C implementado com rotinas MPI, iniciando a execução da modelagem distribuindo o número total de modelos solicitados pelos 16 nós para o próximo

elemento da fila. Para cada enzima da via do ácido chiquímico, foi gerado e analisado um total de 1000 modelos, ampliando o espectro de análise com o objetivo de se obter melhores modelos.

Logo que a modelagem é finalizada, a avaliação de cada modelo é feita automaticamente utilizando programas como: PROCHECK (LASKOWSKI *et al.*, 1993), WHATCHECK (HOOFT *et al.*, 1996), VERIFY3D (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992) e X-PLOR (BRÜNGER, 1992) para avaliarmos o RMSD da geometria ideal de cada proteína. Algumas das propostas dos programas PROCHECK e WHATCHECK são (i) determinar erros grosseiros nas estruturas, tais como cadeias laterais deslocadas, (ii) checar anormalidades locais da estereoquímica e (iii) produzir critérios para a qualidade estereoquímica global. Os parâmetros de cortes estabelecidos nesta etapa foram:

- Análise estereoquímica usando o programa PROCHECK: as porcentagens das regiões mais favoráveis do gráfico de Ramachandran tinham que estar acima de 90% e os valores para o G-factor tinham que estar acima de -0.5.
- Análise da compatibilidade da estrutura 3D com sua seqüência primária usando o programa VERIFY 3D: foi considerado um escore próximo ou acima de 45%, pois abaixo deste valor às estruturas poderiam conter erros em seu dobramento.
- Análise do RMSD da geometria ideal usando o programa X-PLOR: foram utilizados os valores de $< 0,06$ para o comprimento de ligação em Å e < 10 para o ângulo de ligação em graus como parâmetro de corte para garantir a qualidade da estrutura.

Esses softwares de avaliação química de modelos gerados por modelagem molecular comparativa, nos dão maior confiabilidade nas estruturas geradas, possibilitando propor simulações de *docking* contra bibliotecas de ligantes (*screening* virtual) com o objetivo de selecionar alvos terapêuticos para desenho de drogas baseado em estrutura. Todas as análises feitas para os modelos gerados das enzimas que estão no banco SKPDB e podem ser salvas através dos *links* existentes na página com o resultado das análises.

A figura 19 mostra um fluxograma da formatação dos dados e da integração feita para construção da base de dados SKPDB.

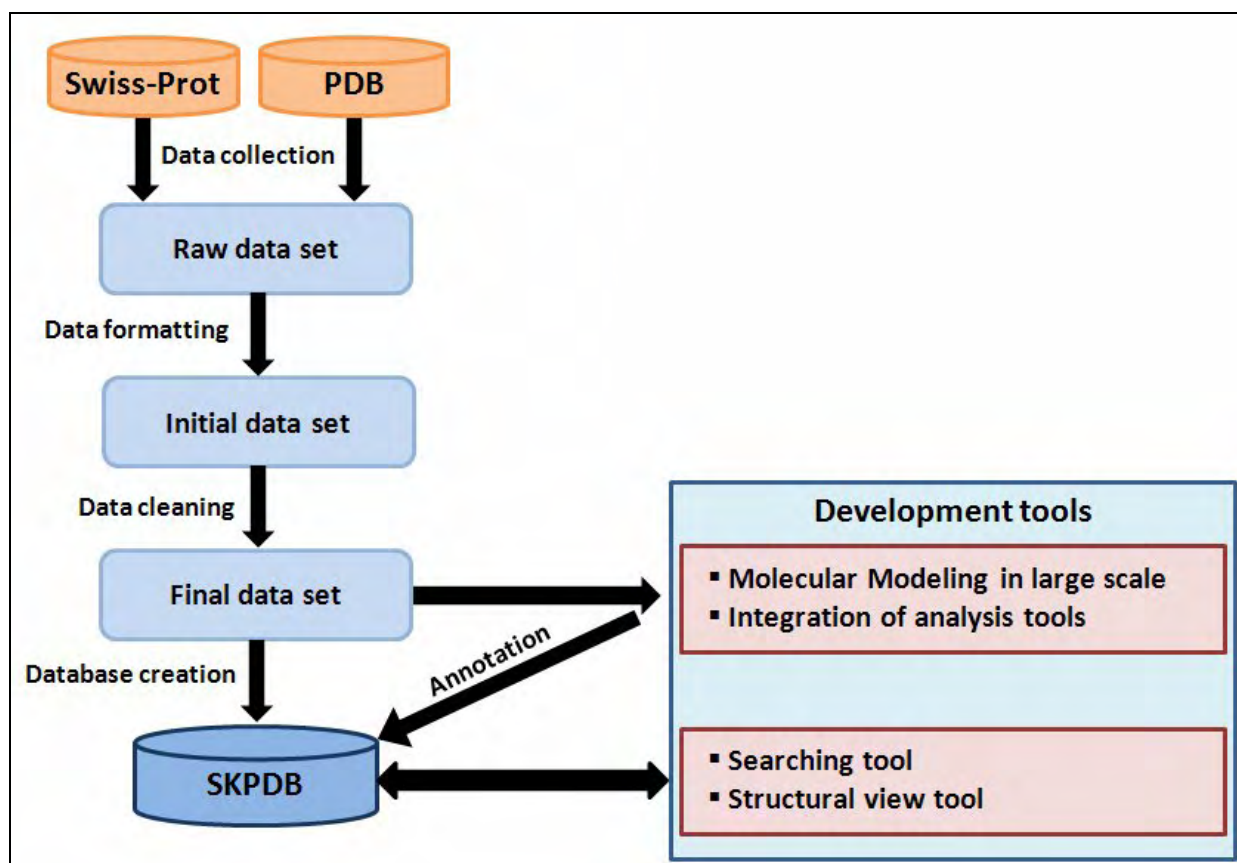


Figura 19. Fluxograma da formatação dos dados e construção do banco de dados SKPDB.

4.2. SKPDB

Neste tópico será apresentado o banco de dados SKPDB que oferece interfaces gráficas para visualizar as análises dos modelos gerados, informações sobre as seqüências primárias (alvo/molde) e sobre os moldes utilizados para modelagem molecular comparativa em larga escala e *downloads* dos dados.

4.2.1. Implementação e descrição do banco de dados SKPDB

Desde o advento da *World Wide Web*, os bancos de dados biológicos se tornaram uma parte vital da literatura biológica. Saber localizá-los e fazer *download* de informações dos repositórios

centrais de dados biológicos é atualmente uma habilidade tão importante para os pesquisadores quanto à pesquisa na literatura tradicional.

O banco de dados SKPDB foi implementado em MySQL versão 4.0.20 (<http://www.mysql.com>), um sistema de gerenciamento de banco de dados de médio desempenho, relativamente simples e disponível gratuitamente para uso não comercial e sua interface gráfica em CGI, a integração entre o banco de dados e o CGI foi escrita em PERL versão 5.8.4 (Figura 20).

O banco de dados foi desenvolvido em um sistema operacional Linux-Debian, em um AthlonXP 2100+ com 80 Gb de HD, 1 Gb RAM e placa de rede 3Com de 100 Mbits, tendo grande capacidade para armazenamento de dados e rapidez nos processos de busca e aquisição das informações e foi suprido com as informações obtidas sobre as seqüências primárias das enzimas da via, com a análise dos resultado dos modelos gerados pela modelagem em larga escala.

A figura 20 mostra um desenho esquemático da integração entre CGI e o banco de dados, utilizando a linguagem de programação PERL para interação com o banco de dados criado para as enzimas da via metabólica do ácido chiquímico identificada nos genomas de microorganismos e plantas completos e a forma a qual o CGI solicita as informações ao servidor, organizando suas respectivas enzimas.

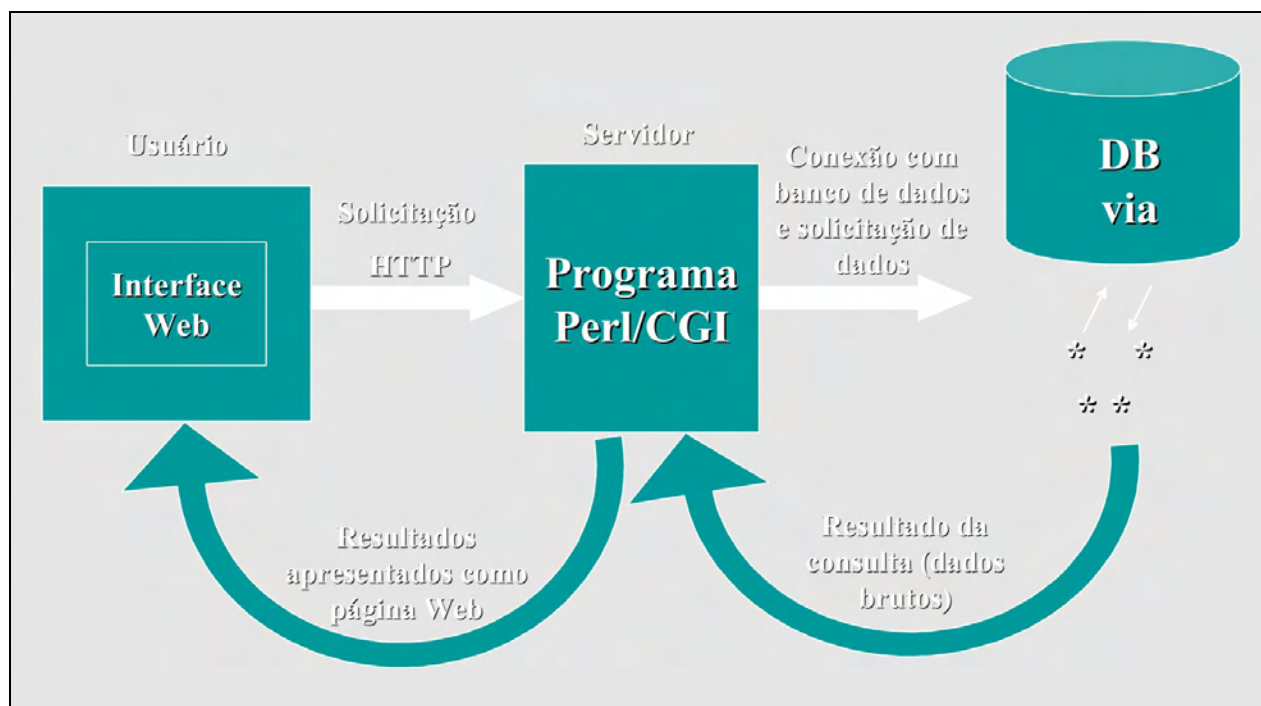


Figura 20. Diagrama esquemático de como a programação CGI interage com o banco de dados SKPDB, organizando e filtrando os dados devolvidos ao usuário.

Foram instalados e configurados alguns softwares para que os dados solicitados pelos usuários fossem apresentados de forma interativa, por exemplo: na página onde se tem acesso ao modelo gerado ou ao molde utilizado. Um dos programas instalados foi o *Jmol* que é um programa que serve para visualização de moléculas, o usuário pode girar, aumentar, diminuir e escolher qual a forma de representação da estrutura tridimensional que deseja ver, entre outras facilidades que o programa fornece. O programa pode ser baixado no site: <http://jmol.sourceforge.net>, é gratuito e funciona em vários tipos de sistema operacional, como Linux, Windows, Mac, entre outros.

O programa *Jmol* necessita da máquina virtual java (JVM - <http://java.sun.com>) para poder funcionar, pois seu código fonte é baseado na linguagem Java, a versão instalada foi 1.5. O JVM é um mecanismo que permite executar código Java em qualquer plataforma, podemos entender JVM como “uma máquina imaginária implementada via software ou hardware que executa instruções vindas de *bytecodes*”.

A página que apresenta a análise dos modelos disponibiliza dinamicamente os resultados em relação ao modelo gerado, apresentando via *web* relatórios e gráficos das análises feitas para o G-

factor, RMSD da geometria ideal e o gráfico de Ramachandran. Estas informações sobre a qualidade da estrutura são indispensáveis para a utilização da estrutura em simulações de *docking* caso seja esta a necessidade do usuário. Todos os *links* externos no SKPDB fornecem informações adicionais em relação a seqüência primária e anotações funcionais e estruturais sobre as enzimas.

4.2.2 Tabelas e Relacionamentos

Para o armazenamento dos dados foi construída uma base de dados contendo as seguintes tabelas: *sequence*, *model*, *template* e *analysis*. Os dados foram inseridos nas tabelas através de outros *scripts* em Perl e do uso de módulos específicos desta linguagem (DBD::mysql) que permitem a interação com a base de dados, dentro de um esquema (estrutura) proposto, que reflete as relações existentes entre os seus diferentes componentes, figura 21.

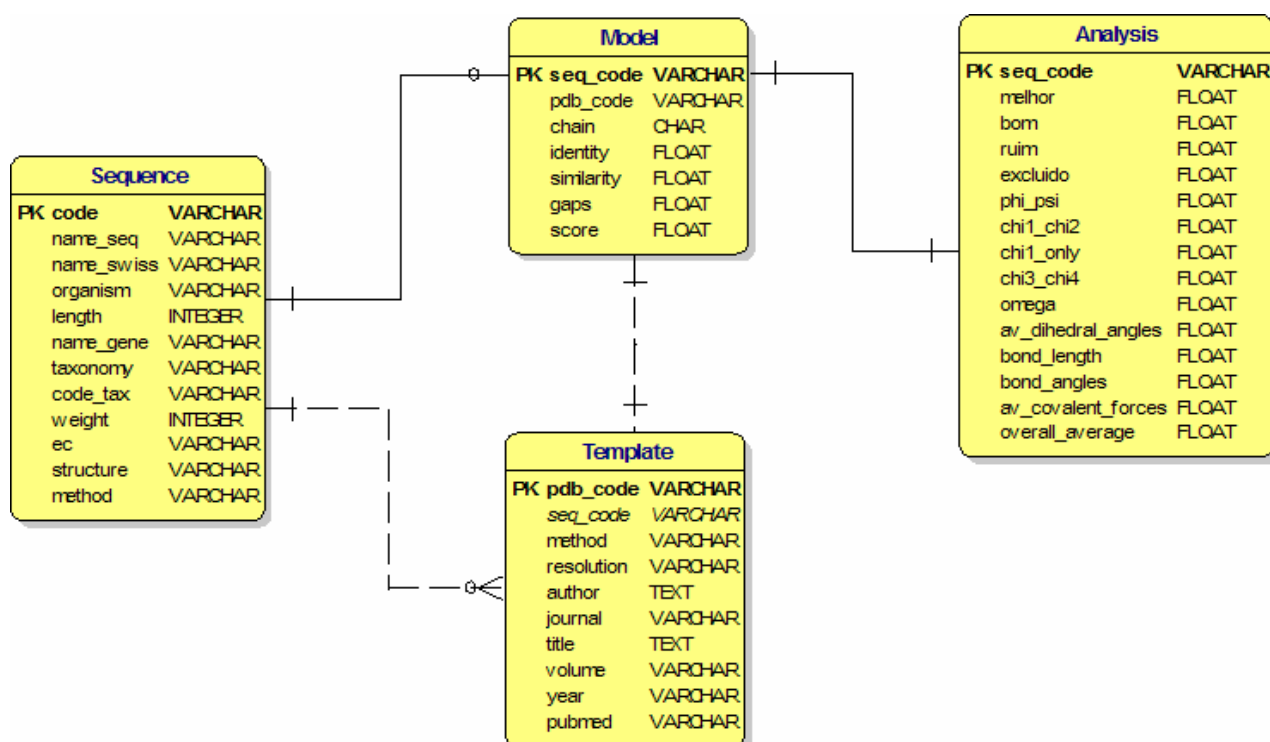


Figura 21. Diagrama de relacionamento existente entre as tabelas que compõem o banco de dados SKPDB.

As quatro tabelas mencionadas acima se relacionam entre si através do código de acesso do Swiss-Prot e contêm os seguintes campos e informações:

✓ **Sequence:** contém a abreviatura usada para cada seqüência depositada no Swiss-Prot (campo code), o nome de entrada no Swiss-Prot (campo name_sws), nome da proteína (campo name_seq), o nome do organismo (campo organism), o tamanho da seqüência em quantidade de aminoácidos (campo length), o nome do gene que expressa a proteína (campo name_gene), a sua taxonomia (campo taxonomy), o código da sua taxonomia (campo code_taxonomy), o peso molecular (campo weight), a classificação bioquímica (campo ec), para saber se a seqüência primária tem estrutura tridimensional (campo structure) e método utilizado na obtenção da estrutura (campo method).

✓ **Model:** contém o código da proteína (campo seq_code), código de acesso ao PDB da estrutura tridimensional que foi utilizada como molde (campo pdb_code), a cadeia (campo chain), a identidade entre as seqüências primárias alvo e molde (campo identity), a similaridade entre as seqüências primárias alvo e molde (campo similarity), porcentagem de *gaps* que o alinhamento teve (campo gaps) e score do alinhamento (campo score).

✓ **Template:** contém o código de acesso ao PDB da estrutura tridimensional (campo pdb_code), código de acesso da seqüência primária no Swiss-Prot (campo seq_code), o método de resolução da estrutura (campo method), a resolução em angstroms no caso de estruturas resolvida por cristalografia (campo resolution), sobre o artigo publicado deste molde (campos author, journal, title, volume, year) e o código do link para acesso ao resumo e o artigo no PUBMED (campo pubmed).

✓ **Analysis:** contém o código da proteína (campo seq_code), os valores das regiões do gráfico de ramachandran gerados para o modelo, região favorável (campo melhor), região permitida (campo bom), região adicionalmente permitida (campo ruim) e não permitida (campo excluído), o valor médio e total do G-factor para o modelo (campos phi_psi, chi1_chi2, chi1_only, chi3_chi4, omega, av_dihedral_angles, av_covalent_forces e overall_average) e o valor da geometria ideal para o modelo do ângulo de torção e comprimento de ligação (campos bond_length e bond_angles).

4.2.3. Acesso e interface do banco de dados

O SKPDB fornece uma interface com menus amigáveis, uma vez que todas as informações podem ser impressas. A pagina inicial (Home Page - Figura 22) oferece para o usuário várias formas de pesquisa no banco de dados, como descritas a seguir:

- ✓ **Pelo código de acesso Swiss-Prot+TrEMBL ou nome do organismo:** na opção “Search with your keyword in SKPDB o usuário deve digitar o código de acesso do Swiss-Prot ou o nome do organismo como palavras chave e selecionar o tipo de campo que vai ser pesquisado “Swiss-Prot Access code” ou “Organism”, depois deve selecionar a opção “Only Swiss-Prot access code or Organism” em “Options for your search”.
- ✓ **Pelo nome do organismo e da enzima:** na opção “Restrict your search in SKPDB with options and” o usuário pode refinar sua busca, buscando através do nome da enzima e do organismo caso não se lembre ou não saiba do código de acesso do Swiss-Prot. O usuário deve preencher os campos “organism” e “enzyme” na opção “Restrict your search in SKPDB with options and” e escolher a opção “and” em “Options for your search”.

A pagina inicial além da função de busca no banco de dados SKPDB também dá ao usuário uma visão geral sobre os objetivos e o que o usuário pode encontrar de informações sobre a via do ácido chiquímico no banco de dados, figura 22.

SKPDB - ShiKimate Pathway DataBase

Tuesday, 2009 February 10 14:35:25

Welcome to SKPDB!

SKPDB is a relational database of protein structures predicted by comparative modeling. We focused our study in enzymes of the shikimate pathway identified in all genome of microorganisms. The main goal of the present database is to provide structural models to be used in docking simulations and drug design. The SKPDB user interface provides user friendly menus, so that all information can be printed in one step from any standard web browser. All structure files are available for downloading. The database is organized to allow identification of the metabolic pathway and include other information about the modeled protein such as, sequence, animated gif for the structure, stereochemical quality and details about the modeling procedure.

Search with your keyword in SKPDB

Organism Swiss-Prot access code

Restrict your search in SKPDB with option AND

Organism Enzyme

Options for your Search

AND

Only Swiss-Prot access code or Organism

Supported by:

FAPESP UNESP UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO" Campus Rio Claro L S B Z Laboratório de Biologia Estrutural e Bioquímica Centro de Estudos de Interf. Sociais - CES Instituto de Biociências bioinfo.zip.net

Note: To see next pages, you need to install Java Plug-in, which happens automatically when you install the [Java \(TM\) SE JRE or JDK](#). You can find more information on the [Java Plug-in home page](#).

Site melhor visualizado com resolução 1024 x 768 pixels!
Copyright © 2008. All rights reserved

Figura 22. Página inicial do banco de dados SKPDB.

Já no menu à esquerda da página inicial encontram-se vários botões onde o usuário encontra as seguintes informações como:

- ✓ **Shikimate Pathway:** onde existem informações sobre a via metabólica do ácido chiquímico e sobre cada enzima da via, são várias informações como função, descrição da estrutura química, ligantes (coenzimas e substrato) pertencentes a cada enzima, *links* externos, entre outras informações (Figura 23);
- ✓ **Group:** informações sobre as pessoas que compõe o nosso grupo de pesquisa;
- ✓ **Links:** contem os *links* externos para as ferramentas e bancos de dados utilizadas neste trabalho;

✓ **Contact:** meu *email* para: sugestões, críticas e possíveis defeitos que o banco de dados possa vir a apresentar.

