

Cécile Fleury

**Ferramentas de Bioinformática
dedicadas ao estudo
das Relações Estrutura-Função-Antigenicidade
em Toxinas Peptídicas Animais**

Orientador : Prof. Dr. Paulo Sérgio Lacerda Beirão, LAMEX,
Departamento de Bioquímica e Imunologia, ICB, Universidade
Federal de Minas Gerais, Brasil

Co-orientador : Dr. Franck Molina, SysDiag, UMR3145
CNRS/Bio-Rad, Montpellier, França

Tese em Bioinformática,
submetida ao Departamento de Bioquímica e Imunologia,
Instituto de Ciências Biológicas,
Universidade Federal de Minas Gerais,
como requisito parcial para a obtenção do título de
Doutor em Ciências.

Belo Horizonte
Minas Gerais – Brasil
2009

AGRADECIMENTOS

Esperando não cair no melodrama, preciso contar aqui que tenho muita gente a agradecer ao concluir um projeto como este. Pois não se trata “somente” de um projeto de pesquisa, mas também de um importante pedaço de vida. Quatro anos riquíssimos em experiências diversas, encontros, desencontros e reencontros com inúmeras pessoas (agora me parecem bilhões... como farei para não esquecer ninguém?), aprendizagens em disciplinas da ciência e da vida, tão múltiplas quanto remotas, esperanças e desesperos, desânimo e motivação, frustrações e alegrias, e tudo muito muito muito intenso. Somando tudo o que vivi aqui no Brasil e lá na França, é um monte de felicidade que me enche e que levo embora comigo ao virar esta página de vida.

Por isso e por tudo o que minha vinda para cá proporcionou, agradeço a todas as pessoas que, de perto ou de longe, em qualquer medida, bem ou mal, contribuíram para fazer da minha vida profissional e pessoal o que ela é hoje e o que será no futuro...

Primeiramente, agradeço ao Dr. Molina e ao Prof. Beirão por ter possibilitado este projeto, por ter me mandado e recebido (respectivamente) neste maravilhoso Brasil, por ter me confiado este trabalho e por ter me orientado e apoiado. Fico imensamente grata por tudo que fizeram.

Agradeço também aos Profs. Carlos Chávez, Evanguedes, Maria-Elena, Adriano Pimenta, no Brasil, e ao Prof. Claude Granier na França, por terem construído e enriquecido nossa colaboração científica dedicada ao estudo das toxinas animais. Agradeço também pelo jeito pelo qual me receberam e cuidaram de tudo para que eu me adaptasse rapidamente, assim como pelo exemplo profissional.

Agradeço ao Dr. Frédéric Ducancel, ao Prof. Pedro Pascutti, ao Prof. Frederico Campos Filho e ao Prof. Carlos Chávez por terem aceitado participar da banca da minha defesa. Me sinto honrada pela disponibilidade e interesse dos membros dessa banca.

Agradeço...

... Aos meus amigos do Brasil, todos tão queridos e insubstituíveis. Já sinto uma saudade imensa de todos que dói... Por ordem aproximada de aparição: Juliana, minha “mãezinha” querida, que me recebeu, ajudou, ensinou, cuidou de mim com um imenso carinho quando cheguei; Chico Lobo, meu amigão, meu parceirão de conversas e desconversas intermináveis e apaixonantes, debates e afins, em inglês no início e depois em mineirês com gírias, “sô”; Adriano, meu amigo queridíssimo, meu

colega de aula, de curso, de congresso, de festas, de dança (para mim, o melhor dançarino do Brasil) e de viagens para a Serra do Cipó, uma das primeiras pessoas que teve a paciência de decifrar meu néo-português; Chicão, amigo e objeto de um intenso estudo meu, cujo resultado hoje é admiração, surpresa, carinho, gosto, reflexão perpétua e mudança, assim como tantas outras coisas boas que me proporcionaram essa convivência especial; Larissa, minha amiga linda que entende tudo, sente tudo, tem as palavras certas na hora certa, sabe e compartilha com gosto, empolga e transmite alegria; Bernardo, o amigo mais bem humorado e carinhoso que alguém possa sonhar, o anjo biólogo, a pessoa mais otimista, divertida, calorosa e amada que conheci nesse lindo país; Liza, a mulher mais generosa, luminosa, fofa, um amor de pessoa; Miguelito, meu grande amigo, sempre por perto para dar apoio ou dançar forró, conselheiro e animado, muito querido; Fabiana, “muié”, minha *roommate* e amiga, a preferida do nosso chefe (a razão!), cinéfila e colega da “promoção casal” de pipoca no Del Rey, enquanto sentadas contra a parede no lado direito da sala de cinema; Cris, minha formigamiga elétrica, minha segunda mãezinha, quase irmã também, até espelho as vezes, amiga inteira e indispensável... e Gui, a quem agradeço pela paciência, pelos biscoitos de queijo (7 para cada um), e pelos palpites sobre A Favorita!; Carolinda, a amiga mais sensata, franca e objetiva, amiga rara e preciosa; Alexandra, minha amiga do Portugal, generosa e atenciosa, além de engraçadíssima, e rara demais também (fala português todo errado, infelizmente); Cris Espetáculo, amiga adorável, generosíssima, engraçadíssima, e muito pouco louca; Ciça, minha amiga confidente, inteligente e interessante, muito querida, discreta e importante; Máira, *last but not least*, recém-chegada mas em tão pouco tempo já muito amiga, divertida e sempre disposta, como já lhe disse, arrependo que esteja chegando a hora de ir embora, pelo tanto de coisas que eu queria ainda poder compartilhar e viver aqui...