SKPDB Shikimate Pathway DataBase

Wednesday, 2008 November 26 17:24:36

SKPDB Home
Shikimate Pathway
Member Group
Links
Contact

Shikimate Pathway DAHPS DHQS DHQD SDH SK EPSPS CS References

Shikimate-5-Dehydrogenase (SDH)

The SDH functions to catalyze the reduction of 3-dehydroshikimate to shikimate using the cofactor NADH (Figure 1). There are two types of SDH from bacteria have been characterized to date, AroE (EC 1.1.1.25) and YdiB (EC 1.1.1.282). This enzyme SDH 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, and iron-activated alcohol dehydrogenases and long chain dehydrogenases [17].

The synonyms of SDH type AroE: dehydroshikimic reductase; shikimate oxidoreductase; shikimate:NADP⁺ oxidoreductase; 5-dehydroshikimate reductase; shikimate 5-dehydrogenase; 5-dehydroshikimic reductase; DHS reductase; shikimate:NADP⁺ 5-oxidoreductase.

3-Dehydroshikimate $\xrightarrow[\text{NADPH} + \text{H}^+]{\text{SDH}}$ D-Shikimate + NADP⁺

Figure 1. The reaction catalyzed by Shikimate-5-Dehydrogenase (in bacteria).

In plants SDH is associated with type I DHQD to form a bifunctional enzyme, this enzyme forms the C-terminal, whereas in fungi, such as *Neurospora crassa*, this enzyme forms the fifth domain of the pentafunctional AROM polypeptide, which catalyzes five of seven steps of the shikimate pathway [18,19].

Although AroE has been established as the enzyme responsible for flux through the main trunk of the shikimate pathway, YdiB has been implicated in a branch point involving the metabolism of quinate. YdiB was first characterized in fungi and was shown to play a significant role in the quinate utilization pathway rather than in the shikimate pathway itself. This pathway consists of three enzymes (quinate/shikimate dehydrogenase, 3-dehydroquinase, and dehydroshikimate dehydratase) that catabolize quinate into protocatechuic acid. In addition, quinate/shikimate dehydrogenase can also catalyze the reduction of 3-dehydroshikimate to shikimate [20].

Links to other databases SDH type AroE: [BRENDA](#), [EXPASY](#), [KEGG](#), [IUBMB](#), [MetaCyc](#), CAS registry number: 9026-87-3.

Links to other databases SDH type YdiB: [BRENDA](#), [EXPASY](#), [KEGG](#), [IUBMB](#), [MetaCyc](#).

Site melhor visualizado com resolução 1024 x 768 pixels!
Copyright © 2008. All rights reserved

Figura 23. Página sobre as enzimas e a via.

Na página inicial quando o usuário faz a consulta no banco de dados SKPDB pelo código de acesso do Swiss-Prot será exibido a este usuário uma com as seguintes informações sobre a sequência primária alvo, o modelo gerado e sobre molde usado na modelagem da enzima consultada quando a enzima tiver sido modelada (Figura 24). Se a enzima consultada já tiver sido resolvida por RMN ou cristalografia a página exibida terá as informações sobre a sequência primária e uma sobre os dados do experimento com link externo para o PDB (Figura 25).

SKPDB ShiKimate Pathway DataBase

Tuesday, 2009 February 10 14:53:51

SKPDB Home
Shikimate Pathway
Member Group
Links
Contact

Sequence Info | Model Info | Template Info | Model Analysis

Swiss-prot access code: Q3R2Y0

Swiss-prot ID: Q3R2Y0_XYLFA

Protein name: 3-dehydroquinate dehydratase class II (DHQD)

Organism: Xylella fastidiosa Ann-1

Gene name: ORFNames=XfasoDRAFT_0910

Taxonomy: Bacteria, Proteobacteria, Gammaproteobacteria, Xanthomonadales, Xanthomonadaceae, Xylella.

Taxonomy ID: NCBI_TaxID=155920

Sequence length: 155 a.a. residues Sequence weight: 16667 Daltons

Primary sequence in fasta format:

```
>Q3R2Y
MPLDLSMAHLLLLHGPNLNLLGTREPEIYGRITLPQIDAALAEAAAAGHGLSSLQNSA
EHLVIERIHATREDGTAFILINPQAFHTSVALRDALLAVALPFVEIHLNPHTRPFRH
HSYLADKALGVVCGFGVDSYRIALEGVIALRGSVDV
```

Click here to download of the sequence (Q3R2Y0.fasta)

Site melhor visualizado com resolução 1024 x 768 pixels!
Copyright © 2008. All rights reserved

Figura 24. Página com os resultados sobre busca efetuada com a enzima tendo sua estrutura 3D determinada por modelagem molecular comparativa.

SKPDB ShiKimate Pathway DataBase

Wednesday, 2008 November 26 17:53:52

SKPDB Home
Shikimate Pathway
Member Group
Links
Contact

Sequence Info | Model Info

An experimentally determined protein structure is available for this sequence:

Swiss-prot Accession: P0A4Z2

Protein Name: SHIKIMATE KINASE

Organism: Mycobacterium tuberculosis

PDB Accession: 1L4U resolved by method X-RAY DIFFRACTION with resolution 1.79 Å

Description Citation: Title: Crystal structure of shikimate kinase from Mycobacterium tuberculosis reveals the dynamic role of the LID domain in catalysis.
Author(s): Gu, Y., Reshetnikova, L., Li, Y., Wu, Y., Yan, H., Singh, S., Ji, X.
Journal: J.Mol.Biol.,v.319, 2002.

Swiss-prot Accession: P0A4Z2

Protein Name: SHIKIMATE KINASE

Organism: Mycobacterium tuberculosis

PDB Accession: 2IYZ resolved by method X-RAY DIFFRACTION with resolution 2.29 Å

Description Citation: Title: Mechanism of Phosphoryl Transfer Catalyzed by Shikimate Kinase from Mycobacterium Tuberculosis.
Author(s): Hartmann, M.D., Bourenkov, G.P., Oberschall, A., Strizhov, N., Bartunik, H.D.
Journal: J.Mol.Biol.,v.364, 2006.

Site melhor visualizado com resolução 1024 x 768 pixels!
Copyright © 2008. All rights reserved

Figura 25. Página com os resultados sobre busca efetuada com a enzima tendo sua estrutura 3D determinada experimentalmente.

ma última hipótese também foi levada em consideração e se a enzima não tiver estrutura resolvida experimentalmente e um molde apropriado não foi encontrado, aí a página apresentada ao usuário irá conter as informações sobre a seqüência primária e um aviso dizendo que a estrutura 3D desta enzima ainda não está disponível, figura 26.

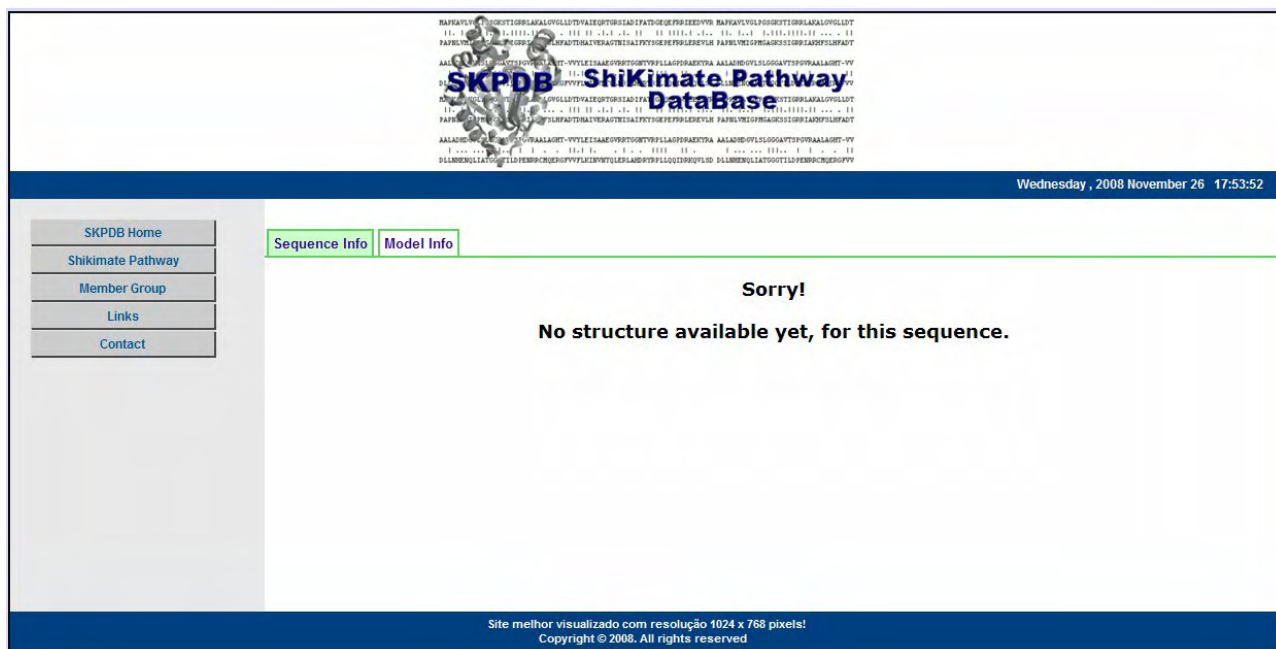
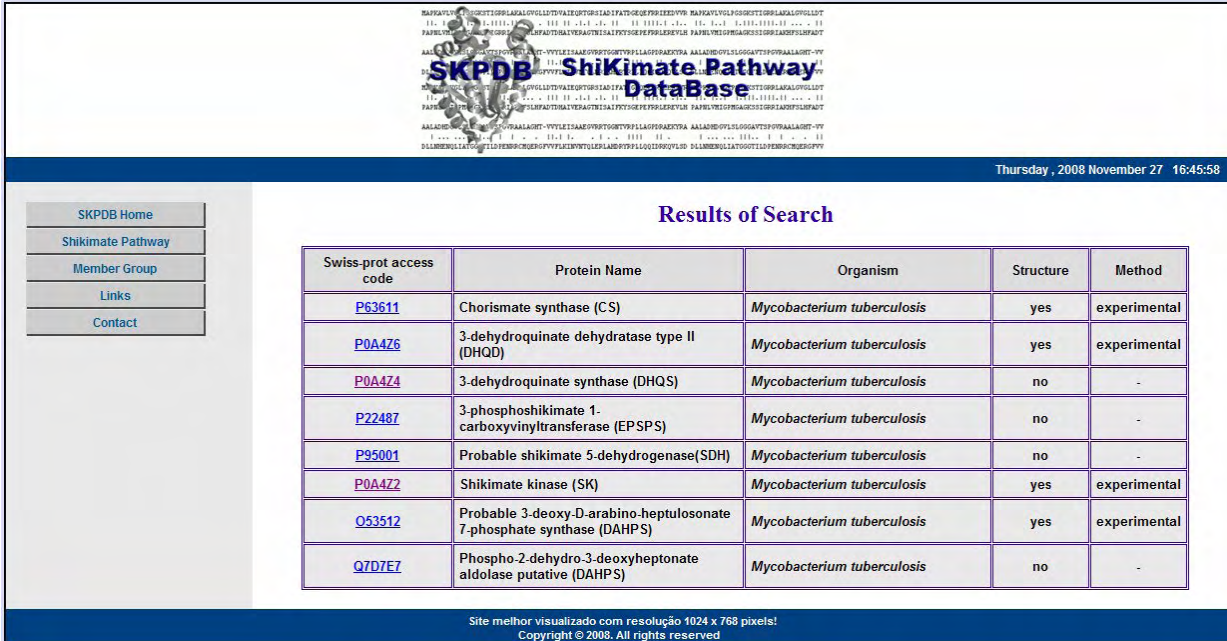


Figura 26. Página com os resultados sobre busca efetuada com a enzima não tendo sua estrutura 3D determinada experimentalmente ou por modelagem molecular comparativa.

Se o usuário na página inicial (Figura 22) faz a consulta no SKPDB pelo nome do organismo será exibida uma lista de todas as enzimas da via metabólica do ácido chiquímico que estão depositadas no SKPDB, para que o usuário possa escolher qual enzima ele precisa dos dados (Figura 27), depois da escolha feita às páginas apresentadas ao usuário vão variar de acordo com as informações disponíveis na base de dados do SKPDB, ou seja, a página exibida com os resultados da busca feita será uma das mostradas acima, figuras 24, 25 ou 26.



SKPDB Shikimate Pathway DataBase

Thursday, 2008 November 27 16:45:58

Results of Search

Swiss-prot access code	Protein Name	Organism	Structure	Method
P63611	Chorismate synthase (CS)	<i>Mycobacterium tuberculosis</i>	yes	experimental
P0A4Z6	3-dehydroquinate dehydratase type II (DHQD)	<i>Mycobacterium tuberculosis</i>	yes	experimental
P0A4Z4	3-dehydroquinase (DHQS)	<i>Mycobacterium tuberculosis</i>	no	-
P22487	3-phosphoshikimate 1-carboxyvinyltransferase (EPSPS)	<i>Mycobacterium tuberculosis</i>	no	-
P95001	Probable shikimate 5-dehydrogenase (SDH)	<i>Mycobacterium tuberculosis</i>	no	-
P0A4Z2	Shikimate kinase (SK)	<i>Mycobacterium tuberculosis</i>	yes	experimental
O53512	Probable 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAHPS)	<i>Mycobacterium tuberculosis</i>	yes	experimental
Q7D7E7	Phospho-2-dehydro-3-deoxyheptonate aldolase putative (DAHPS)	<i>Mycobacterium tuberculosis</i>	no	-

Site melhor visualizado com resolução 1024 x 768 pixels.
Copyright © 2008. All rights reserved

Figura 27. A interface acima representa a página com a lista de resultados sobre busca efetuada através do nome do organismo na base de dados.

As informações anotadas na base de dados são divididas em 3 partes e são apresentadas ao usuário em uma só página separada por abas, que são:

- ✓ **Aba sequence info:** são disponibilizadas informações gerais sobre a enzima como peso molecular, organismo, taxonomia, tamanho da enzima (a.a.), *download* da seqüência primária, e ainda existe um *link* para o Swiss-Prot/TrEMBL para aumentar a acessibilidade a outros tipos de informações sobre a enzima consultada, figura 24.
- ✓ **Aba model info:** são disponibilizadas informações gerais sobre a modelagem molecular da enzima como as informações sobre o *molde* utilizado (código de acesso no PDB, a cadeia escolhida e um link externo para o PDB), sobre o alinhamento (identidade, similaridade, gaps e score) entre as seqüências primárias (alvo/molde), e sobre o modelo (estruturas secundárias e superfície eletrostática) que pode ser visualizado através do programa Jmol, figura 28.

SKPDB ShiKimate Pathway DataBase

Tuesday, 2009 February 10 14:58:49

SKPDB Home
Shikimate Pathway
Member Group
Links
Contact

Sequence Info	Model Info	Template Info	Model Analysis
Swiss-prot Accession	Q3R2Y0 (Q3R2Y0_XYLFA)		
Protein Name	3-dehydroquinase class II (DHQD)		
Organism	<i>Xylella fastidiosa</i> Ann-1		
PDB Accession	1UQR Chain id A		
Alignment	Identity: 39.6% between target and template		
	Similarity: 66% between target and template		
	Gaps: 61.5, Score: 424		

Jmol View

Structure Color by:

Mode View Structure:

Surface:

Background:

Animation spin

This image may be rotated and zoomed.

[Click here to download of the structure modelling \(Q3R2Y0.pdb\)](#)

Instructions for using Jmol to display molecules:

1. Choose which molecules to display by selecting them using the popup menus or clicking on the appropriate link.
2. ROTATE the image by holding down the mouse button while moving the cursor over the image.
3. ZOOM by holding down the shift key while moving the cursor up (decrease magnification) or down (increase magnification) on top of the image.
4. Other options are available in the control menu accessible by holding the mouse button down while the cursor is over "Jmol" in the lower right corner (right click also works on a multibutton mouse).
5. For more info about Jmol go to www.jmol.org.

Site melhor visualizado com resolução 1024 x 768 pixels!
Copyright © 2008. All rights reserved

Figura 28. Página da aba “Model Info”, que mostra os dados sobre o modelo gerado.

✓ **Aba template info:** são disponibilizadas informações gerais sobre o *molde* como nome da proteína, organismo, faixa do alinhamento, método experimental, resolução (se tiver), dado sobre a publicação se tiver (autor, jornal, título, volume, página, ano e link de acesso externo ao PUBMED), figura 29.

SKPDB ShiKimate Pathway DataBase

Tuesday, 2009 February 10 15:04:38

SKPDB Home	Sequence Info	Model Info	Template Info	Model Analysis
Shikimate Pathway	Swiss-prot Accession	P43877		
Member Group	Protein Name	3-DEHYDROQUINATE DEHYDRATASE		
Links	Organism	Actinobacillus pleuropneumoniae		
Contact	PDB Accession	1UQR resolved by method X-RAY DIFFRACTION with resolution 1.70 Å		
	Description Citation	<p>Title: Structural Study of the Type II 3-Dehydroquinate Dehydratase from Actinobacillus Pleuropneumoniae</p> <p>Author(s): Maes, D., Gonzalez-Ramirez, L.A., Lopez-Jaramillo, J., Yu, B., De Bondt, H., Zegers, I., Afonina, E., Garcia-Ruiz, J.M., Gulnik, S.</p> <p>Journal: Acta Crystallogr., Sect. D, v.60, 2004.</p>		

Site melhor visualizado com resolução 1024 x 768 pixels!
Copyright © 2008. All rights reserved

Figura 29. Página da aba “Template Info” que mostra os dados sobre o molde utilizado na construção da estrutura 3D do modelo.

✓ **Aba analysis model:** são disponibilizadas as informações sobre as análises feitas no modelo, utilizando as ferramentas mencionadas no tópico 3 e também o *download* destes arquivos, figura 30.

SKPDB ShiKimate Pathway DataBase

Thursday, 2008 November 27 17:28:14

SKPDB Home	Sequence Info	Model Info	Template Info	Model Analysis
Shikimate Pathway	<p>Ramachandran Statistics Generate Ramachandran Plot</p>			
Member Group				
Links	<p>Residues in most favoured regions: 96.2%</p> <p>Residues in additional allowed regions: 3%</p> <p>Residues in generously allowed regions: 0%</p> <p>Residues in disallowed regions: 0.8%</p>			
Contact	<p>Dihedral angles score</p> <p>Phi-psi distribution: 0.22</p> <p>Chi1-chi2 distribution: -0.18</p> <p>Chi1 only: 0.01</p> <p>Chi3 and chi4: 0.48</p> <p>Omega: -0.22</p> <p>Average score: 0.01</p>		<p>Main-chain covalent forces score</p> <p>Main-chain bond lengths: -0.09</p> <p>Main-chain bond angles: -0.27</p> <p>Average score: -0.19</p>	
	<p>G - Factor Click here to download</p>			
	<p>RMSD from Ideal Geometry Click here to download</p>			
	<p>Overall Average score: -0.07</p> <p>Bond Lengths: 0.019 Å</p> <p>Dihedral Angles: 23.415 °</p> <p>Bond Angles: 2.259 °</p> <p>Improper Angles: 1.868 °</p>			

Site melhor visualizado com resolução 1024 x 768 pixels!
Copyright © 2008. All rights reserved

Figura 30. Página da aba “Model Analysis”, que mostra os resultados sobre a qualidade do modelo.

Caso a enzima consultada tenha sua estrutura 3D resolvida por métodos experimentais somente será mostrada ao usuário somente a informação sobre a seqüência e sobre a estrutura 3D. E se a enzima não tiver estrutura será mostrado ao usuário somente as informações sobre a seqüência.

4.3. Cristalização, difração de raios X e resolução da estrutura

4.3.1. Cristalização da enzima chiquimato quinase

Para a cristalização da mtCQ na forma apo e dos complexos foram feitas todas as condições dos artigos mencionados no tópico 3, mas infelizmente não obtivemos sucesso. Somente para uma condição a descrita por Krell *et al.*, 2001 obtivemos micro-cristais. Então foi feita a variação da concentração das soluções de cristalização e pH do Tris HCl na faixa de tamponamento de 7,0 à 8,6 mas os mini-cristais não cresceram; somente mudaram de forma como podemos ver na figura 31. A condição inicial utilizada na cristalização foi 23% de PEG 3350 e 0,1M de Tris HCl pH 8,0.

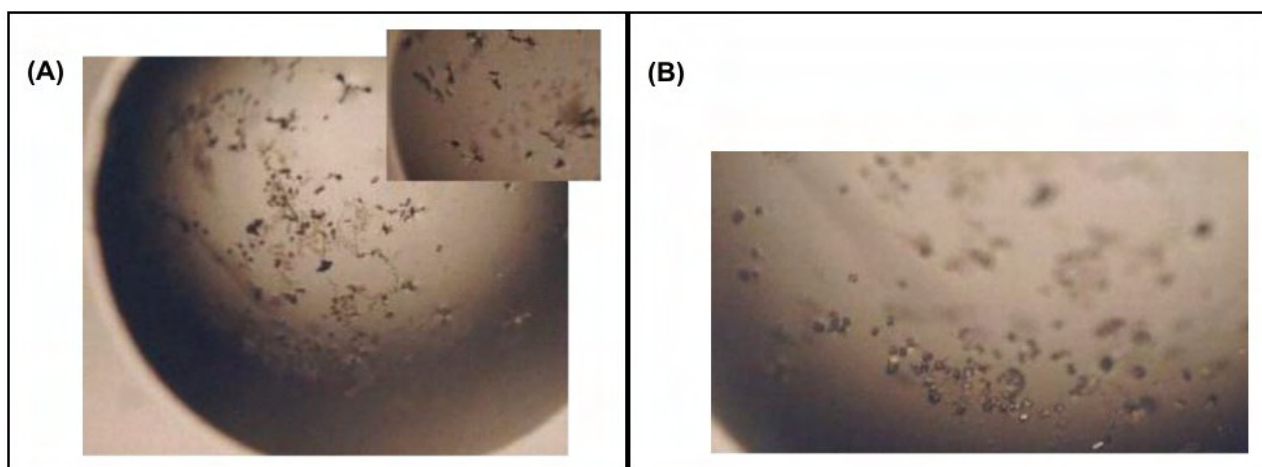


Figura 31. Cristais da enzima mtCQ. (A) Forma dos mini-cristais do pH 7,0 à 7,9 e (B) Forma dos mini-cristais do pH 8,0 à 8,6.