... Aos meus amigos da França, também extremamente queridos e também excepcionais, especialmente: Robin, Gilles, Mikaël, Barthélémy, Floriane, Mathieu, Maya, Anthony, Anne, Ryselenn, Sabrina, Virginie, Philippe, Nabil, Vincent.

... Aos amigos e colegas de Montpellier: Violaine, minha primeira orientadora, hoje colaboradora e também amiga muito querida, agradeço por tudo; David, essencial e discreto, agradeço pela preciosa ajuda mas também pela confiança e pelos bons momentos no trabalho, as conversas e os cafés; Sabine, amiga a primeira vista, agradeço por ter me recebido e apoiado, e também pela super companhia; Benjamin pelos ensinamentos e pelo carinho; Pierre, pelo exemplo, pelo contra-exemplo (“rm*”, por enquanto não apaguei todos os meus dados!), por BioΨ e pela convivência na torre do laboratório; Christophe, agradeço “quadradamente” pelo bom humor e pelo convite na sala dos químicos asiáticos.

... Aos meus colegas de trabalho (e de diversão) do ICB, do LAMEX, da Bioinformática, da UFMG inteira e do Sysdiag, “randomicamente”: Maurício, Rodrigo Vêras, Stéphanie, Quelé, Eve, Dung, Esquisito, Lionel, Rodriban, Mário, Bellinha, Jean-Daniel, Raoni, Michael, Jota, Ricardo, Isabelle, Priscilão, Laurence, Moscão, Aline, Sandrine, Anita, Gabrielzim, Rafaela, Dudu, Clara, Alain, Raquel, Eva, Carol, Karine, Daiane, Nicolas, Luara, João, assim como todos os membros dos laboratórios do quarto andar do ICB e da UMR3145.

... Aos amigos da moradia universitária e à FUMP!

... À Prof^a Glória, ao Prof. Ronaldo Nagem, ao Prof. Marcelo Santoro e ao Prof. José Miguel Ortega por todos os conselhos e ensinamentos.

... Ao secretário da Bioinformática, Carlos, pelos serviços prestados tão bem e tão rapidamente, agradeço e peço desculpas pelos transtornos ocasionados.

Sei que não lerão essas linhas, mas quero também agradecer à Serra do Cipó, ao Rio de Janeiro, à lagoa da Pampulha, ao Buteco da Bio, à Lavras Novas, ao mar, à goiabada, ao pão de queijo e ao cafezim... pois sem eles não haveria tese alguma!

E finalmente agradeço à minha família...

... Aos meus pais, que me incentivaram, aconselheram e apoiaram. Ao meu pai, devo o meu gosto pela ciência e pela bossa nova, à minha mãe devo meu gosto pelas viagens.

... Aos meus irmãos Alexis e Pierre, e a minha irmã Marine. Agradeço a eles por ter nascido na mesma família que eu, e por serem essas pessoas divertidas e diferentes.

... Aos meus avôs e avós queridas: ao meu avô Charles de quem sinto muita falta, à minha avó Paulette que eu queria ver muito mais, à minha avó Denise, e também ao meu Vô, por serem essas pessoas apaixonantes que têm mil histórias novas para contar cada dia, mil receitas de cozinha para provar, mil piadinhas e mil sorrisos que descomplicam a vida.

... Aos meus tios, especialmente Gilles, e minhas tias Yvonne e Cathy, pelo exemplo, pelos conselhos e pelo carinho.

... Aos meus primos, especialmente ao Charles-David, pela convivência ao longo do meu mestrado em Paris, pelas conversas, pelos passeios, pelas idéias.

E ao Rodrigão, pela paciência, pelo interesse, pelo carinho, pela calma, pelas músicas, pela *extreme-biology*, pela aleatoriedade, pela atenção, pelos cafés, pelas conversas, pelo amor e pelo futuro.

RESUMO

Os venenos de aracnídeos contêm uma grande quantidade de neurotoxinas peptídicas que constituem um sério problema de saúde pública no Brasil e em outras partes do globo. Por outro lado, elas representam ferramentas valiosas para estudar os canais iônicos e apresentam um grande potencial para aplicações biotecnológicas.

Essas toxinas diferem entre si em termo de sequências, estruturas tridimensionais (3D), e alvos moleculares. O conhecimento das suas estruturas 3D é fundamental para entender seus mecanismos de ação e caracterizar suas interações com os receptores. Infelizmente, somente poucas estruturas são disponíveis até agora, devido às limitações experimentais.

Em grande maioria, suas sequências são notavelmente ricas em resíduos de cisteína. O fato que o número de *fold*s de toxinas de aranhas e escorpiões seja pequeno, junto à observação que padrões comuns de distribuição de resíduos de cisteína e padrões comuns de conectividade correspondem geralmente a um enovelamento similar, nos levou a propor que, apesar da variabilidade de sequências, um modelo molecular confiável pode ser produzido para cada toxina, baseado na similaridade em termos de padrão de cisteína.

Após ter removido as sequências de toxinas fragmentadas ou desprovidas de resíduos de cisteína, 630 sequências de toxinas peptídicas de aranhas e escorpiões foram selecionadas e inseridas num novo banco de dados chamado SiliTox, dedicado às investigações das relações estrutura-antigenicidade-função das toxinas peptídicas animais. Padrões de cisteína e pontes dissulfeto foram designados, e grupos estruturais foram estabelecidos de acordo com esses. Das 630 sequências, 350 pertencem a grupos estruturais que incluem pelo menos uma estrutura 3D que tenha sido previamente determinada experimentalmente e pode ser usada como molde (*template*) molecular. Um programa de modelagem automático chamado SiliMod foi desenvolvido para efetuar a modelagem comparativa das toxinas, baseada na similaridade dos padrões de cisteína e de conectividade. SiliMod usa os dados estruturais armazenados no banco SiliTox e utiliza o programa Modeller.

SiliMod produziu 350 modelos estruturais com sucesso e o método foi avaliado. Duas outras ferramentas de predição foram criados para analisar as potencial regiões de interação nas superfícies das toxinas. Os programas foram chamados PEPOP e SiliPop. Análises das 630 toxinas de aracnídeos presentes em SiliTox foram realizadas e levaram à predição de 2863 peptídeos mimetizando regiões das toxinas possivelmente envolvidas em interações proteína-proteína. O modelo 3D e a superfície de interação propostos das toxinas Tx2-5 e Tx2-6 da aranha *Phoneutria nigriventer* foram confrontados com resultados experimentais obtidos no nosso laboratório. Os resultados evidenciaram uma região provavelmente envolvida na ligação dessas toxinas nos canais de sódio dependentes de voltagem.

ABSTRACT

Venoms of arachnids contain a large amount of peptide neurotoxins that represent a serious health issue in Brazil and in other parts of the globe. Besides, they are valuable tools to study ion channels and show a great potential for biotechnological applications.

These toxins differ among each other in terms of sequence, three-dimensional (3-D) structure and molecular target. Knowing their 3-D structure is fundamental to understand their mechanisms of action and characterize their interactions with receptors. Unfortunately, only few structures are available to date, due to experimental limits.

Their sequences are at their huge majority notably rich in cysteine residues. The fact that the number of spider and scorpion toxin known folds is limited, together with the observation that common cysteine distribution and connectivity patterns generally correspond to a similar folding lead us to propose that, despite the sequence variability, a reliable molecular model can be produced for each toxin based on the similarity in term of cysteine pattern.

After removing fragment and without-cysteine-residue toxins, 630 spider and scorpion peptide toxin sequences were selected and inserted into a new database called SiliTox, dedicated to investigations of the animal toxin structure-antigenicity-function relationships. Cysteine patterns and disulfide bridge connectivity were assigned, and structural groups were established according to them. Out of the 630 sequences, 350 toxin sequences fell into a structural group that included one or more experimentally determined template structure(s). An automated modeling program called SiliMod has been developed to perform the toxin comparative modeling based on cysteine pattern and connectivity similarity. SiliMod uses the structural data stored into SiliTox and Modeller software.