Após muitas tentativas de cristalização, acredita-se que a obtenção de cristais da mtCQ em complexo com os ligantes citados acima seja difícil, pois tentamos todas as condições descritas na literatura (CHENG *et al.*, 2005; GAN *et al.*, 2006; GU *et al.*, (2002); KRELL *et al.*, 2001; PEREIRA *et al.*, 2004) inclusive fazendo variações destas condições mas sem sucesso. Para tentar

solucionar esta dificuldade tentamos cristalizar a chiquimato quinase na forma apo para podermos utilizar a técnica de *soaking* com os ligantes, mas também não obtivemos sucesso. Então buscamos na literatura algo que justificasse essa dificuldade encontrada, já que não havia nenhum estudo destes ligantes com a chiquimato quinase.

Uma das justificativas encontradas é que os ligantes *roscovitine* e *olomoucine* apesar de serem análogos de ATP apresentam um perfil hidrofóbico ao contrário do que ocorre com o ATP que tem um perfil hidrofílico. Por serem hidrofóbicos não diluem no tampão como o ATP, somente são solúveis em clorofórmio e dimetilsulfoxido (DMSO) e para este trabalho utilizamos o solvente orgânico DMSO por ser menos hidrofóbico que o clorofórmio.

Inclusive simulações de *docking* molecular foram feitas para verificar como estes ligantes possivelmente se ligam a enzima mtCQ e estes resultados foram comparados com a estrutura cristalográfica depositada no PDB código de acesso 2IYZ. E os resultados mostram que estes ligantes *olomoucine* e *roscovitine* se ligam de forma irregular ao sítio do ATP, estes resultados são discutidos com mais detalhes no tópico 4.8.

Uma outra justificativa é que o DMSO diminui ou até mesmo pode inibir a afinidade da chiquimato quinase pelos ligantes. Estudos feitos com este solvente orgânico utilizando proteínas dependentes de ATP (THE & HASSELBACH, 1977; OZAL & VAN DER VEGT, 2006; AL-SHAWI & SÊNIOR, 1992; WESTH, 2004; BEHARRY & BRAGG, 2001) mostram que o DMSO pode causar perturbações nas interações hidrofóbicas entre a proteína e o ligante por mudar o ambiente do meio e passa a funcionar como um agente perturbador (caotrópico) por ser de baixa dieletricidade e por aumentar a entropia deste meio.

Para verificar a extensão destes efeitos do DMSO encontrados na literatura na chiquimato quinase, foi feito experimentos de CD tanto de quantização de estrutura quanto de estabilidade térmica da enzima na forma apo e na presença dos ligantes *olomoucine*, *roscovitine* e ADP, e foi verificado que o DMSO pode estar interferindo na afinidade entre a enzima e os ligantes, os

resultados são mostrados com mais detalhes nos tópicos 4.4 e 4.5.

4.3.2. Cristalização da enzima corismato sintase

Para a cristalização da mtCS na forma apo e na presença dos ligantes foram feitas todas as condições dos artigos mencionados no tópico 3 e foi obtido cristais em várias condições de cristalização diferentes, mas nem sempre estes cristais eram bons para coleta, por serem pequenos demais, por não apresentarem arestas e formas bem definidas para coleta de dados e por difratarem a uma resolução superior a 3,0 Å.

Então neste trabalho será apresentado somente as melhores condições em que os cristais estavam ideais para coleta de dados e que apresentaram os melhores conjuntos de dados durante a resolução da estrutura, as condições otimizadas estão descritas abaixo:

- ✓ Variação de uma das condições do kit cristalização (5): 3M formato de sódio e 0,1M de bis-tris-propano, foi utilizado os cristais somente na forma apo da proteína mtCS (Figura 32A).
- ✓ Variação da condição (4) de cristalização 0,7 de citrato de sódio e 0,1M de Hepes pH 7,5 descritas por QUEVILLON-CHERUEL *et al.*, 2004: a condição otimizada de cristalização foi de 1M de citrato de sódio e 0,1M de Hepes pH 7,5, foi utilizado cristais da proteína mtCS na forma apo e na presença do ligante CA (produto da reação catalisada pela enzima) (Figura 32B e 32C).

Os cristais utilizados para a coleta de dados tinham um tamanho entre 0,7 a 1 mm de comprimento, como podemos ver nas figuras abaixo apresentavam arestas e forma bem definidas.

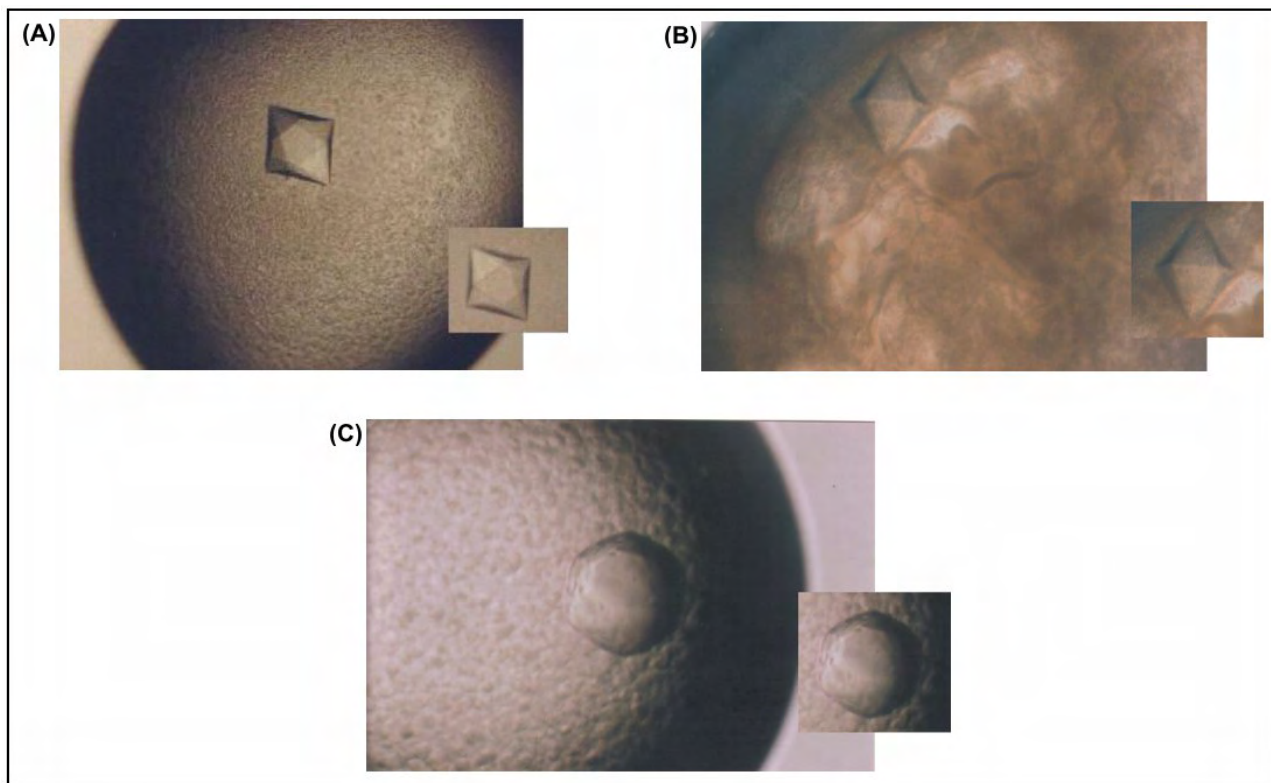


Figura 32. Cristais da enzima mtCS. (A) Na forma apo na condição (5), (B) Na forma apo na condição (4) e (C) Na presença do produto AC na condição (4).

4.3.3. Coleta de dados e difração de raios X da enzima corismato sintase

Os dados processados pelo programa MOSFILM e escalonados pelo programa SCALA determinaram uma resolução de 2,00 Å para o cristal mtCS-AC conseguido na condição (5). Quando indexados, mostraram pertencer ao grupo de Laue primitivo ortorrômbico P6422, com parâmetros de cela $a=129,08\text{Å}$, $b=129,08\text{Å}$, $c=155,83\text{Å}$ e os ângulos $\alpha=90^\circ$, $\beta=90^\circ$, $\gamma=120^\circ$. Os dados apresentaram ainda completude de 99,84%, $R_{\text{sym}}=5,1\%$ que está abaixo do limite de 20, indicando que é um conjunto de dados que pode ser utilizado para a resolução do modelo tridimensional.

Para o cristal mtCS-FMN-AC conseguido na condição (4) a resolução determinada foi de 2,13 Å. Quando indexados, mostraram pertencer ao grupo de Laue primitivo ortorrômbico P6422, com parâmetros de cela $a=130,6\text{Å}$, $b=130,06\text{Å}$, $c=160,9\text{Å}$ e os ângulos $\alpha=90^\circ$, $\beta=90^\circ$, $\gamma=120^\circ$. Os dados apresentaram ainda completude de 99,86% e $R_{\text{sym}}=7,1\%$ que está abaixo do limite de 20,

indicando que é um conjunto de dados que pode ser utilizado para a resolução do modelo tridimensional. As imagens dos dados processados são mostradas na figura 33.

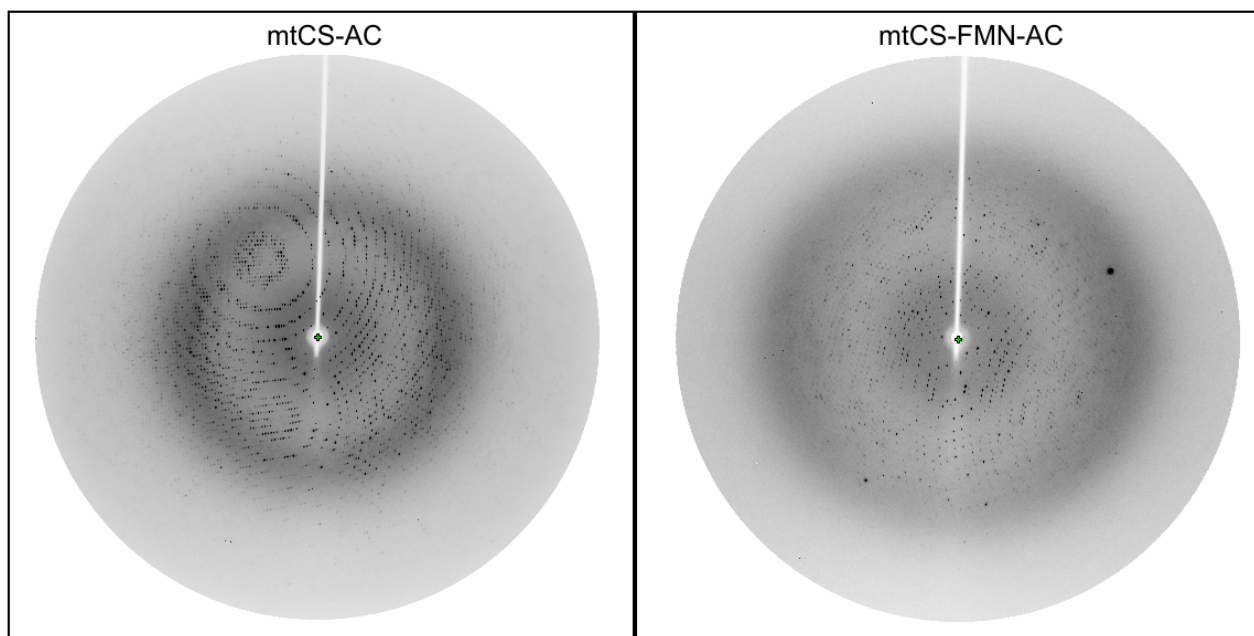


Figura 33. Imagens de difração por raios X do cristal mtCS-AC na condição (5) e mtCS-FMN-AC na condição (4).

4.3.4. Resolução e refinamento do modelo da estrutura 3D da enzima corismato sintase

Para substituição molecular foi utilizada a estrutura cristalográfica depositada no PDB com o código de acesso: 2QHF como modelo para a resolução do problema da fase, da qual foi retirada às moléculas de água cristalográficas para o procedimento de substituição molecular. E foram utilizadas as estruturas cristalográficas depositadas no PDB com o código de acesso: 2QHF, 2O12 e 1QXO para retirar as coordenadas cristalográficas dos ligantes.

A tabela 6, a seguir apresenta os resultados obtidos com o programa AMoRe. São apresentados os ângulos de Euler (α , β e γ), que fornecem a rotação necessária para orientar a estrutura corretamente na unidade assimétrica do cristal. Também são mostradas as coordenadas fracionárias, que indicam a correto posicionamento dentro da unidade assimétrica, que é realizada por tentativa e erro.

Na tabela 6 estão, ainda, apresentados o coeficiente de correlação (CC), assim como o Fator

R (R_{fator}) para a estrutura. Os valores ideais do Fator R e CC são de abaixo de 40% e acima de 50%, respectivamente, nessa fase.

Tabela 6. Resultados obtidos na substituição molecular.

	mtCS-FMN-AC	mtCS-AC
Ângulos de Euler		
α (°)	56,99	57,68
β (°)	69,34	69,78
Γ (°)	87,80	88,39
Coordenadas Fracionárias		
T_x	0,4181	0,4178
T_y	-0,1113	-0,1118
T_z	0,3883	0,3878
Coefficiente de correlação (%)	68,2	81,93
* R_{fator} (%)	39,6	28,5

* $R_{\text{fator}} = 100 \times \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum (F_{\text{obs}})$, somas obtidas sobre todas as reflexões com corte de $F/\sigma(F) > 2$.

Na etapa de refinamento, o objetivo é encontrar uma melhor concordância entre o modelo proposto para a estrutura da molécula e sua estrutura real. O acompanhamento da qualidade desse processo de refinamento é feito através do cálculo do R_{fator} . Partindo de uma estrutura com R_{fator} por volta de 50%, um bom refinamento diminui esse valor para 20% ou menos. Na tabela 7 estão as estatísticas de refinamento da mtCS.

Tabela 7. Estatísticas de refinamento para mtCS.

	mtCS-FMN-AC	mtCS-AC
Faixa de resolução (Å)	37,27-2,13	37,27- 2,00
* R_{fator} (%)	23,8	21,6
** R_{free} (%)	26,7	23,6
R.M.S.D. a partir da geometria ideal		
Distância de ligações (Å)	0,016	0,018
Ângulos de ligações (°)	2,4	2,0

* $R_{\text{fator}} = 100 \times \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum (F_{\text{obs}})$, somas obtidas sobre todas as reflexões com corte de $F/\sigma(F) > 2$

** R_{free} (%) = R_{fator} para 10% dos dados, que não foram incluídos no refinamento

As estruturas refinadas, oriundas da substituição molecular, convergiram para bons valores de R_{fator} e R_{free} . A análise comparativa dos valores do R_{fator} e do R_{free} , entre essa estrutura de mtCS e as demais já determinadas e publicadas, mostra boa qualidade e seus parâmetros de refinamento estão dentro dos valores aceitáveis.

Após a obtenção das estruturas 3D da enzima corismato sintase para cada um dos complexos (mtCS-AC e mtCS-FMN-AC) pode-se verificar a ausência de densidade eletrônica para os referidos ligantes, figura 34. Como os experimentos de cristalização não foram muito satisfatórios, resolveu-se fazer experimentos com *docking* molecular, pois dessa forma poder-se-ia prever a localização do sítio ativo para a FMN e o AC.

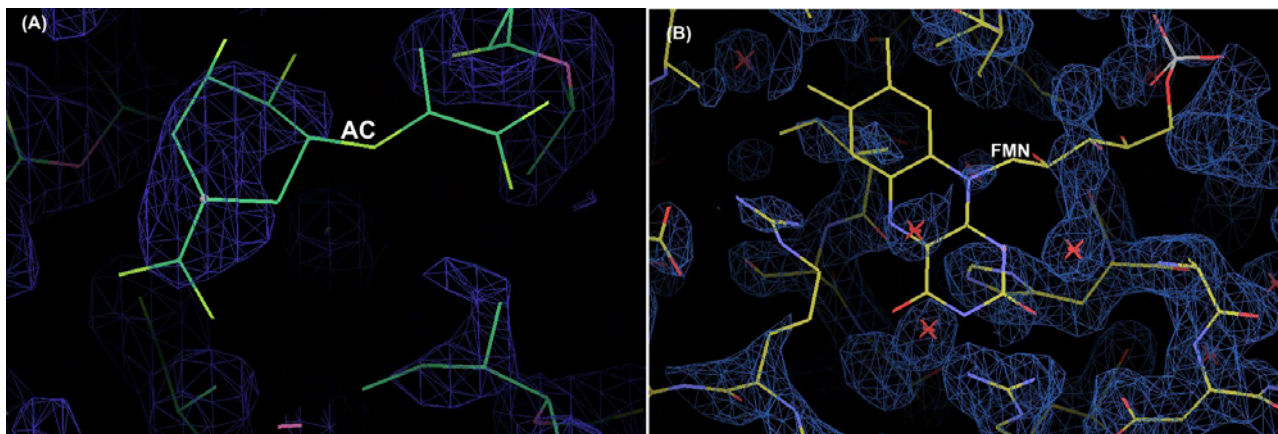


Figura 34. Mapa de densidade eletrônica na região de ligação dos ligantes, mostrando a ausência parcial ou total da densidade eletrônica para os referidos ligantes.

4.3.5. Análise da estrutura 3D

O modelo final da estrutura da enzima corismato sintase pode ser observado na figura 35 gerada pelo programa Pymol (DeLano, 2002). A estrutura mtCS possui topologia semelhante à do molde. Sua topologia é do tipo β - α - β , uma arquitetura típica de CS, apresentando 9 α -hélices e 15 fitas- β .

O melhor modelo estrutural para a mtCS foi obtido após 10 ciclos de refinamento, seguido de análise da estrutura 3D associada ao mapa de densidade eletrônica (figura 36). O mapa de densidade mostra que a resolução foi suficiente para identificação precisa dos átomos.

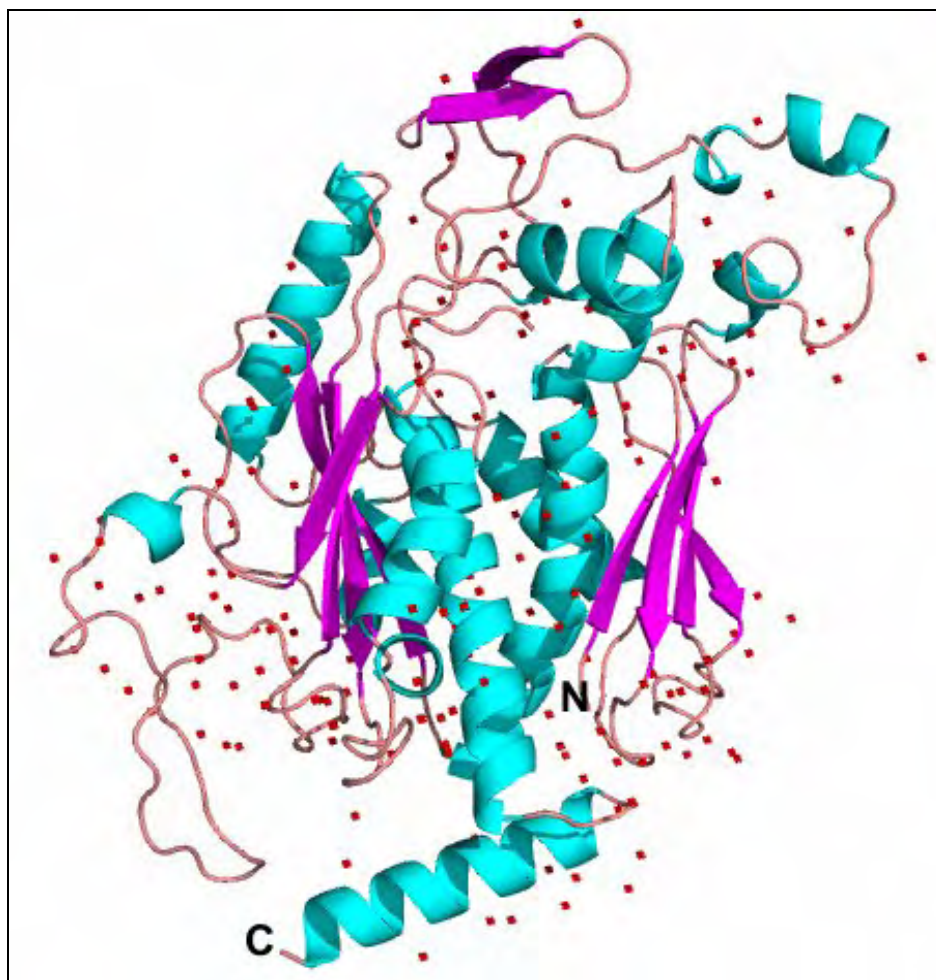


Figura 35. Estrutura da mtCS com as moléculas de água (em vermelho), gerada pelo programa Pymol. Colorida de acordo com a estrutura secundária: α -hélices (verde) e fitas- β (roxa) voltas e alças (rosa).