SiliMod successfully produced 350 structural models and the method was evaluated. Two other prediction tools have been designed to analyze the potential regions of interaction on the toxin surfaces. The programs were called PEPOP and SiliPop. Analysis of the 630 arachnid toxins present in SiliTox has been carried out and led to the prediction of 2863 peptides mimicking toxin regions likely to be involved in protein-protein interactions. The 3-D model and the predicted interacting surface of *Phoneutria nigriventer* spider toxins Tx2-5 and Tx2-6 were confronted to experimental results carried out in our laboratory. The results point out a region likely to be involved in the binding of these toxins to voltage-gated sodium channels.

LISTA DAS ABREVIACÕES

Na ⁺	sódio
Ca ⁺	cálcio
K ⁺	potássio
Cl ⁻	cloreto
Na _v	canal de sódio dependente de voltagem
Na _v 1.2	canal de sódio dependente de voltagem subtipo 1.2
ASIC	canal iônico sensível ao ácido
MSC	canal mecano-sensível
nAChR	receptor nicotínico para acetilcolina
AaH II	toxina II do escorpião <i>Androctonus australis</i> Hector
Lqh2	toxina 2 do escorpião <i>Leiurus quinquestriatus hebraeus</i>
LqhαIT	toxina αIT do escorpião <i>Leiurus quinquestriatus hebraeus</i>
BmK M1	toxina M1 do escorpião <i>Buthus martensi</i> Karsch
Css4	toxina 4 do escorpião <i>Centruroides suffusus suffusus</i>
Ts VII	toxina VII do escorpião <i>Tityus serrulatus</i>
Lqhβ1	toxina β1 do escorpião <i>Leiurus quinquestriatus hebraeus</i>
AahIT	toxina IT do escorpião <i>Androctonus australis</i> Hector
Bj-xtrIT	toxina xtrIT do escorpião <i>Buthotus judaicus</i>
LqqIT2	toxina IT2 do escorpião <i>Leiurus quinquestriatus quinquestriatus</i>
BjIT2	toxina IT2 do escorpião <i>Buthotus judaicus</i>
PnTx2-5	toxina Tx2-5 da aranha <i>Phoneutria nigriventer</i>
PnTx2-6	toxina Tx2-6 da aranha <i>Phoneutria nigriventer</i>
HaTx1	hanatoxina 1 da aranha <i>Grammostola spatulata</i>
GrTx	ω-gramotoxina SIA da aranha <i>Grammostola rosea</i>

SFI8	toxina I8 da aranha <i>Segestria florentina</i>
ATXII	toxina II da anémoma do mar <i>Anemonia sulcata</i>
HEL	" <i>Hen Egg Lysozyme</i> "
cDNA	ácido desoxiribonucléico complementar
Ag	antígeno
Ig	imunoglobulina
Ab	anticorpo
mAb	anticorpo monoclonal
DIV	domínio IV de um canal iônico
S1	segmento 1 de um canal iônico
ICK	" <i>Inhibitory Cystine Knot</i> "
DDH	" <i>Disulfide-Directed β-Hairpin</i> "
2-D	bidimensional
3-D	tridimensional
Å ²	Angström quadrado
kDa	quilo Dalton
LD50	dose letal 50
MOF	" <i>Modeller Objective Function</i> "
RMSD	" <i>Root Mean Square Deviation</i> "
C α	carbono álfa
C or Cys	cisteína
W or Trp	triptofano
R or Arg	arginina
Y or Tyr	tirosina
P or Pro	prolina
HPLC	" <i>High Performance Liquid Chromatography</i> "

NMR	espectrometria de Ressonância Magnética Nuclear
CD	Dicroísmo Circular
SQL	<i>"Structured Query Language"</i>
UML	<i>"Unified Modeling Language"</i>
BF	<i>"Biological Function"</i>
BA	<i>"Biological Activity"</i>
BEA	<i>"Basic Element of Action"</i>
FPS	<i>"Flanking Protein Sequence"</i>
NN	<i>"Nearest Neighbor "</i>
ONN	<i>"Optimized Nearest Neighbor path"</i>
CASP	<i>"Critical Assessment of techniques for protein Structure Prediction"</i>
CAFASP	<i>"Critical Assessment of Fully Automated Structure Prediction"</i>
PDB	<i>"Protein Data Bank"</i>