Todas as águas foram observadas e analisadas quanto à estabilização da estrutura 3D por ligações de hidrogênio e as distâncias pertinentes entre essas moléculas e as cadeias laterais dos aminoácidos. Este modelo refinado apresentou um Fator R de 16,8% (abaixo do esperado de 20% para estruturas com alta qualidade e resolução) e um R_{free} de 23,2%. As coordenadas atômicas foram salvas em um arquivo formato pdb.

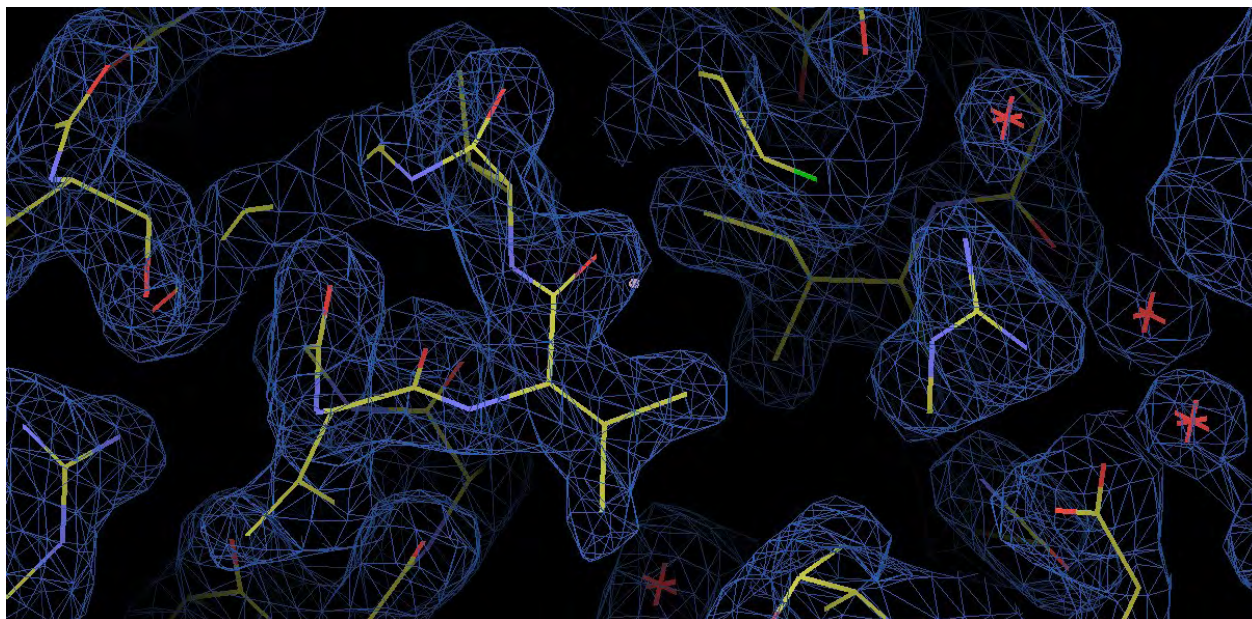


Figura 36. Mapa de densidade eletrônica para uma das estruturas de mtCS.

A qualidade estereoquímica da estrutura mtCS foi analisada por meio do Gráfico de Ramachandran que apresenta os ângulos de ligação phi (Φ) e psi (ψ) da cadeia principal, tais ângulos são referentes às ligações C^{α} -N e C^{α} -C. O gráfico foi gerado pelo programa PROCHECK (LASKOWSKI et al., 1993), figura 37.

Análises do diagrama de Ramachandran Φ - Ψ para as estruturas indicaram que os resíduos se encontram > de 93,5% nas regiões mais favoráveis, < 5,6% nas regiões adicionalmente permitidas, < de 0,9% em regiões generosamente permitidas e nenhum resíduo em regiões não permitidas do diagrama.

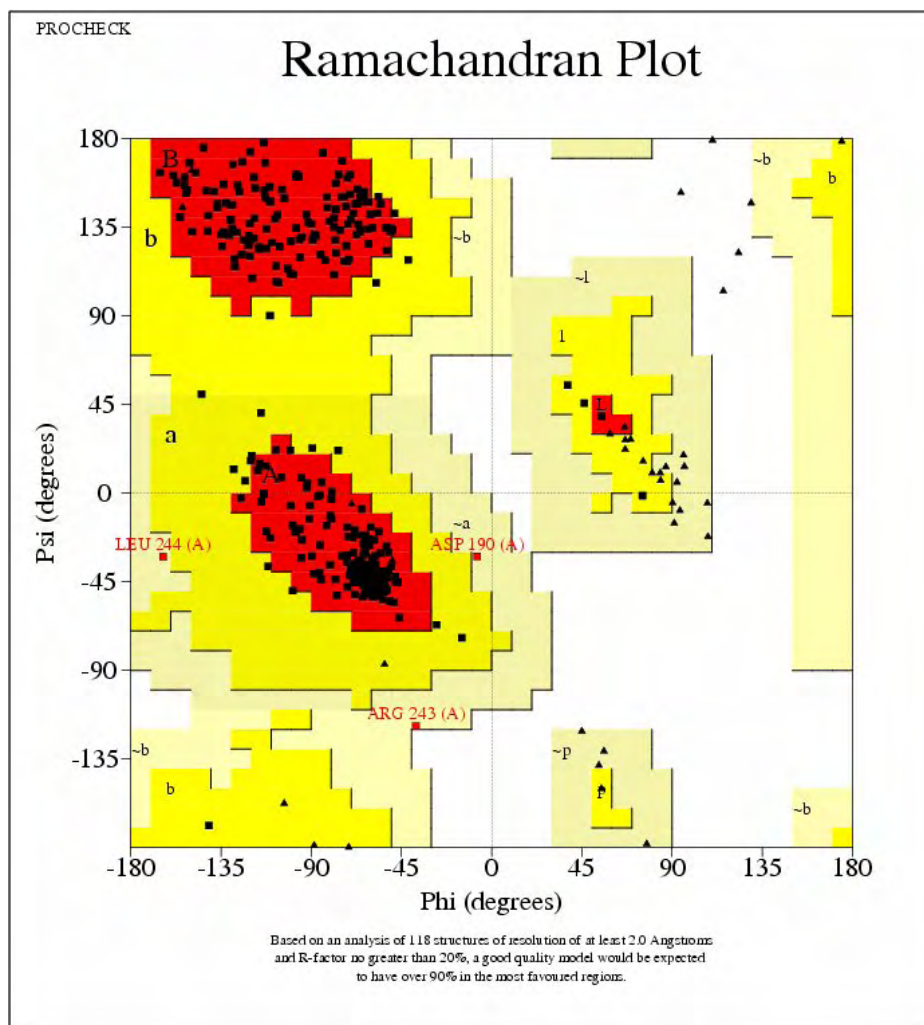


Figura 37. Gráfico de Ramachandran com a posição de cada resíduo de aminoácido para um dos modelos finais da estrutura 3D da mtCS.

4.4. Experimentos de CD para estimar as frações de estrutura secundária das enzimas

A análise dos espectros de CD das enzimas que determina a fração da estrutura secundária foi obtida pelo programa CONTILL (SREERAMA & WOODY, 2000), este programa segue oito referências, das quais sete são geralmente usadas na análise do CD e uma é construída com base na estrutura terciária. As oito referências do programa CONTINLL são: SP29, SP22X, SP37, SP43, SP37A, SDP42, SDP48, CLSTR.

O programa CONTINLL representa no seu banco de dados uma quantidade determinada de proteínas com comprimento de onda num intervalo estabelecido e ainda as estruturas secundárias: αR , αD , βR , βD , T, U, P2. Uma exceção é a referência CLSTR, a qual apresenta uma quantidade

variada de proteínas e que determina a fração da estrutura terciária.

4.4.1. Chiquimato quinase

Foi utilizado como base para o experimento de CD os seguintes artigos publicados: KRELL *et al.*, 2001, CERASOLI *et al.*, 2003 e KRELL *et al.*, 1998. A enzima CQ pertence à família das nucleosídeo monofosfato (NMP) quinases que sofrem grandes mudanças conformacionais durante a catálise (KRELL *et al.*, 1998). Segundo estudos de Krell e seus colaboradores (1998) que fizeram análises de CQ de *Escherichia coli* em complexo γ N-ATP e ácido chiquímico (SA) foi observado que com a ligação do γ N-ATP e SA houve uma mudança conformacional que corresponde aproximadamente a um decréscimo de 10% de hélice alfa e um aumento de 10% de folha β com um erro experimental de $\pm 5\%$, o que seria normal por causa das mudanças conformacionais que ocorrem durante catálise.

Os espectros de CD da mtCQ na presença e ausência do ADP e ácido chiquímico (SA) são mostrados na figura 38. As porcentagens obtidas dos elementos de estruturas secundárias podem ser observadas na tabela 8. A porcentagem das estruturas secundárias obtidas através dos dados de CD da mtCQ demonstram uma boa concordância com os resultados obtidos através das estruturas cristalográficas depositadas no PDB levando em consideração que proteína se encontra em um estado de empacotamento e que os experimentos de CD foram feitos em solução onde as moléculas se encontram em um estado mais livre.

Tabela 8. Porcentagem dos elementos de estrutura secundária calculado para os dados experimentais de CD da mtCQ.

	mtCQ na forma apo	mtCQ-ADP- MgCl ²⁺	mtCQ- <i>roscovitine</i> - MgCl ²⁺	mtCQ- <i>olomoucine</i> - MgCl ²⁺
Folhas β	26%	28%	39%	37%
Hélices α	70%	68%	48%	56%
Estr. aleatória	4%	4%	13%	7%

Os dados de CD para a proteína mtCQ em complexo com os ligantes: *olomoucine* e

roscovitine mostraram uma diminuição bastante significativa na porcentagem de hélices α e um aumento nas estruturas aleatórias e folhas β , já os dados de CD para a proteína mtCQ na presença do ADP mostram uma diminuição discreta na porcentagem de hélices α e um aumento também discreto na porcentagem de folhas β .

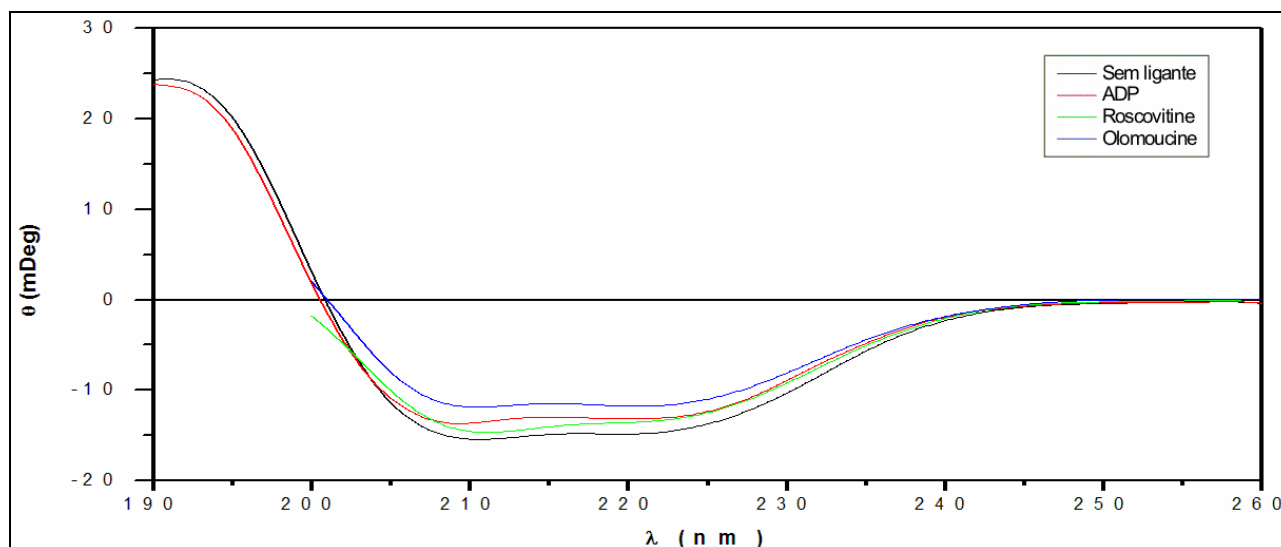


Figura 38. Sobreposição dos espectros de dicroísmo circular da mtCQ sem e na presença de ligantes

Baseado na figura 38 e na tabela 8 pode-se afirmar que o complexo mtCQ-ADP-MgCl²⁺ apresenta resultados similares aos obtidos por Krell e seus colaboradores (1998) e com as estruturas depositadas no PDB, mas nos outros dois complexos os resultados encontrados ficam bem distantes destes valores, o que nos leva a concluir que o DMSO pode estar interferindo no meio causando uma diminuição da afinidade da proteína pelos ligantes *olomoucine* e *roscovitine* ou interferindo diretamente na estabilidade da mtCQ.

Um outro motivo para este resultado seria que como os ligantes têm um perfil hidrofóbico poderiam estar se ligando de forma errada na molécula, ou com pouca afinidade o que resultaria numa instabilidade da mtCQ, esta verificação foi feita através de *docking* molecular.

4.4.2. Corismato sintase

Os espectros de CD da mtCS na presença e ausência de FMN e CA são mostrados na figura 39. As porcentagens obtidas dos elementos de estruturas secundárias podem ser observadas na tabela 9. A porcentagem das estruturas secundárias obtidas através dos dados de CD da mtCS apresentam uma grande concordância com os resultados obtidos através das estruturas cristalográficas depositadas no PDB levando em consideração que proteína se encontra em um estado de empacotamento e que os experimentos de CD foram feitos em solução onde as moléculas se encontram em um estado mais livre.

Tabela 9. Porcentagem dos elementos de estrutura secundária calculado para os dados experimentais de CD da mtCS.

	mtCS na forma apo	mtCS-FMN	mtCS-AC	mtCS-AC-FMN
Folhas β	18	18	14	14
Hélices α	40	36	46	48
Estr. aleatória	42	46	38	36

A presença de FMN oxidada causa pequenos efeitos no espectro do CD como mostra o artigo Macheroux e seus colaboradores (1998) que fizeram experimentos de CD com a CS de *E.coli* (MACHEROUX *et al.*, 1998). Mas quando adicionamos o produto da reação AC foi observado outra mudança no espectro de CD, o que indica que a mtCS se compacta para fazer a catalise, somente quando o substrato (EPSP) se liga a esta enzima, como pode ser observado na figura 38.

As moléculas de AC e de FMN fornecem interessantes alvos para o desenvolvimento de drogas baseado em estrutura, não somente contra tuberculose, mas também para outros patógenos. Inibidores podem ser competitivos com o produto e cofatores, levando a inativação da enzima.

A molécula de 5-deazaflavina (LAUHON & BARTLETT, 1994), um análogo de FMN onde o nitrogênio da posição 5 é substituído por um carbono, e (6R)-6-fluoro-EPSP, onde o hidrogênio-6R do EPSP é substituído por um flúor, resulta na ausência de atividade da corismato sintase (MACHEROUX, 1998; BORNEMANN *et al.*, 2003). Estes são exemplos de moléculas que foram

racionalmente modificadas e possuem a capacidade de inibição da enzima.

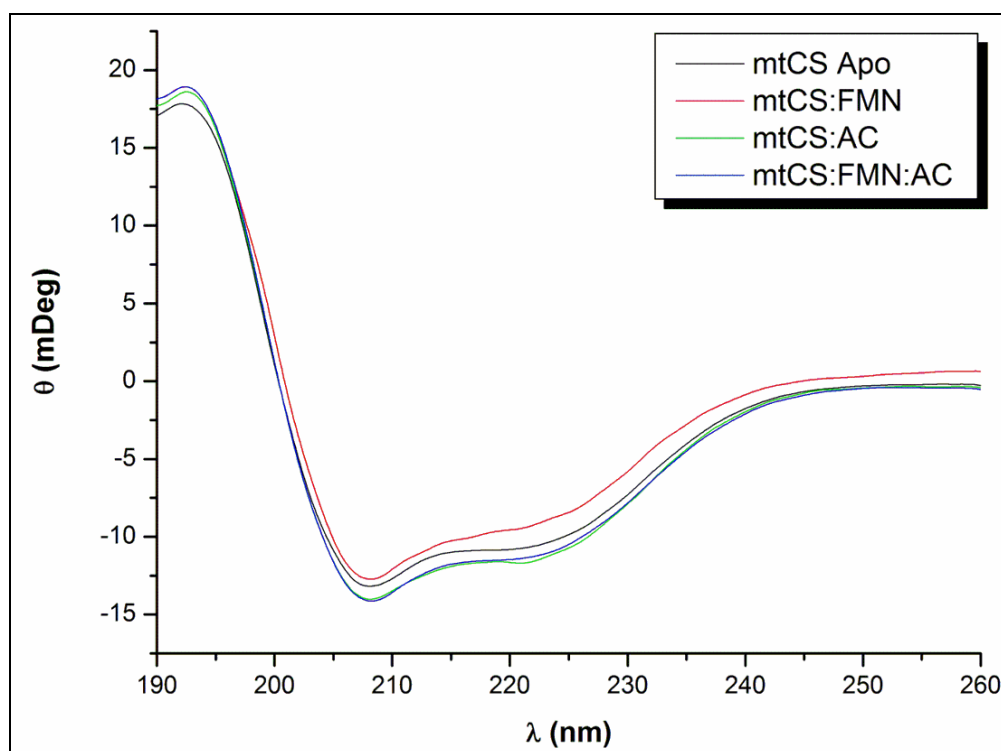


Figura 39. Sobreposição dos espectros de dicroísmo circular da mtCS sem e na presença de ligantes.

4.5. Experimentos de CD para verificar a estabilidade das enzimas

O estudo de estabilidade térmica das enzimas na forma apo e na presença dos ligantes foi usada para averiguar a natureza das transições de desenovelamento e renovelamento térmico e o quanto os ligantes interferiam nestas transições. A desnaturação térmica das enzimas na forma apo e na presença dos ligantes mostrou-se um processo irreversível, uma vez que as amostras aquecidas até 70 °C não voltaram a apresentar espectro de CD semelhante ao seu estado nativo, mesmo após rápido ou lento resfriamento para 25 °C.

Para o mtCQ foi utilizado tanto a desnaturação térmica com controle do tempo manual quanto a por controle do aparelho, com a finalidade de comparar os valores obtidos e desta forma verificar se estes valores eram confiáveis, já que se tratava de ligantes que ainda não foram estudados com esta enzima. Neste trabalho será mostrado somente o valor obtido pela desnaturação

térmica por controle manual de tempo, pois os valores obtidos pelo controle de tempo pelo aparelho ficaram muito próximos aos valores obtidos pelo controle manual do tempo, sendo assim mostrar os valores para os dois experimentos ficaria repetitivo e não traria informação adicional nenhuma para este trabalho.

4.5.1. Chiquimato quinase

As figuras 40 a 42, mostram a mtCQ na forma apo e na presença dos ligantes *olomoucine* e *roscovitine* e na presença do cofactor ADP em diferentes temperaturas num pH 7,4. O espectro de CD da mtCQ na forma apo, característico de proteína contendo elementos em hélice alfa, foi preservado no intervalo de temperatura entre 283,15 e 318,15 K, figura 40.

No espectro de CD da mtCQ Na presença do ligante *olomoucine* o intervalo de temperatura característico de proteína contendo elementos em hélice alfa, foi entre 283,15 e 308,15 K, figura 41. E no espectro de CD da mtCQ na presença do ligante *roscovitine*, característico de proteína contendo elementos em hélice alfa, foi preservado no intervalo de temperatura entre 283,15 e 303,15 K, figura 42.

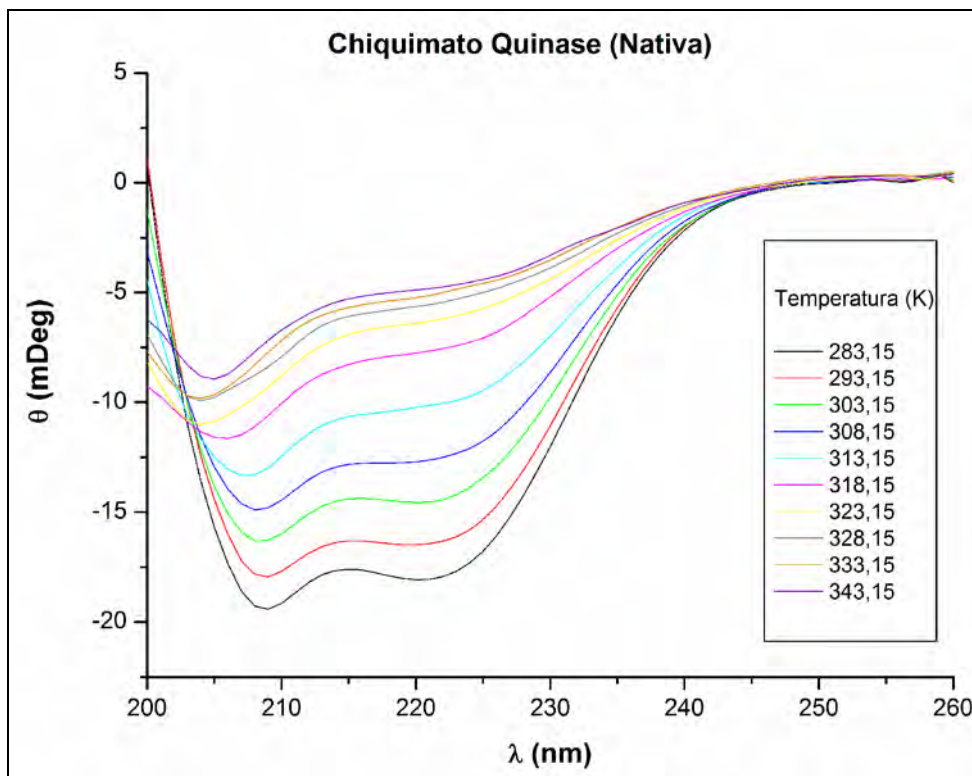


Figura 40. Espectros de CD da mtCQ na forma apo em diferentes temperaturas

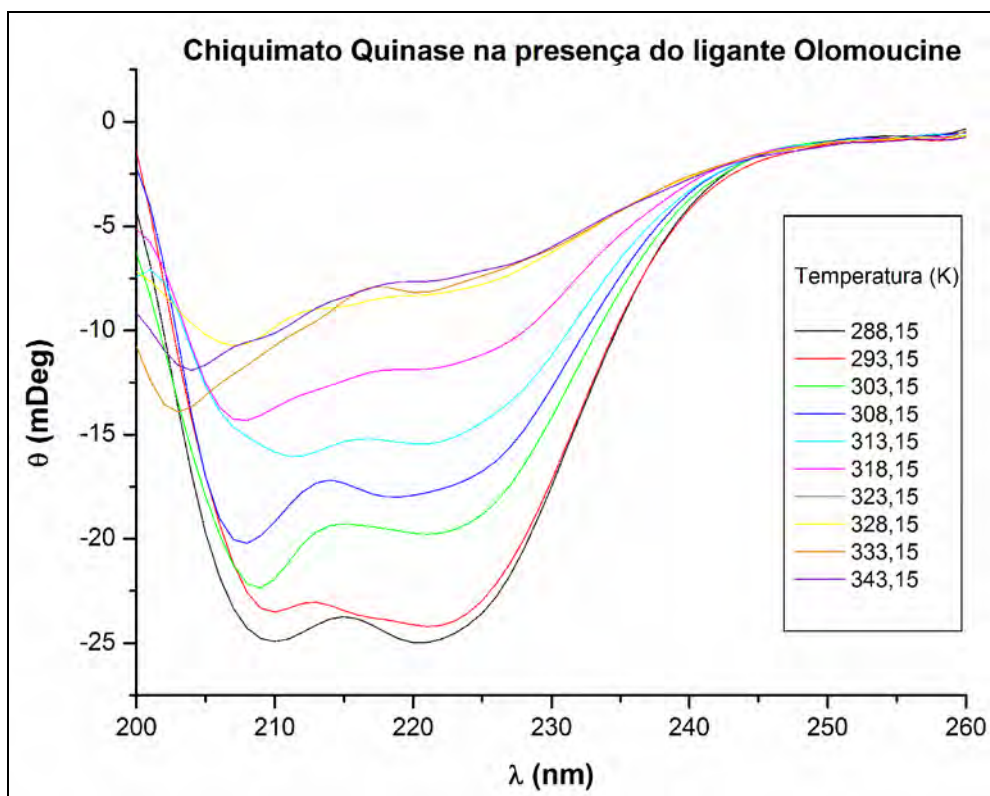


Figura 41. Espectros de CD da mtCQ em complexo com *olomoucine* em diferentes temperaturas

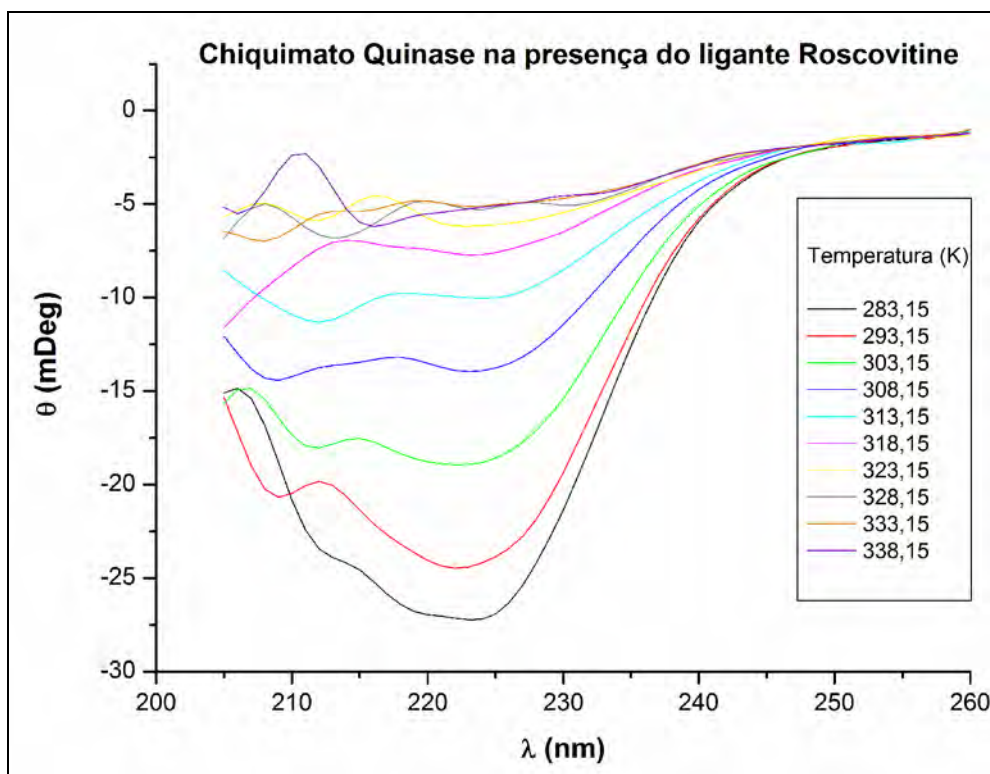


Figura 42. Espectros de CD da mtCQ em complexo com *olomoucine* em diferentes temperaturas

A curva sigmoidal da mtCQ na forma apo e em complexo com os ligantes indica que este processo envolve aparentemente uma transição de dois estados. A temperatura de transição do estado enovelado para o desenovelado da proteína, T_m , foi de aproximadamente 310K na forma apo (Figura 43), de aproximadamente 309K (Figura 44) e de aproximadamente 305K, Figura 45.

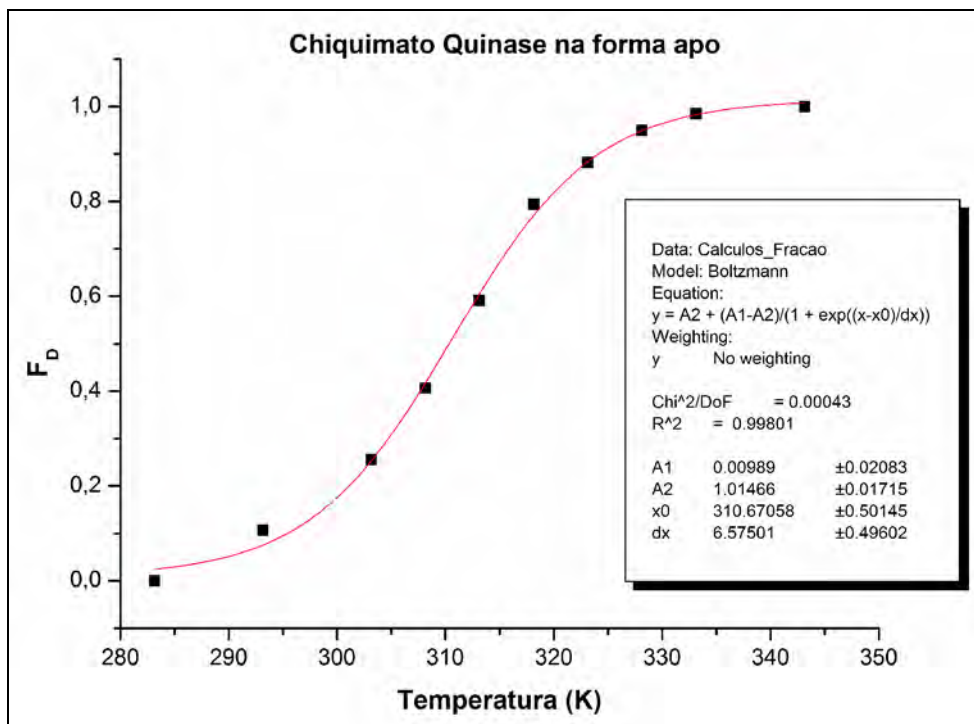


Figura 43. Transição da desnaturação térmica da mtCQ na forma apo

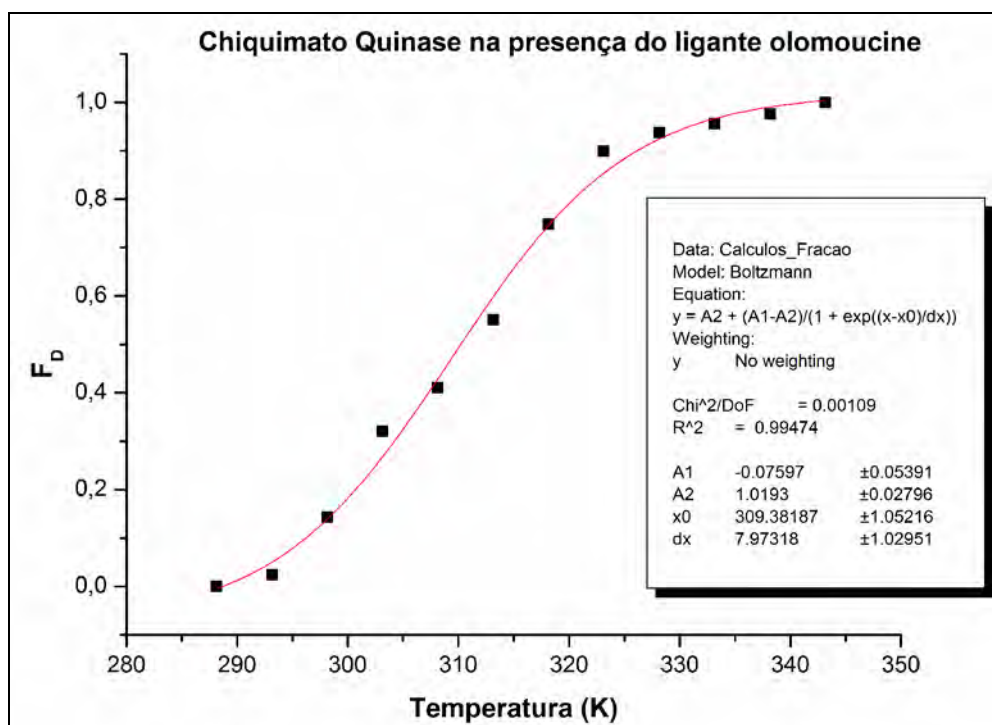


Figura 44. Transição da desnaturação térmica da mtCQ na presença do ligante *olomoucine*

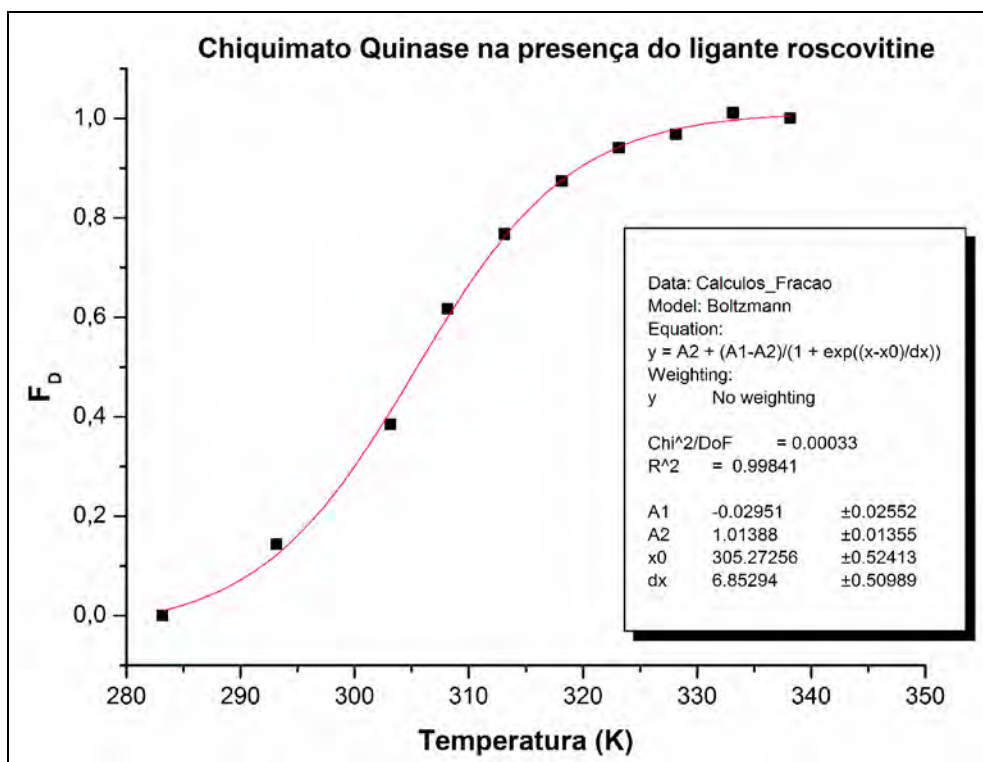


Figura 45. Transição da desnaturação térmica da mtCQ na presença do ligante *roscovitine*

Foi feita também a desnaturação do cofator ADP por controle de tempo, somente para verificar o T_m e se havia mudança no tipo de transição do estado de desnaturação, mas a curva sigmoidal indicou que o processo envolve aparentemente uma transição de dois estados também. A temperatura de transição do estado enovelado para o desenovelado da proteína, T_m , foi de aproximadamente 311K, Figura 46.

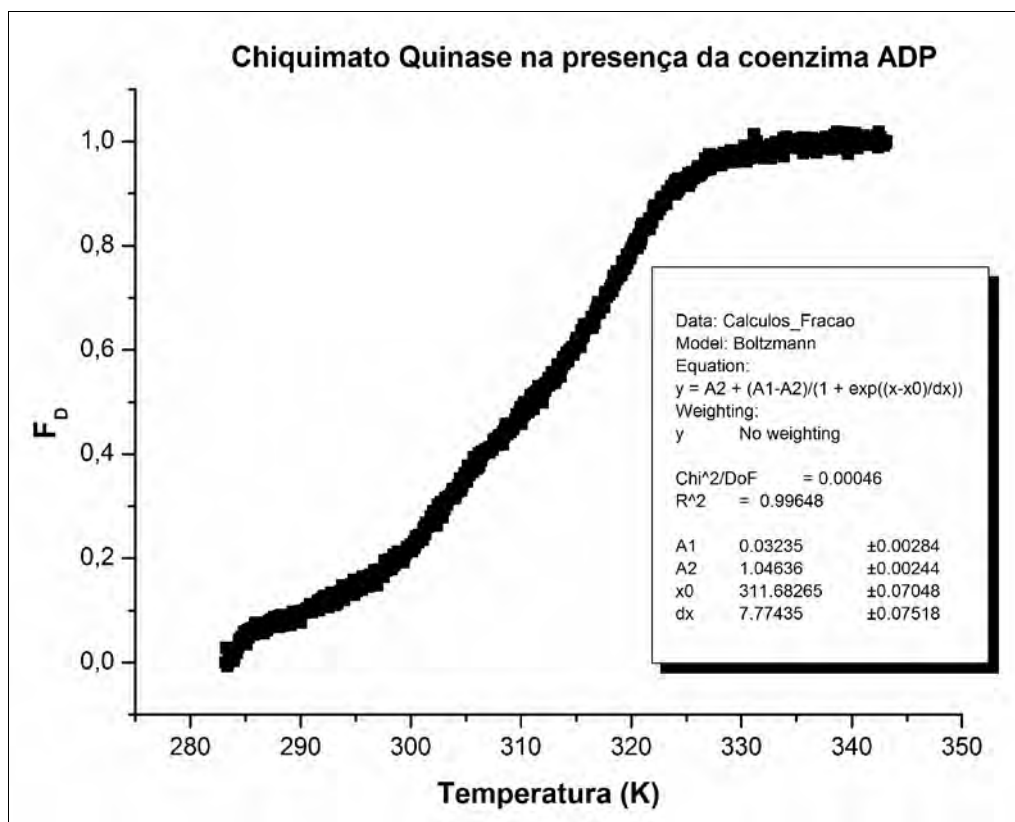


Figura 46. Transição da desnaturação térmica da mtCQ na presença da coenzima ADP

Foi observado que com os ligantes o processo de desenovelamento acontece a uma faixa de temperatura bem menor do que a mtCQ na forma apo e com ADP, isso se dá ao fato das perturbações causadas pelo DMSO e pelos próprios ligantes por terem um perfil hidrofóbico. Os resultados mostram que a enzima mtCQ é relativamente sensível a alteração do meio e que este fator poderia estar interferindo na obtenção dos cristais, que os ligantes não alteram o modo do estado de transição da enzima mtCQ e que a coenzima ADP estaria dando uma maior estabilidade a enzima pois houve um sutil aumento na T_m , como apresentado no trabalho de Krell (1998 e 2001) e Cerasoli (2003).

4.5.2. Corismato sintase

A curva sigmoidal indica que este processo envolve aparentemente uma transição de dois estados. A temperatura de transição do estado enovelado para o desenovelado da proteína, T_m , foi de aproximadamente 314K (Figura 47) no estado nativo e 314K (figura 48) na presença de ligantes.

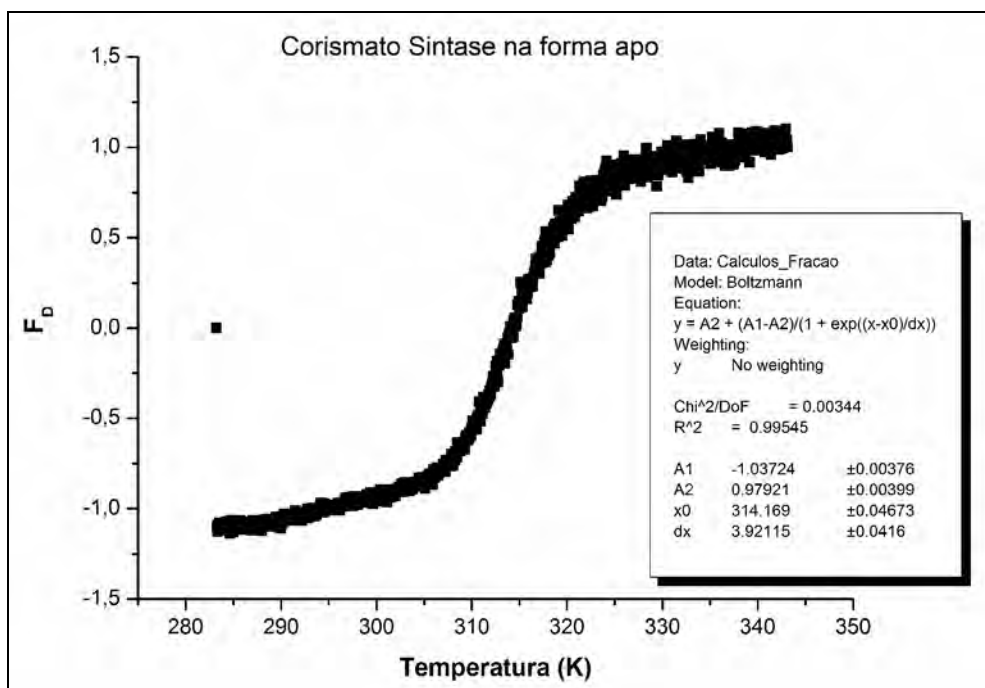


Figura 47. Transição da desnaturação térmica de mtCS na forma apo

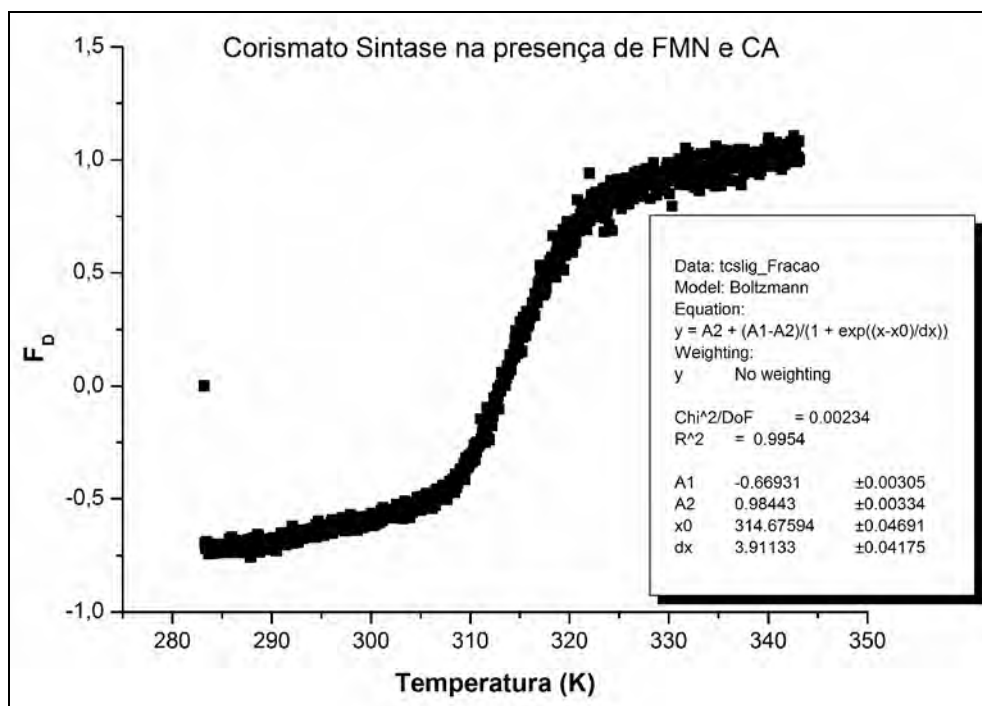


Figura 48. Transição de desnaturação térmica da mtCS na presença dos ligantes FMN e AC