LISTA DAS FIGURAS

Figura 1. Os canais iônicos dependentes de voltagem	17
Figura 2. Representação em duas dimensões da organização transmembrana da subunidade alfa do canal de sódio dependente de voltagem.....	20
Figura 3. Representação do motivo ICK	25
Figura 4. Epitopos contínuos e descontínuos	36
Figura 5. Precisão e aplicações da modelagem por homologia.....	52
Figura 6. Diagrama de classes do banco SiliTox.....	57
Figura 7. Esquemática do algoritmo do SiliMod.....	69
Figura 8. Descrição do algoritmo do PEPOP.....	74
Figura 9. Descrição esquemática dos métodos de construção dos peptídeos empregados por PEPOP e SiliPop.....	77
Figura 10. Modelagem molecular de PnTx2-5 e Pn2x2-6	83
Figura 11. Resultados da avaliação das estruturas teóricas de PnTx2-5 e PnTx2-6.....	85
Figura 12. Testes experimentais do modelo de PnTx2-6 proposto	88
Figura 13. Alinhamento múltiplo das sequências das toxinas pertencendo aos grupos estruturais “Kozclass1403”, “Kozclass1502”, “8C_ICK2.10.2”, “6C_N3.x.x.5.4”	89
Figura 14. Árvore de Neighbor-Joining das sequências de toxinas aracnídicas membros do grupo de motivo cisteína 6C_Nx.3.x.4.1	91
Figura 15. Página principal do PEPOP.....	94
Figura 16. Saída da análise da toxina de aranha PnTx2-6 por PEPOP	95
Figura 17. Desenho de peptídeo antigênico por PEPOP	96
Figura 18. Resultado da análise por PEPOP das estruturas modeladas de PnTx2-5 e PnTx2-6	101
Figura 19. Organização dos dados na tabela <i>BioMolecule</i> do banco SiliTox	103
Figura 20. Organização dos dados nas tabelas (A) <i>Model</i> , (B) <i>Peptide</i> , e (C) <i>SegInPep</i> do banco SiliTox.	104
Figura 21. Páginas de pesquisa do banco SiliTox e de resultado de uma procura....	105
Figura 22. Entrada do SiliTox para a biomolécula PnTx2-6.....	107
Figura 23. Informações estruturais do SiliTox sobre a biomolécula SFI8	108
Figura 24. Informações do SiliTox sobre a estrutura determinada experimentalmente da biomolécula δ -paluIT2	109
Figura 26. Informações funcionais do SiliTox sobre a biomolécula PnTx2-6.....	112
Figura 27. Comparação das superfícies bioativas das neurotoxinas modificadoras de canal de sódio se ligando ao sítio 3 dos canais sódicos de mamíferos	117

LISTA DAS TABELAS

Tabela 1. Algumas toxinas de aranhas, escorpiões e caramujos marinhos com potencial terapêutico.....	14
Tabela 2. Sítios receptores de neurotoxinas nos canais de sódio dependentes de voltagem	21
Tabela 3. Resultados estatísticos da avaliação pelo diagrama de Ramachandran das estruturas dos <i>templates</i> e dos modelos	86
Tabela 4. Resultados da comparação entre os modelos teóricos produzidos pelo SiliMod e as estruturas experimentais	92
Tabela 5. Distribuição dos resíduos do epítipo HEL nos <i>clusters</i> identificados por PEPOP	98
Tabela 6. Avaliação do desempenho do PEPOP	99

ÍNDICE

RESUMO.....	1
ABSTRACT	2
LISTA DAS ABREVIACÕES.....	3
LISTA DAS FIGURAS	6
LISTA DAS TABELAS.....	7
INTRODUÇÃO	12
I. As toxinas animais.....	12
I.1. As toxinas animais na pesquisa científica	12
I.1.A. Um problema de saúde pública	12
I.1.B. Potencial terapeutico	13
I.1.C. Aplicações biotecnológicas	15
I.2. Aspectos funcionais.....	15
I.2.A. Canais iônicos	16
I.2.A.i. Função	16
I.2.A.ii. Estrutura	16
I.2.B. Mecanismos de interação toxinas/canais iônicos	18
I.2.B.i. Toxinas de escorpiões	18
I.2.B.ii. Toxinas de aranhas	19
I.3. Aspectos estruturais.....	22
I.4. Relação estrutura-atividade	24
I.4.A. Especificidade das toxinas animais	24
I.4.B. Superfícies bioativas das toxinas animais	26
I.5. Métodos experimentais.....	28
I.6. Recursos existentes sobre as toxinas animais	29
II. Interações proteína-proteína.....	30
II.1. Caracterização da interface proteína-proteína	31
II.1.A Características topológicas	31
II.1.B. Propriedades físico-químicas.....	32
II.1.B.i. Hidrofobicidade	32
II.1.B.ii. Composição em aminoácidos.....	33
II.2. Peculiaridades da interação anticorpo-antígeno.....	34
II.2.A. Sobre o anticorpo	34
II.2.B. Epitopo e paratopo.....	35
II.2.C. Antigenicidade e imunogenicidade.....	37
II.2.D. Reação cruzada e poliespecificidade.....	37
II.3 Métodos de estudo	38
II.3.A. Métodos experimentais.....	38