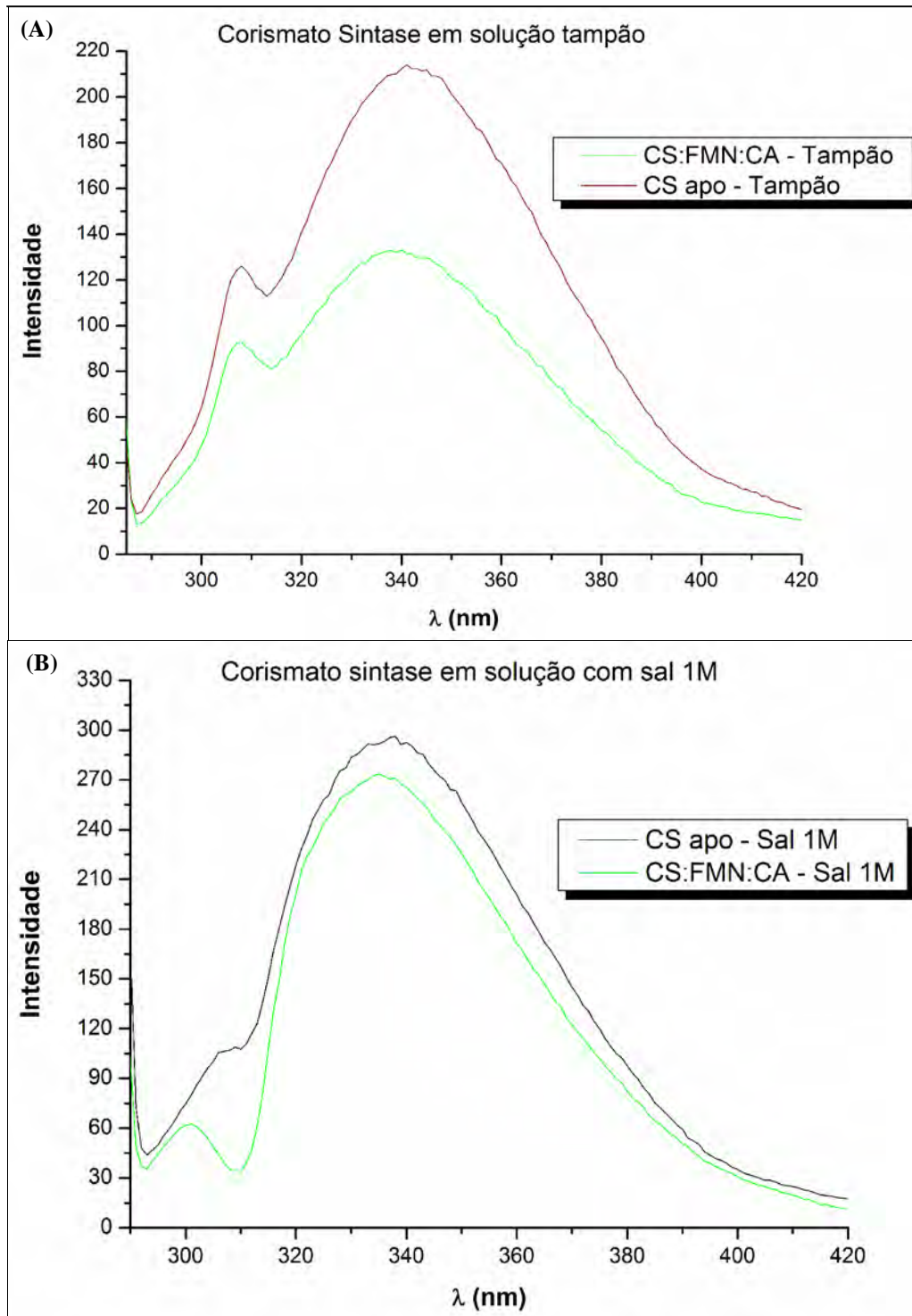
Os resultados mostram que a proteína mtCS é estável a alteração do meio com ou sem ligantes em função do aumento da temperatura, baseado nos resultados possivelmente os ligantes desta enzima não tem a função de estabilizar a enzima para a catalise, como é o caso da chiquimato quinase.

4.6. Experimento de Fluorescência da corismato sintase

O ambiente dos resíduos de Trp e Try presentes na seqüência primária da mtCS foram monitorados pela espectroscopia de emissão de fluorescência a fim de complementar e confirmar a estabilidade da mtCS em diferentes concentrações de sais.

Os espectros de emissão de fluorescências da mtCS na forma apo e na presença dos ligantes (FMN e AC) em diferentes concentrações de sais e em solução tampão é mostrado nas figuras 49A – D. É possível observar que mudanças na intensidade de fluorescência aconteceram com mudanças na concentração do sal no meio, isto também foi observado na obtenção dos cristais, quanto menor

a quantidade de sal na solução, menor eram os cristais e a formação destes cristais também eram alteradas.



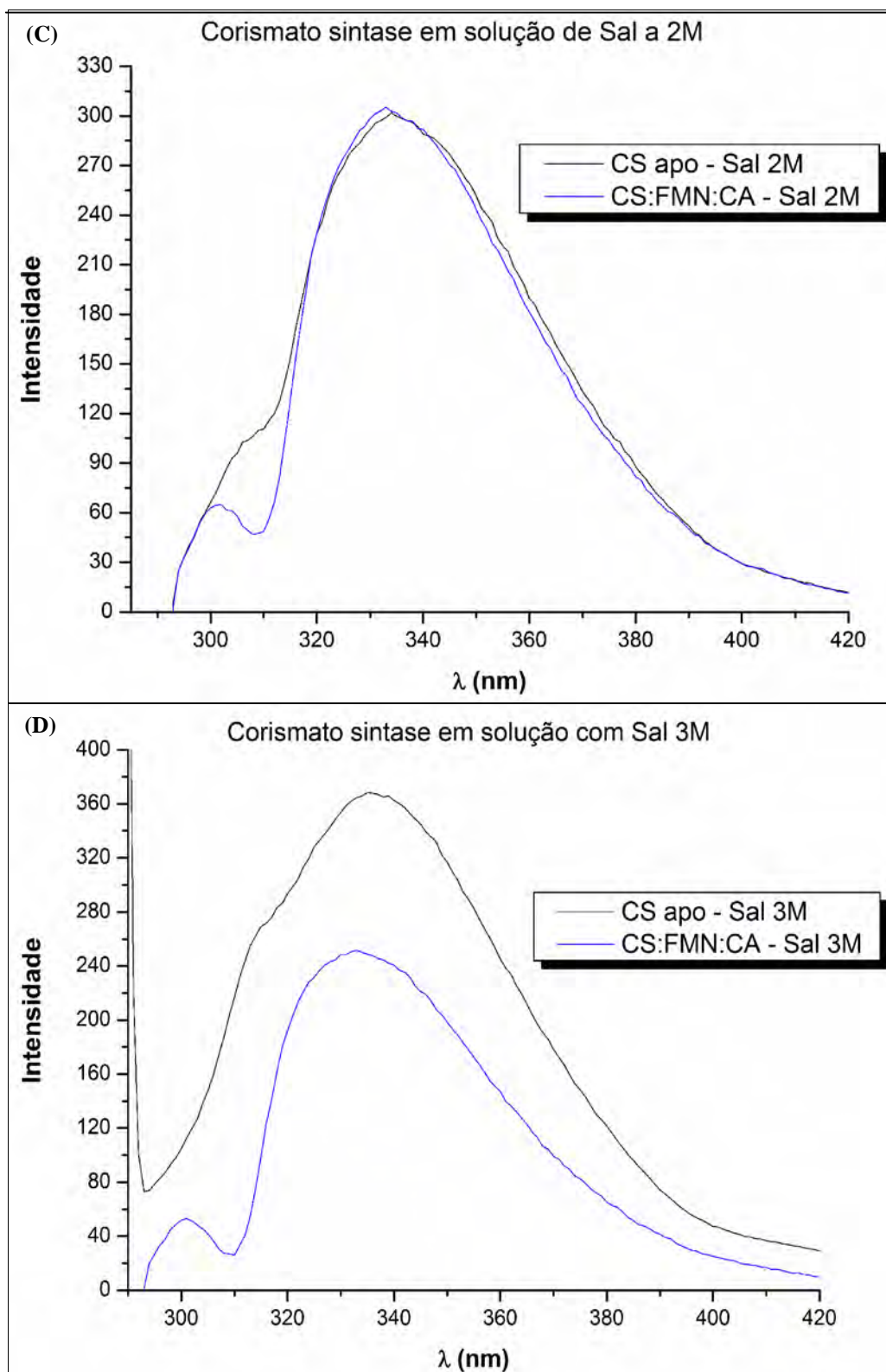


Figura 49. Espectros de fluorescência da mtCS sem e na presença dos ligantes FMN e CA

Uma supressão de fluorescência intrínseca pôde ser atribuída onde a redução do rendimento de fluorescência foi resultante da distribuição de cargas e da mudança da carga líquida da proteína

mtCS. Baseado nas figuras acima pode-se afirmar que a quantidade de sal interfere tanto na afinidade dos ligantes como mencionado acima quanto no empacotamento da enzima mtCS, pois ocorre mudança na intensidade e deslocamento do pico.

4.7. Experimento de SAXS da corismato sintase

Analisando as curvas de espalhamento obtidas por SAXS calculamos os seguintes parâmetros geométricos da proteína como máxima dimensão ($D_{\text{máx}} = 100,00 \text{ \AA}$), raio de giração ($R_g = 41,67 \text{ \AA}$), superfície da molécula ($S = 42140 \text{ \AA}^2$) e o volume hidratado ($V_h = 210 \times 10^3 \text{ \AA}^3$); estes parâmetros indicam que esta proteína se apresenta em solução sobre a forma de um tetrâmero, e esses resultados estão em bom acordo com os parâmetros calculados computacionalmente para o modelo cristalográfico tetramérico onde ($R_g = 40,97 \text{ \AA}$) volume ($V = 199184 \text{ \AA}^3$) e superfície ($S = 39702 \text{ \AA}^2$) foi avaliado utilizando o programa CRY SOL (SVERGUN *et. al.*1995) e a máxima dimensão ($D_{\text{máx}} = 92,94 \text{ \AA}$).

Foi gerado computacionalmente o modelo do tetrâmero a partir das coordenadas cristalográficas do monômero e assim sobreposto ao modelo obtido por SAXS e observou-se uma sobreposição muito boa que é uma indicação de que realmente a proteína mtCS encontra-se sob a forma de um tetrâmero em solução.

Os dados de SAXS foram devidamente tratados e a correção das curvas de espalhamento por efeitos geométricas do feixe, assim como o cálculo da função de distribuição de distância $p(r)$ foi obtido através da aplicação da função de transformada indireta de Fourier utilizando do programa GNOM (SVERGUN 1992).

O raio de giração, R_g ($R_g = 2.628\sqrt{-\alpha}$), foi obtido através da aproximação da lei de Guinier $I(q) = I(0).e^{\frac{-R_g^2 q^2}{3}}$ (Guinier and Fournet, 1955) para valores de q tendendo a zero mediante a inclinação α da região linear do gráfico de $\log I$ versus q^2 .

O volume e área da superfície externa da proteína foram calculados pelas equações $V = 2\pi^2 I(0)/Q$ e $S = \pi V \lim(I(q)q^4)/Q$ respectivamente, onde Q representa o invariante $Q = \int_0^\infty I(q)q^2 dq$, a intensidade da curva extrapolada $I(0)$ foi obtida da lei de Guinier.

Na figura 50 é apresentado o gráfico da região de Guinier, que mostra claramente que o sistema estudado é caracterizado como monodisperso.

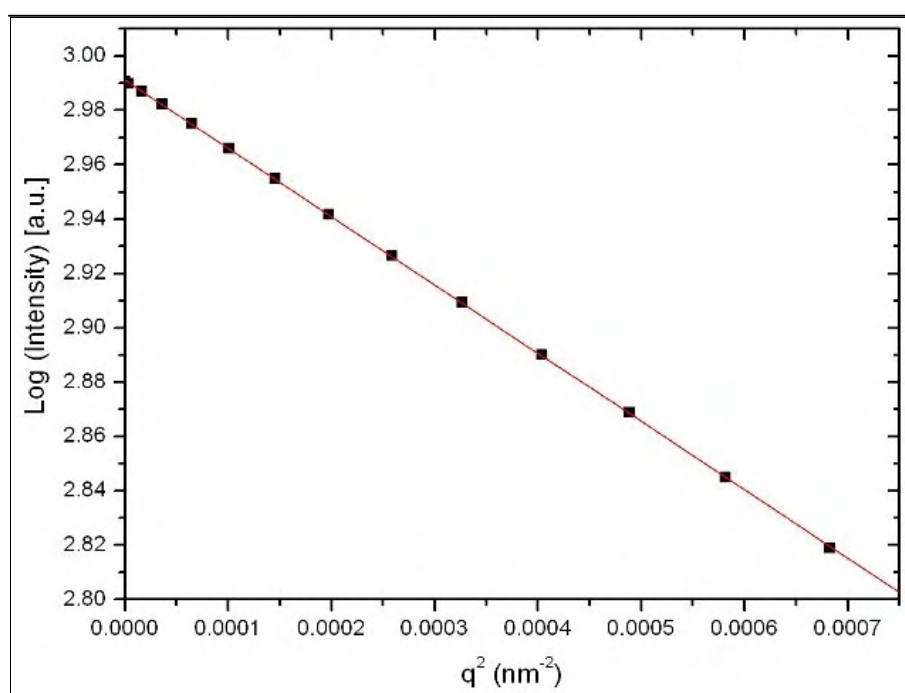


Figura 50. Gráfico de $\log I(q)$ versus q^2

Na figura 51 esta contida a curva de espalhamento extrapolada a concentração zero a qual foi utilizada para calcular a função de distribuição de distâncias $\rho(R)$ mostrada na figura 52.

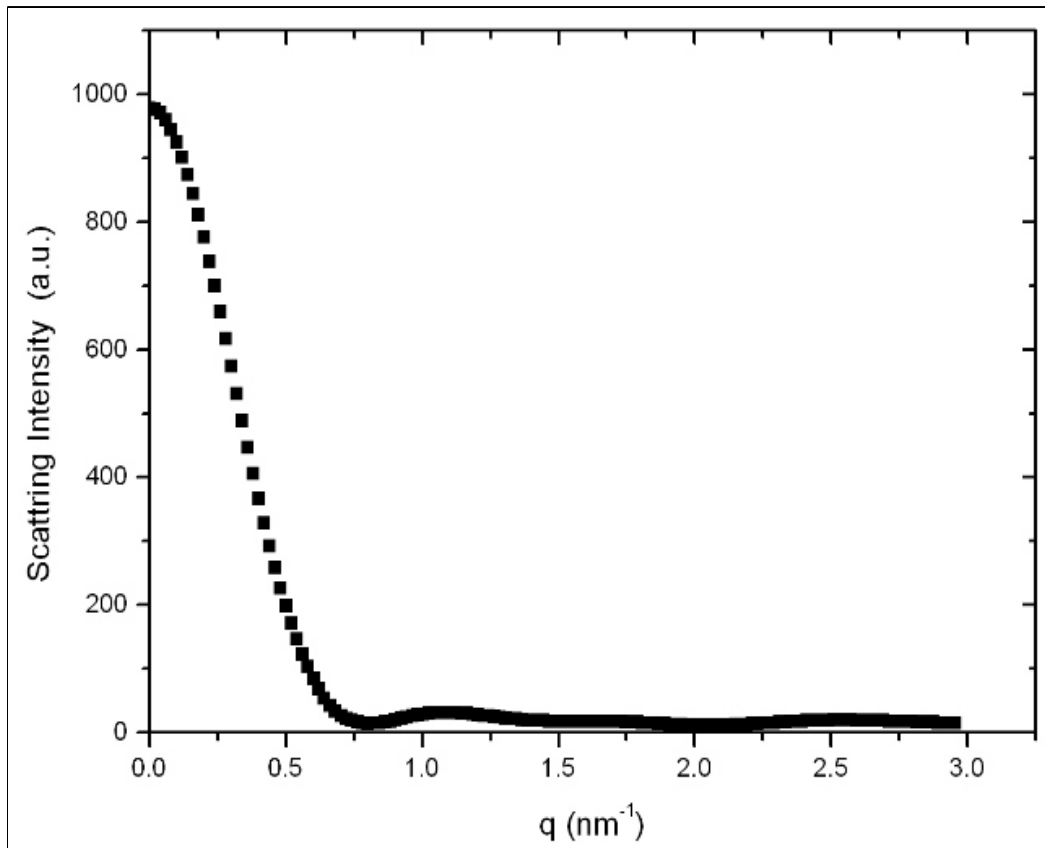


Figura 51. Curva de espalhamento da mtCS corrigida a feixe pontual e extrapolada a concentração e a intensidade zero.

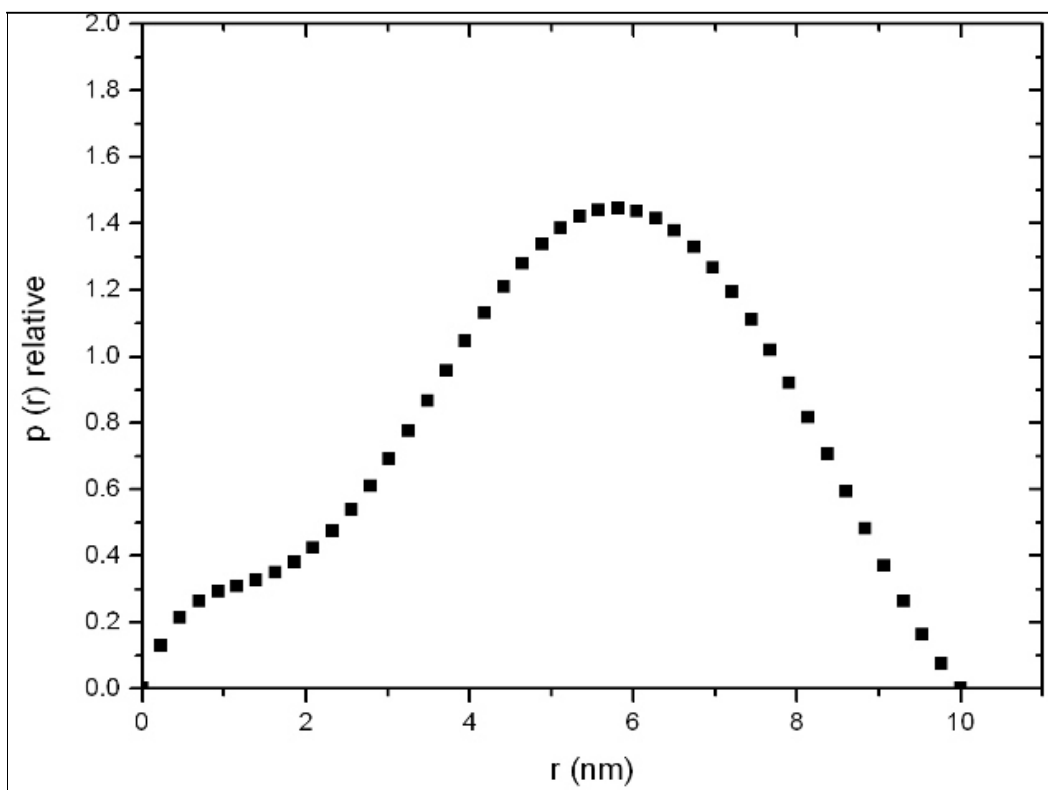


Figura 52. Função de distribuição de distância $p(r)$, com a máxima dimensão de 10nm.

Esta função $p(r)$ mostra um perfil típico de uma proteína com a forma de um esferóide achatado similar a um elipsóide oblato. Finalmente a sobreposição da função envelope obtida pelo método *ab initio* e o modelo tetramérico cristalográfico não deixa dúvidas que a proteína mtCS esta de acordo com a forma tetramérica em solução, nas figuras 53A – C onde são apresentados os modelos *ab initio* e a suas sobreposições com coordenadas cristalográficas.