II.3.A.i. Interações proteína-proteína.....	38
II.3.A.ii. Identificação experimental do epítopo.....	39
II.3.B. Métodos preditivos	40
II.3.B.i. Interações proteína-proteína	40
II.3.B.ii. Interações Ac/Ag e localização do epítopo	41
III. Determinação da estrutura protéica	42
III.1. Métodos experimentais.....	42
III.1.A. Cristalografia de raios X	43
III.1.B. Espectroscopia RMN	43
III.1.C. Dicroísmo circular	43
III.1.D. Limitações e dificuldades	44
III.1.E. Progressos na área.....	44
III.2. Métodos preditivos.....	45
III.2.A. Predição <i>ab initio</i>	45
III.2.B. <i>Fold recognition</i>	46
III.2.C. Modelagem comparativa.....	47
III.2.C.i. Comparação sequência-sequência	47
III.2.C.ii. Comparação perfil-sequência e sequência-perfil.....	48
III.2.C.iii. Comparação perfil-perfil.....	48
III.2.D. Ferramentas de modelagem.....	49
III.2.E. Métodos de avaliação da modelagem	49
III.2.F. Avaliação do modelo	51
III.3. Coleções e classificações de estruturas 3D	51
OBJETIVO	55
ESTRATÉGIA	55
OBJETIVOS ESPECÍFICOS	55
MATERIAL & MÉTODOS	56
I. Banco de dados : SiliTox	56
I.1. Dados	56
I.2. Estrutura do banco de dados	58
I.2.A. Classes ' <i>BioEntity</i> '	58
I.2.B. Classes BioΨ	58
I.2.C. Classes ' <i>RefBiblio</i> ', ' <i>Species</i> ', ' <i>Origin</i> ', ' <i>NaturalEffect</i> ' e ' <i>Symptom</i> '	59
I.2.D. Classe ' <i>Application</i> '	59
I.2.E. Classes ' <i>Disulfide</i> ' e ' <i>CysPattern</i> '	59
I.2.F. Classes ' <i>TempGroup</i> ' e ' <i>Model</i> '	61
I.2.G. Classes ' <i>Segment</i> ', ' <i>Distance</i> ', ' <i>Cluster</i> ' e ' <i>Peptide</i> '	61
I.3. Implementação do banco de dados.....	62
I.3.A. Hierarquia das classes	62
I.3.B. Relações múltiplas	62
I.3.C. Tabelas e atributos.....	63

II. Modelagem das estruturas 3D das toxinas : SiliMod	63
II.1. As toxinas Tx2-5 e Tx2-6 da aranha <i>Phoneutria nigriventer</i> como caso de estudo	63
II.1.A. Seleção de <i>templates</i>	63
II.1.A.i. Baseada na similaridade de sequências	63
II.1.A.ii. Baseada no reconhecimento estrutural (<i>fold recognition</i>).....	64
II.1.A.iii. Baseada na composição de aminoácidos	64
II.1.A.iv. Baseada na similaridade de padrões de cisteínas	64
II.1.A.v. Seleção final.....	65
II.1.B. Alinhamento das estruturas <i>templates</i>	65
II.1.C. Alinhamento das sequências alvo- <i>template</i> baseado no alinhamento das estruturas dos <i>templates</i>	65
II.1.D. Modelagem e restrição das pontes dissulfeto	65
II.1.E. Avaliação dos modelos e seleção	66
II.1.F. Fluorescência do triptofano e dicroísmo circular	66
II.2. Constituição dos grupos estruturais baseados no padrão de cisteínas para a modelagem automática das toxinas	67
II.2.A. <i>Templates</i> disponíveis	67
II.2.B. Composição dos grupos	67
II.3. O algoritmo do SiliMod.....	68
II.3.A. Conjunto de alvos	68
II.3.B. Seleção de <i>templates</i>	68
II.3.C. Alinhamento estrutural dos <i>templates</i> (optativo)	70
II.3.D. Alinhamento das sequências <i>template</i> (s)-alvo baseado na(s) estrutura(s).....	70
II.3.E. Obrigação das pontes dissulfeto e construção dos modelos	70
II.3.F. Seleção do melhor modelo	71
II.4. Validação do método.....	71
III. Ferramentas de análise da superfície antigênica das proteínas: PEPOP & SiliPop.....	72
III.1. Algoritmo do PEPOP	72
III.1.A. Agrupamento (<i>Clustering</i>) dos segmentos acessíveis na superfície	72
III.1.B. Notação dos segmentos.....	73
III.1.C. Notação dos <i>clusters</i>	75
III.1.D. Classificação dos segmentos e dos <i>clusters</i>	75
III.1.E. Desenho dos peptídeos antigênicos e/ou imunogênicos	75
III.2. O algoritmo do SiliPop.....	78
III.3. Avaliação do desempenho da ferramenta.....	78
III.4. Análise da superfície antigênica das toxinas	79
IV. Implementação das ferramentas	80
RESULTADOS.....	82
I. Modelagem da estrutura 3D das toxinas	82
I.1. PnTx2-5 e PnTx2-6.....	82
I.2. Estruturas teóricas de PnTx2-5 e PnTx2-6.....	82
I.2.A. Descrição das estruturas	84
I.2.B. Avaliação dos modelos	84
I.3. Grupos estruturais.....	87
I.4. Resultados SiliMod.....	90

I.4.A. Modelos tridimensionais.....	90
I.4.B. Validação do método	93
II. Análise da superfície antigênica das toxinas : <i>PEPOP</i> & <i>SiliPop</i>	93
II.1. PEPOP	93
II.1.A. Saída do programa PEPOP.....	93
II.1.B Desempenho do PEPOP sobre epitopos conhecidos.....	97
II.2. Predição de peptídeos antigênicos nas toxinas PnTx2-5 e PnTx2-6.....	100
II.3. SiliPop.....	102
III. Banco de dados de toxinas : <i>SiliTox</i>	102
III.1. Estocagem dos dados	102
III.2. Recuperação dos dados.....	102
DISCUSSÃO	113
CONCLUSÃO & PERSPECTIVAS	119
REFERÊNCIAS BIBLIOGRÁFICAS	122
APÊNDICES	151
APÊNDICE 1: Dicionário de dados do <i>SiliTox</i>	152
APÊNDICE 2: <i>Scripts</i> adaptados do Modeller.....	172
APÊNDICE 3: Grupos estruturais para <i>SiliMod</i> - alinhamentos	174
APÊNDICE 4: Grupos estruturais para <i>SiliMod</i> - árvores NJ.....	188

INTRODUÇÃO

I. As toxinas animais

I.1. As toxinas animais na pesquisa científica

As toxinas animais são de grande interesse em biologia. Além do fato que elas começaram a serem estudadas por causa do seu envolvimento nos acidentes de envenenamento, elas são usadas como sondas moleculares para identificar ou modificar processos biológicos chave. Estes estudos revelam novos potenciais terapêuticos, e um interesse crescente para suas propriedades estruturais foi também observado na área do desenho de fármacos nos últimos 15 anos.

I.1.A. Um problema de saúde pública

Os animais venenosos constituem uma causa significativa de mortalidade mundial, especialmente nos continentes tropicais e subtropicais. As picadas de serpentes são responsáveis por uma mortalidade significativa em várias partes da África, Ásia e nas Américas. Populações da Austrália, da Nova Guiné e de alguns países europeus também são vulneráveis às ataques de animais peçonhentos. Dentro desses últimos, o envenenamento por aracnídeos tem um impacto significativo na saúde pública, especialmente as picadas de escorpião que podem causar um severo envenenamento sistêmico. Conhece-se por volta de 50 espécies perigosas pertencendo à família Buthidae. As picadas de aranha também representam um problema de saúde pública e a maioria dos casos medicamente importantes pode ser atribuída às aranhas viúva negra (*Latrodectus* spp.), reclusa (*Loxosceles* spp.) e armadeira (*Phoneutria* spp.). Acidentes causados por animais marinhos (caramujos e anêmona) são comuns, mas os efeitos menos severos. O envenenamento causa uma série de efeitos clínicos locais, notadamente a necrose dos tecidos; e efeitos sistêmicos, incluindo paralisia, distúrbio hemostático, choque, aumento da permeabilidade capilar, dano miocárdico, rhabdomiólise e falha renal aguda (Warrell 1993). O soro hiperimune específico (antiveneno) é o suporte principal do tratamento médico dos

envenenamentos severos, porém não existem soros para todos os organismos perigosos. Tratamentos auxiliares como a ventilação assistida, restauração do volume circulante, diálise renal e limpeza cirúrgica dos tecidos necróticos são necessários em alguns casos. Drogas vasodilatadoras são importantes para contrapor-se aos efeitos da liberação massiva de catecolamina (White 2000). Infelizmente, os antivenenos não são um produto economicamente viável, assim o desenvolvimento e a manufatura desses agentes são limitados. Há necessidade de melhorar a prevenção e a gestão dessas condições, e de desenvolver novos suprimentos de antivenenos (White et al. 2003).