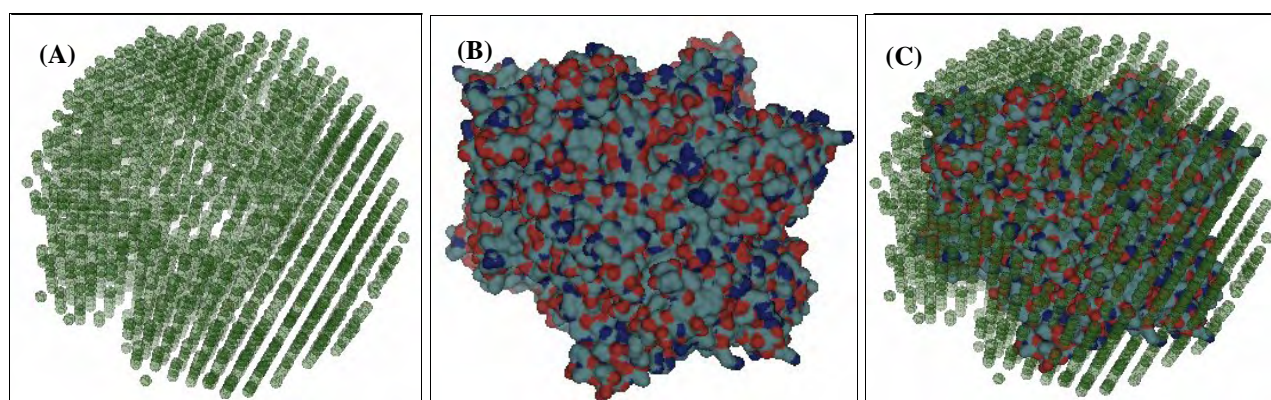


Figura 53. Modelo *ab initio* e sua sobreposição com a estrutura cristalográfica da mtCS na forma apo. (A) Modelo *ab initio* obtido por SAXS, (B) Modelo tetramérico das coordenadas cristalográficas geradas pelo CCP4 e (C) Sobreposição dos modelos *ab initio* com as coordenadas cristalográficas.

4.8. Docking molecular

A seleção dos melhores resultados foi realizada a partir da análise dos resultados dos parâmetros dos *poses* e das análises das interações de caráter hidrofóbico, eletrostático e tamanho das interações. Os resultados das cinco melhores posições foram verificados no próprio programa Molegro e aqui neste trabalho foi colocado o melhor dos cinco resultados.

O programa Molegro possui uma ferramenta que colore a enzima e o substrato dependendo de sua característica hidrofóbica (verde), hidrofílica (laranja) e neutra (branca). Utilizando essa ferramenta percebe-se como é a distribuição das regiões hidrofóbicas, hidrofílicas e neutras, nos ligantes e no sitio ativo das enzimas: mtCQ e mtCS.

4.8.1. Chiquimato sintase

Para a realização do *docking* do *olomoucine* e *roscovitine* na estrutura da mtCQ, foi utilizada a estrutura depositada no PDB com código de acesso 2IYZ, onde a molécula de ADP foi retirada, para a realização do *docking*, foi utilizado o programa Molegro na realização de ambos os *dockings* moleculares.

Na figura 54 pode-se ver como o *docking* previu o encaixe dos ligantes na cavidade do sitio ativo do ATP. Percebe-se que o ADP está bem encaixado na cavidade (primeira molécula mostrada na figura), já o ligante *olomoucine* apresenta somente algumas partes que se encaixaram, ou seja, ocorre a ligação do ligante, mas de uma forma pouco estável (segunda molécula), e o que mais chamou a atenção é que o *docking* molecular da *roscovitine* mostrou que este ligante se ligaria possivelmente uma parte na cavidade do sitio ativo do substrato D-Chiquimato (SHK) e outra parte da coenzima ATP o que justificaria os dados de CD feitos para este ligante que mostram que este ligante provoca uma maior instabilidade na enzima mtCQ (terceira molécula).

Por causa deste resultado do *docking* molecular da *roscovitine* foi feito ainda um outro *docking* molecular com este ligante, para verificar como ficaria sua ligação junto com o substrato SHK (ultimas moléculas), e foi visto que com a adição do substrato no *docking* molecular o ligante *roscovitine* fica na cavidade do sitio ativo do ATP, o que indica que possivelmente este ligante competiria com o substrato SHK pelo sitio.

Este tipo de competição se daria porque a cavidade do sitio de ligação do substrato SHK é mais hidrofóbica do que a cavidade do sitio de ligação do ATP e como o ligante *roscovitine* apresenta um perfil hidrofóbico, tende a procurar a região onde possa ficar o mais escondido possível do meio.

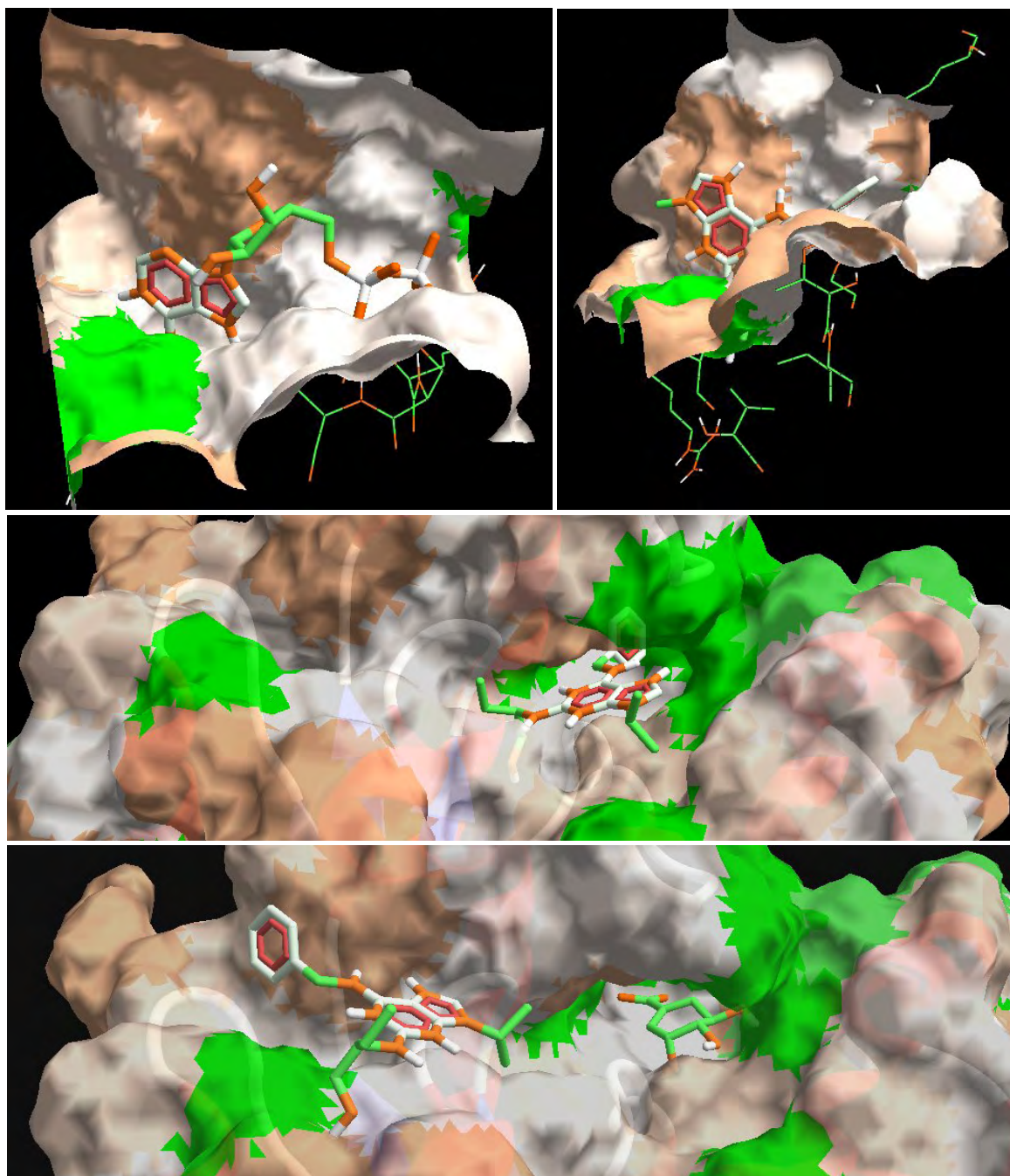


Figura 54. Diagrama das hidrofobicidades dos ligantes com mtCQ.

A interação entre os ligantes e a mtCQ também pode ser analisada através da superfície eletrostática da cavidade do sitio ativo de ligação. A figura 55 mostra a superfície eletrostática da região do sito ativo do ATP. A previsão das cores que estão na figura também é feita pelo programa Molegro, em azul é a região carregada positivamente, em vermelho é a região carregada negativamente e em branco é a região neutra.

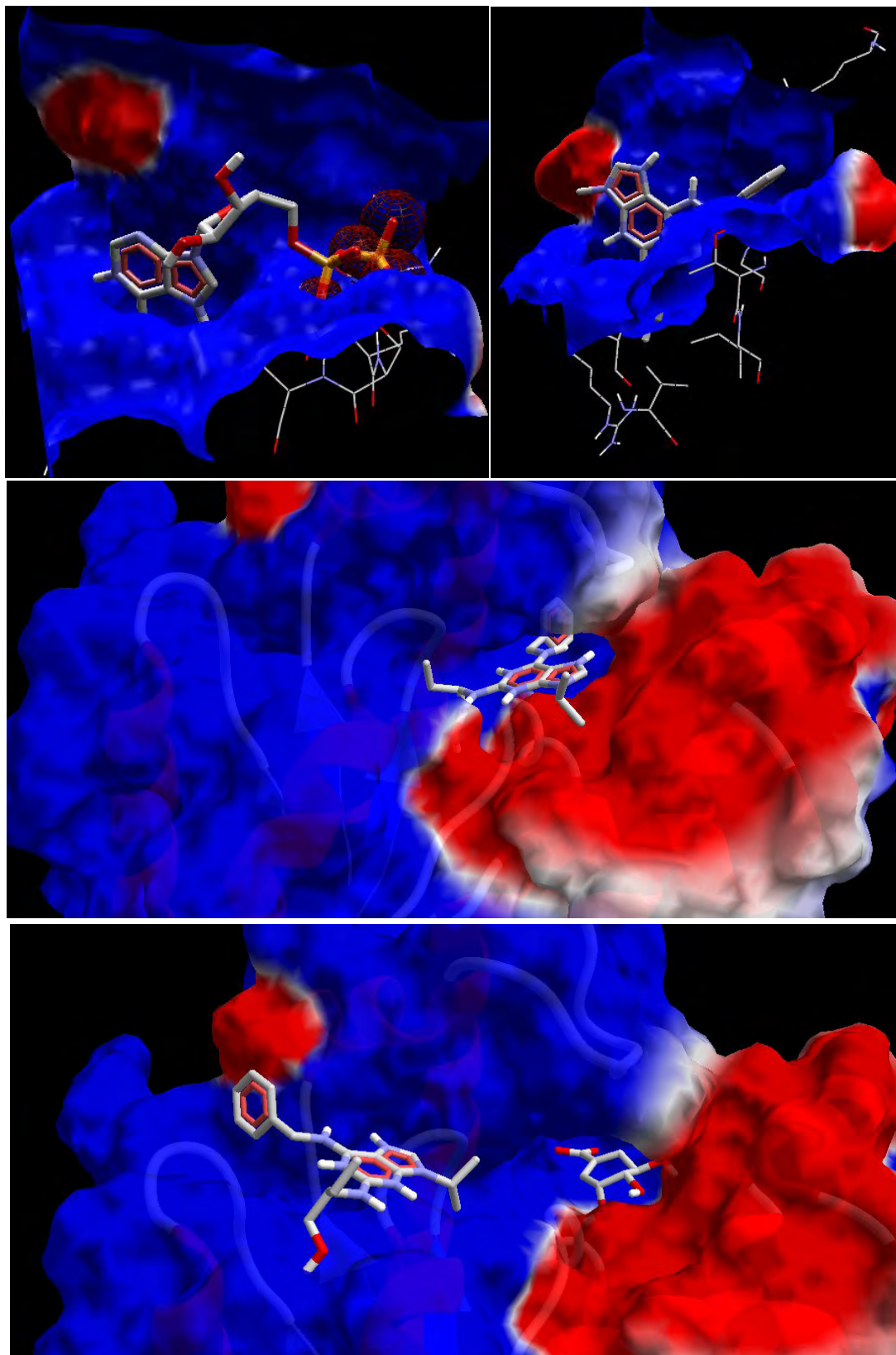


Figura 55. Diagrama das superfícies eletrostáticas dos ligantes com mtCQ.

Na figura 55 pode-se ver como o *docking* previu o encaixe dos ligantes na cavidade do sítio

ativo do ATP. Percebe-se que o ADP está bem encaixado na cavidade (primeira molécula mostrada na figura) por ter complementaridade de cargas no sitio ativo do ATP da mtCQ, já o ligante *olomoucine* e *roscovitine* não têm complementaridade de cargas no sitio ativo do ATP na mtCQ.

4.8.2. Corismato sintase

Para a realização do *docking* do EPSP e do FMN na estrutura da mtCS, foi utilizada a estrutura resolvida a 2.0 Å a ser depositada no PDB, foi utilizado o programa Molegro.

Pela figura 56 observa-se que as cores atribuídas ao substrato se encaixam exatamente na região da cavidade que possui as mesmas cores. A previsão feita pelo Molegro da característica de hidrofobicidade aparentemente está correta, pois como se sabe as hidroxilas são altamente hidrofílicas e elas receberam a coloração laranja (hidrofílico) e estas mesmas hidroxilas estão próximas das regiões hidrofílicas da cavidade do sitio ativo.

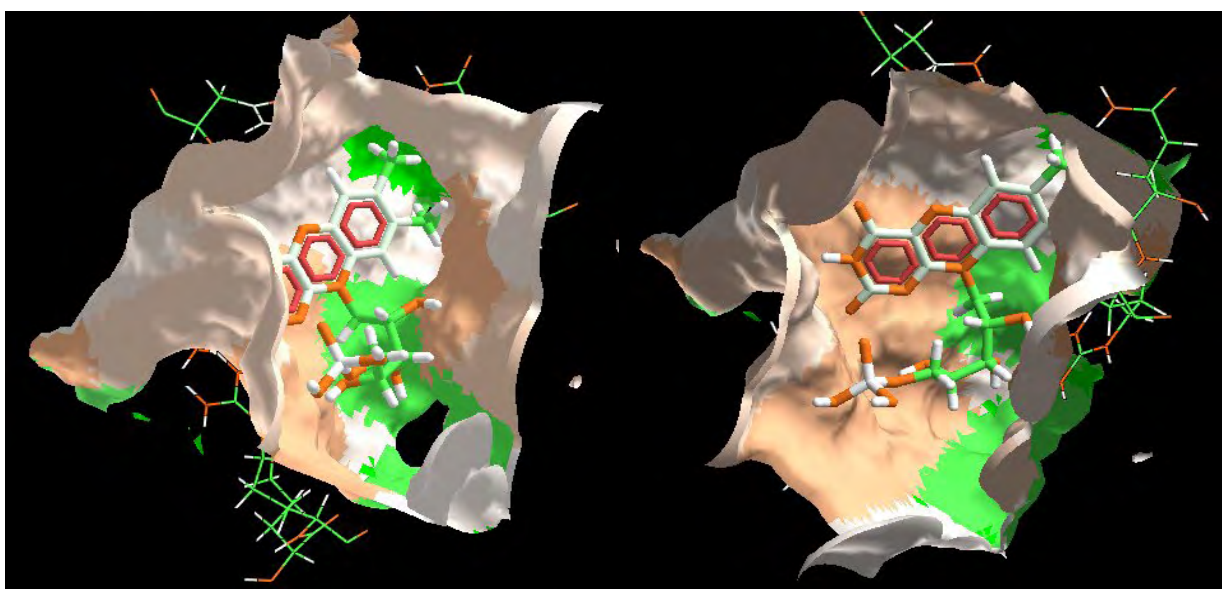


Figura 56. Diagrama da hidrofobicidade da região de interação da molécula de mtCS com o FMN. As cores tanto da molécula de mtCS quanto da molécula de FMN indicam: verde: região hidrofóbica, laranja: região hidrofílica e branco: região: neutra. Percebe-se que cada cor do substrato está na região da molécula de mesma cor.

O FMN parece se ligar na mtCS em uma região carregada positivamente devido a presença de vários resíduos de Arg. Através da figura 57 da superfície eletrostática da possível da região do

sito ativo da FMN pode ser verificada essa interação do FMN com resíduos carregados positivamente. A previsão das cores que estão na figura também é feita pelo programa Molegro, em azul é a região carregada positivamente, em vermelho é a região carregada negativamente e em branco é a região neutra.

Pode-se perceber também que a molécula de FMN está carregada negativamente, principalmente na região do fosfato. Dessa forma, a previsão do *docking* está de acordo com a análise da interação eletrostática, pois sendo a região da cavidade do sitio ativo carregada positivamente e a molécula de FMN em sua maior parte carregada negativamente, deve ocorrer uma atração entre as cargas opostas.

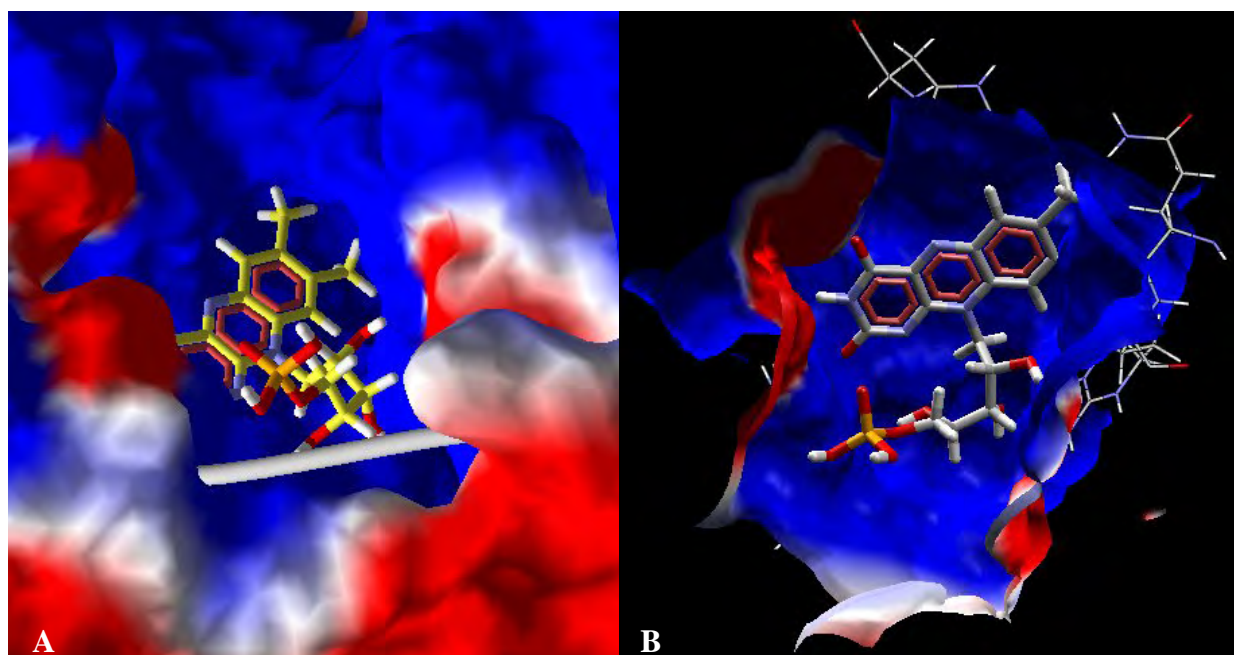


Figura 57. Figura da superfície eletrostática da região em que o substrato FMN se encaixa na mtCS. Azul: região positiva e vermelho: região negativa. (A) Zoom da superfície eletrostática da predição do sitio ativo do FMN. (B) Superfície eletrostática da interação do substrato FMN com a mtCS.

A molécula de EPSP é quase que completamente hidrofóbica, possuindo apenas algumas regiões hidrofílicas. Pela figura 58 pode-se perceber que a porção hidrofóbica do EPSP se encaixa na parte hidrofóbica da cavidade e apenas suas poucas porções hidrofílicas se encaixam na cavidade hidrofílica.

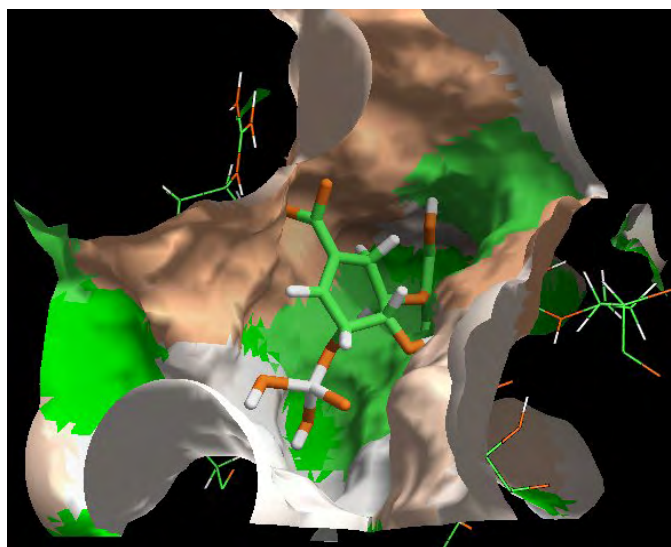


Figura 58. Mapa da hidrofobicidade da região de interação da molécula de mtCS com o EPSP, a molécula de EPSP está encaixada na região hidrofóbica. As cores tanto da molécula de mtCS quanto da molécula de EPSP indicam: verde: região hidrofóbica, laranja: região hidrofílica e branco: região neutra.

A interação entre o substrato e a enzima também pode ser analisada através da superfície eletrostática da cavidade do possível sítio de ligação. Na figura 59 em que a cavidade da enzima está na forma de superfície eletrostática (em azul região carregada positivamente, em vermelho região carregada negativamente e em branco região neutra), percebe-se que o EPSP está totalmente encaixado em uma região carregada positivamente, isto está de acordo com a molécula de EPSP que é envolvida por uma carga negativa, principalmente na região do grupamento fosfato.

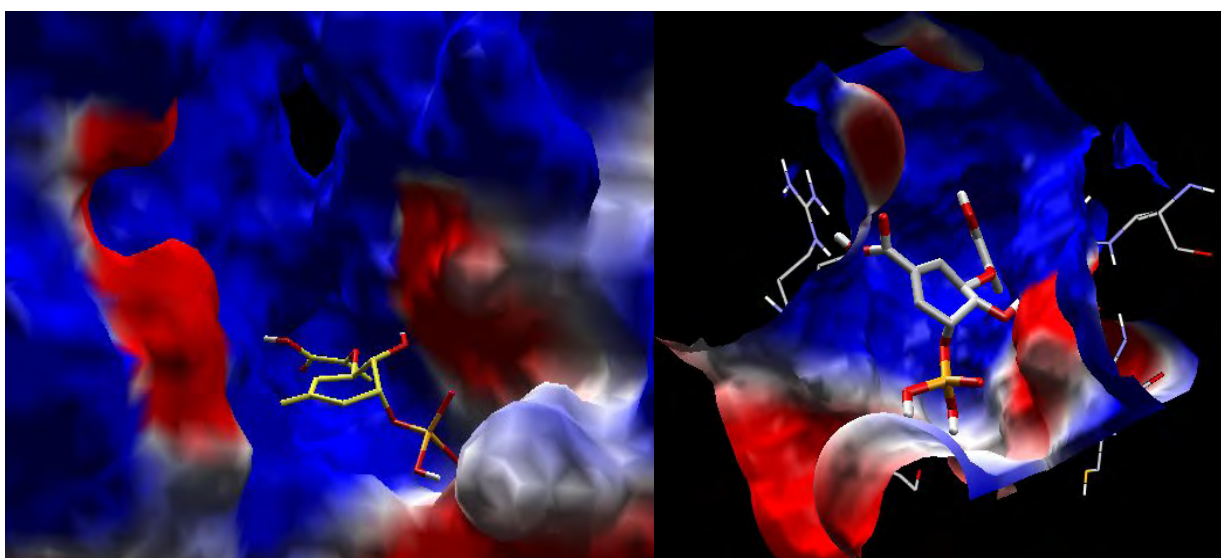


Figura 59. Superfície eletrostática do substrato EPSPS no possível sítio ativo da mtCS

Foi feito experimentos com *docking* para tentarmos prever a localização do sítio ativo para a FMN e o EPSP, já que a localização do sítio de ligação do FMN e do EPSP foram estudados em outros organismos como *Streptococcus pneumoniae*, e ocupam posições adjacentes no sítio ativo, observando que o sítio do FMN se encontra mais enterrado e o do EPSP mais próximo da entrada.

Sendo assim, deve haver uma ordem de ligação ao sítio ativo: primeiramente o FNM entra no sítio e o EPSP entra em seguida, fechando a entrada do mesmo, como sugerido também nos estudos feitos por CD acima discutidos.

Observa-se que a cavidade do sítio ativo das enzimas está localizada entre um dos arranjos de fitas e duas grandes alças que devem ter a função de abertura do sítio e a entrada dos ligantes. A região do sítio é rica em resíduos de Arg podendo então ser explicado o fato de que a região do sítio é quase totalmente carregada positivamente. Não foram encontrados resíduos que se conservassem no bolsão de ligação de ambas estruturas de corismato sintase, sugerindo que a especificidade está relacionada com a característica dos resíduos do sítio ativo, bem como àqueles ao redor do sítio.

5. Conclusões

Nesse trabalho o estudo estrutural realizado teve como finalidade expandir e compreender melhor o modo de interação da mtCS com seus ligantes naturais e da mtCQ com os análogos de ATP.

A modelagem molecular comparativa de proteínas em larga escala, na qual bancos de dados inteiros de seqüências ou genomas completos são usados como entrada em algoritmos de modelagem automatizada, tem sido utilizado por diversos grupos de pesquisa em bioinformática estrutural. Pelo uso de poderosos sistemas de computadores com múltiplos processadores, estes esforços têm permitido a criação de grandes bancos de dados de modelos de proteínas determinadas por modelagem molecular comparativa.

O SKPDB tem acesso público aos dados estruturais e às publicações. Este estudo de proteínas alvo de genomas completos enfatiza que a modelagem molecular é uma ferramenta de uso intensivo em biologia estrutural e que ela pode ser muito valiosa na anotação de seqüências de genomas e contribuir para a genômica estrutural e funcional.

A disponibilidade de modelos precisos tem aplicação em problemas biologicamente significantes tais como especificidade de substratos e ligantes e áreas como desenho de drogas. O desenho de drogas baseado em estrutura tornou-se uma tecnologia altamente desenvolvida que está em uso ativo nas maiores empresas farmacêuticas do mundo. Modelos moleculares de proteínas que são alvos preferenciais para desenho de drogas têm sido utilizados contra bases de dados de ligantes, com o objetivo de estimar sua afinidade usando *screening* virtual e simulações de *docking*, além de avaliar a energia livre de ligação absoluta da interação proteína-ligante tão precisamente quanto possível.

A utilização destes modelos implica em grau de precisão, no qual a identidade, o r.m.s.d., as análises de ambiente químico da proteína são de extrema importância, tendo que obedecer a alguns parâmetros de precisão. Tais critérios foram estabelecidos na seleção dos modelos e inseridos no

SKPDB para fornecer ao usuário modelos precisos para uso em simulações de *docking* e identificação de alvos terapêuticos.

6. Referências bibliográficas

- ALLEN, F. H. *et al.* The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. **Acta Cryst.** v.B35, p.2331-2339, 1979.
- BASSO, L.A.; BLANCHARD, J.S. Resistance to antitubercular drugs. **Adv. Exp. Med. Biol.** v.456, p.115-144, 1998.
- BASSO, L.A. *et al.* Mechanisms of isoniazid resistance in Mycobacterium tuberculosis: enzymatic characterization of enoyl reductase mutants identified in isoniazidresistant clinical isolates. **J. Infect. Dis.** v.178, p. 769-775, 1998.
- BAYAT, A. Science, medicine, and the future: Bioinformatics. **BMJ** v.324, p.1018-1022, 2002.
- BENACH, J.; *et al.* The 2.3Å crystal of the shikimate 5-Dehydrogenase orthologue YdiB from *E.coli* suggests a novel catalytic environment 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. **Crit. Rev. Biochem. Mol.Biol.** v.25, p.307-384, 1990.
- BERMAN, H. M. *et al.* The protein data bank. **Nucleic Acids Research** v.28, n.1, p.235-242, 2000.
- BOECKMANN B *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Res** v. 1; n.31(1), p. 365-370, 2003.
- BOURNE, P.E.; WEISSIG, H. **Structural Bioinformatics** Wiley-Liss; p. 509-524, 2003.
- 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.
- BRENNER, S.E. A tour of structural genomics. **Nat. Rev. Genet.** v. 2, n. 10, p. 801-809, 2001.
- BRÜNGER, A. T.; X-PLOR Version 3.1: A System for Crystallography and NMR, Yale University Press, New Haven, 1992.
- BUJNICKI, J.M. *et al.* LiveBench-1: continuous benchmarking of proteins structure prediction servers. **Protein Sci.** v.10, p. 352-361, 2001.
- BURLEY, S.K. *et al.* Structural genomics: beyond the human genome project. **Nat. Genet.** v. 23, p. 151-157, 1999.
- CARPENTER, E. P. *et al.* Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. **Nature** v.394, n.6690, p.299-302, 1998.
- CAMPBELL, M. M.; SAINSBURY, M.; SEARLE, P. A.; The biosynthesis and synthesis of shikimic acid, chirismic acid, and related compounds. **Synthesis-Stuttgart** v.2, p.179-193, 1993.
- CAMPBELL, S. A. *et al.*, A complete shikimate pathway in *Toxoplasma gondii* an accident eukaryotic innovation. **International Journal for Parasitology** v.34, p.5-13, 2004.

COGGINS, J. R. *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.

COLE, S.T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. **Nature** v.11, n. 393, p. 537-544, 1998.

CORBETT, E.L. *et al.* The growing burden of tuberculosis. **Arch. Intern. Med.** v.163, p. 1009-1021, 2003.

DAVIS, B. D; Aromatic biosynthesis. I. The role of shikimic acid. **Journal of Biological Chemistry** v.191, p.315-325,1950.

DAVIES, G.M. *et al.* (6S)-6-fluoroshikimic acid, an antibacterial agent acting on the aromatic biosynthetic pathway. **Antimicrob. Agents Chemother.** v. 38, p. 403-406, 1994.

DE AZEVEDO, W.F. JR. *et al.*; Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase. **Proc. Natl. Acad. Sci. USA.** v. **93**, n. 7, p. 2735-2740, 1996.

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.

DIAS, M. V. B. *et al.* Crystallization and preliminary X-ray crystallographic analysis of chorismate synthase from *Mycobacterium tuberculosis*. **Acta Crystal.** v. D60, 2003-2005, 2004.

DRENTH, J. **Principles of Protein X-ray Crystallography.** 2. ed. New York: Springer, 1999.

DUJON, B. The yeast genome project: what did we learn? **Trends Genet.** v. **12**, p. 263-270, 1996.

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.

ELY, B.; PITTARD, J. Aromatic amino acid biosynthesis: Regulation of shikimate kinase in *Escherichia coli* K-12. **J. Bacteriol.** v. 138, p. 933-943, 1979.

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; *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.

EYRICH, V.A. *et al.*; EVA: continuous automatic evaluation of protein structure prediction servers. **Bioinformatics** p. 17, v. 242-1243, 2001.

FISER, A.; DO, R. K.; SALI, A.; Modeling of loops in protein structures, **Protein Sci.** v.9, p.1753-1773, 2000.

FEIGIN, L.A.; SVERGUN, D.I. Structure analysis by Small-angle X-ray and Neutron Scattering, Plenum Press, New York, 1987.

FOSTER, M.J. Molecular modelling in structural biology. **Micron** v.33, p.365-384, 2002.

GALPERIN MY. The Molecular Biology Database Collection: 2005 update. *Nucleic Acids Res.* v. 1, p. 33, D5-24.

GANEM, B.; From glucose to aromats: recent developments in natural products of the shikimic acid pathway. **Tetrahedron** v.34, p.3353-3383, 1978.

GARMAN, E. F.; SCHNEIDER, T. R.. Macromolecular cryocrystallography. **Journal of Applied Crystallography**, n. 30, p. 211-237, 1997.

GIACOVAZZO, C. The diffraction of X-rays by crystals. In: GIACOVAZZO, C. *et al.* **Fundamentals of Crystallography**. Oxford: Oxford University Press, 1992. p. 141-228.

GIBSON, F.; The elusive branch-point compound of aromatic amino acid biosynthesis. **Trends in Biochemical Sciences** v.24, p.36-38, 1999.

GLATTER, O.; KRATKY, O. Small Angle X-Ray Scattering, Academic Press Inc. Ltd., New York, 1982.

GREENFIELD, N. J. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. **Anal. Biochem.**, v. 235, p. 1-10, 1996.

GU, Y. *et al.* Cristal Structure of Shikimate Kinase from *Micobacterium tuberculosis* Reveals the Dynamic Role of the LID Domain in Catalysis. *J. Mol. Biol.* v.319, p.779-789, 2002.

GUINIER, A.; FOURNET, G. Small-Angle Scattering of X-Rays, John Wiley & Sons, New York, 1955.

HAGEN JB. The origins of bioinformatics. **Nat Rev Genet** v.1, n. 3, p. 231-236, 2000.

HASLAM, E. **Shikimic acid: metabolism and metabolites**. John Wiley & Sons, Chichester, United Kingdom, 1993.

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, 1978.

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.

HOOFT, R. W. *et al.* Errors in protein structures. **Nature** v.381, p.272, 1996.

HÖLTJE, H. D.; FOLKERS, G. **Molecular Modeling- Basic principles and applications**. New York: VHC, p. 194, 1996.

KEARSLEY, S. K.; On the orthogonal transformation used for structural comparisons. **Acta Crystallogr.** v.45A, p.208-210, 1989.

KELLY, S. M.; PRICE, N. C. The application of circular dichroism to studies of protein folding and unfolding. **Biochim. Biophys. Acta** v.1338, p.161-185, 1997.

KISHORE, G.M.; SHAH, D.M. Amino acid biosynthesis inhibitors as herbicides. **Annu. Rev. Biochem.** v. 57, p. 627–663, 1998.

KITANO, H. Computational system biology. **Nature** v. 420, p. 206–210, 2002.

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.

LAKOWICZ, J.R. Principles of fluorescence spectroscopy, Plenum Press, New York, 341-379, 1983.

LASKOWSKI, R.A.; *et al.*, PROCHECK - A program to check the stereochemical quality of protein structures, **J. Appl. Cryst.** v.26, p.283–291, 1993.

LESK, A.M.; **Introduction to Protein Architecture**, Oxford University Press, Oxford, Great Britain, 2001.

LESLIE, A. G. W.. A reciprocal-space method for calculating a molecular envelope using the algorithm of B. C. Wang. **Acta Crystallographica A**, n. 43, p. 134-136, 1987.

LINDER, P.; ZEMB, T. Neutron, **X-Ray and Light Scattering: Introduction to an Investigative Tool for Colloidal and Polymeric Systems**, New York, Elsevier Science Publisher, 1991.

LIU, J.; ROST, B. Target space for structural genomics revisited. **Bioinformatics** v. 18, n. 7, p. 922-933, 2002.

LUTHY, R.; BOWIE, J. U.; EISENBERG, D.; Assessment of protein models with three-dimensional profiles. **Nature** v.356, p.383-385, 1992.

MACHEROUX, P. *et al.* Evidence for a major structural change in *Escherichia coli* chorismate synthase induced by flavin and substrate binding. **J. Biochem.** v. 335, p. 319–327, 1998.

MACHEROUX, P. *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.

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.

MARTI-RENOM, M.A. *et al.* Alignment of protein sequences by their profiles. **Protein Sci.** v.13, p. 1071-1087, 2004.

MATHEWS, C.K.; VAN HOLDE, K.E. *Biochemistry*, Benjamin Cummings, 1130p., California, USA, 1990.

MCCREE, D.E.; XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* v.125, n.(2-3),p.156-65, 1999.

MELO, F.; FEYTMANS, E. Assessing proteins structures with a non-local atomic interaction energy. *J. Mol. Biol.* v. 277, p. 1141-1152, 1998.

MILLAR, G. *et al.* The cloning and expression of the *aroL* gene from *Escherichia coli* K-12. *J. Biochem.* v. 237, p. 427-437, 1986.

MORGAN, J.; HARITAKUL, R.; KELLER, P. A. *Bioorg. Med. Chem. Lett.*, v. 13, p. 1755-1757, 2003.

MOUSDALE, D.M. *et al.* Purification and characterization of bifunctional dehydroquinase-shikimate: NADP oxidoreductase from pea seedlings. *Phytochemistry* v. 26, p. 2665-2670, 1987.

NAGEM, R.A.P **Otimização das condições para a aquisição de dados de derivados e para a determinação das fases dos fatores de estrutura de cristais de proteínas por meio da difração da luz síncrotron.**, Campinas, 2003. 143p. Tese (Doutorado) - Instituto de Física Gleb Wataghin, UNICAMP.

NAVAZA, J.; AMoRe: an automated package for molecular replacement, *Acta Crystallogr.* v.A50, p.157-163, 1994.

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.

NUNN, P; WILLIAMS, B.; FLOYD, K; DYE, C; ELZINGA, G; RAVIGLIONE, M. *Nature Reviews*, v. 5, p. 819-826, 2005.

POLIKARPOV, I. *et al.*. Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. *Journal of Synchrotron Radiation*, v. 5, p. 72-76, mar. 1998.

POLIKARPOV, I. *et al.*. The protein crystallography beamline at LNLS, the Brazilian National Synchrotron Light Source. *Nuclear Instruments and Methods in Physics Research A*, n. 405, p. 159-164, 1998.

OLIVER, S.G. From DNA sequence to biological function. *Nature (London)* v. 379, p. 597-600, 1996.

ORENGO, C. A. *et al.* The CATH protein family database: A resource for structural and functional annotation of genomes. *Proteomics* v.2, p.11-21, 2002.

OUZOUNIS CA; VALENCIA A. Early bioinformatics: the birth of a discipline--a personal view. *Bioinformatics* v.19, n. 17, p. 2176-2190, 2003.

PELCZAR M. J. JR. *et al.* **Microbiologia – Conceitos e Aplicações**, Makron Books, v. 2, 2ed.,

517p., São Paulo, 1996.

PEREIRA, J. H. *et al.* Structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimic acid. **Acta Cryst.** v.D60, p.2310–2319, 2004.

PETOUKHOV, M.V.; SVERGUN, D.I. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data, *Biophys. J.* 89, 1237-1250, 1995.

PITTARD, A.J. *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. In: F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.) American Society for Microbiology, Washington, D.C., v. 1.1, p. 368–394.

QUE, X. *et al* Cysteine proteinases from distinct cellular compartments are recruited to phagocytic vesicles by *Entamoeba histolytica*. **Mol. Biochem. Parasitol.** v. 119, p. 23-32, 1993.

RAMJEE, M. N. *et al.* Reaction of (6R)-6-F-EPSP with recombinant *Escherichia coli* chorismate synthase generates a stable flavin mononucleotide serni-quinone radical. **J. Am. Chem. Soc.** v.114, p.3151-3153, 1992.

RATLEDGE, C. **Nutrition, growth and metabolism** in: Ratledge, C. and Stanford, J. L. (Eds.), *The biology of the mycobacteria*, Academic Press, London, v. 1, p. 185-271, 1982, 1987.

RING, C.S. *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

ROBERTS, F. *et al.* Evidence for the shikimate pathway in apicomplexan parasites. **Nature** v. 393, p. 801-805, 1998.

RUSSO, A.A. *et al.*, **Nature**, (Lodon) 382, 325-331 (1996)

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.

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

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

SAQI, M.A. *et al.* Misleading local sequence alignments: implications for comparative protein

modelling. **Protein Eng.** v. 11, p. 627-630, 1998.

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. *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. *et al.* Purification and characterization of chorismate synthase from *Euglena gradis*. **Plant Physiol.** v.97, p.1271 -1279, 1991b.

SCHÖNBRUNN, E. *et al.*, Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. **Proc. Natl. Acad. Sci. USA**, v.98, p.1376-1380, 2001.

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

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.

SIMMONS, K. T. *et al.* Ab initio protein structure prediction of casp iii targets using rosetta. **Proteins Suppl** 3, p. 171-176, 1999.

STEINRÜCKEN, H. C.; AMRHEIN, 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-urabino-heptulosonate 7-phosphate synthase from *Escherichia coli*. **J Biol Chem** v.267, p.5762-5767, 1992.

SVERGUN, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria, **J. Appl. Cryst.** v.25, p.495-503, 1992.

SVERGUN, D. I., BARBERATO, C.; KOCH, M. H. J. CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, **J. Appl. Cryst.**, v.28, p.768-773, 1995.

SVERGUN, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. **Biophys. J.** v.76, p.2879-2886, 1999.

SVERGUN, D.I., PETOUKHOV, M.V.; KOCH, M.H.J. Determination of domain structure of proteins from X-ray solution scattering. **Biophys. J.** v.80, p.2946-2953, 2001.

SVERGUN, D.I.; KOCH, M.H.J. Small-angle scattering studies of biological macromolecules in solution, **Rep. Prog. Phys.** v.66, p.1735-1782, 2003.

TAYLOR, W.R. A 'periodic table' for protein structures. **Nature** v. 11, n. 416, p. 657-660, 2002.

TOPHAM, C.M. *et al.* Comparative modelling of major house dust mite allergen Der p I: structure validation using an extended environmental amino acid propensity table. **Protein Eng.** v. 7, p. 869-894, 1994.

TROUILLER, P. *et al.* Drugs for neglected diseases: a failure of the market and a public health failure? **Tropical Medicine and International Health**, v. 6, p. 945-951, 2001.

UCHÔA, H.B. *et al.* Parmodel: a web server for automated comparative modeling of proteins. **Biochem. Biophys. Res. Commun.** v. 325, p. 1481-1486, 2004.

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.

VOLKOV, V.V.; SVERGUN, D.I. Uniqueness of *ab initio* shape determination in small-angle scattering. **J. Appl. Cryst.**, v.36, p.860-864, 2003.

VEERAPANDIAN, P. **Structure-based drug design**, (Dekker, M., ed.), INC. New York., 1997.

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.

WESTBROOK, J.; FENG, Z.; JAIN, S.; *et al.* The Protein Data Bank: unifying the archive. **Nucleic Acids Res** v.30, p.245-248, 2002.

WHIPP, M. J. & PITTARD, A. J. A reassessment of the relationship between *aroK*- and *aroL*-encoded shikimate kinase enzymes of *Escherichia coli*. **J. Bacteriol.**v.177, 1627-1629, 1995.

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.

WHO (2002). Tuberculosis. Fact Sheet n° 104. World Health Organization, Geneva, Switzerland. August/2002.

WOODY, R.W. Circular dichroism. **Methods Enzymol.** v.246, p.34-71, 1995.