I.1.B. Potencial terapeutico

O efeito dessas toxinas em humanos não é limitado ao envenenamento, pois elas mostraram um potencial inestimável como ferramentas em pesquisa e agentes para diagnóstico, e podem até ter um futuro como precursores de agentes terapêuticos (White 2000; Lewis and Garcia 2003) (Tabela 1) (Bosmans et al., in press). 1.5 Bilhão de pessoas (Jacobs 2005) sentindo dor crônica moderada até severa, usam drogas anti-inflamatórias não esteroideais, opióides, anticonvulsivantes e até antidepressivos. Como muitos desses analgésicos são somente moderadamente eficientes contra certos tipos de dor crônica e causam frequentemente efeitos colaterais sérios, a descoberta de novas abordagens terapêuticas é um objetivo importante. O fato que as toxinas animais atuam em moléculas implicadas na nocicepção oferece perspectivas interessantes para a descoberta de novos analgésicos e de novas drogas a serem usadas como neuroprotetores e anticonvulsivantes (Bosmans et al., in press). Porém, muitos desafios permanecem entre a fase inicial de descoberta de drogas e o uso das toxinas num contexto clínico ou até mesmo sua validação como verdadeiras drogas (Lewis and Garcia 2003). Apesar da existência de exemplos em ambos os casos, ainda são poucos até agora aprovados como medicamentos.

Tabela 1. Algumas toxinas de aranhas, escorpiões e caramujos marinhos com potencial terapêutico. (Adaptado de Billen et al., 2008)

	Organism	Toxin	Uniprot Accession	Potential Therapeutic Use or Lead
Spiders	<i>Chilobrachys jingzhao</i>	JZTX I	P83974	cardiovascular pathologies
		JZTX III	P62520	
	<i>Ceratogyrus cornuatus</i>	CcoTx1	P84507	analgesia, neuronal disorders
		CcoTx2	P84508	
		CcoTx3	P84509	cardiovascular pathologies
	<i>Thrixopelma pruriens</i>	ProTx-I	P83480	inflammatory pain
		ProTx-II	P83476	
<i>Phrixotrichus auratus</i>	PaurTx3	P84510	neuronal disorders, pain	
Scorpions	<i>Centruroides limpidus limpidus</i>	CII9	Q8WRY4	epilepsy
	<i>Buthus martensii</i> Karsch	BmK AS	Q9UAC9	analgesia
	<i>Odonthobuthus doriae</i>	OD1	P84646	inflammatory pain
Cone snails	<i>Conus striatus</i>	μ -SIIIA	Q86DU6	neuropathic pain
	<i>Comus kinoshitai</i>	μ -KIIIA	P0C195	neuropathic pain
	<i>Conus marmoreus</i>	μ O-MrVIB	Q26443	analgesia, skeletal muscle diseases
	<i>Conus ermineus</i>	δ -EVIA	P60513	neurological pathologies, multiple sclerosis

I.1.C. Aplicações biotecnológicas

Por causa do seu alto potencial contra insetos, algumas toxinas são consideradas modelos potenciais como biopesticidas e estudos extensivos estão sendo realizados nessa área (de Figueiredo et al. 2001; Wang et al. 2001). Ainda assim, alguns critérios precisam ser respeitados para que toxinas animais possam ter um papel no mercado dos pesticidas (Whetstone and Hammock 2007), especialmente sua baixa toxicidade em organismos não alvo precisa ser confirmada.

Por outro lado, a estrutura compacta natural das toxinas animais representa um esqueleto atraente para a engenharia de novos fármacos (Menez 1998; Gilquin et al. 2002; Blumenthal and Seibert 2003) e a possibilidade de enxertar o sítio funcional de enzimas na estrutura principal dessas mini-proteínas para criar novas moléculas ativas já foi demonstrada (Vita et al. 1995; Mer et al. 1998; Vita et al. 1999; Mourier et al. 2000; Ricciardi et al. 2000).

I.2. Aspectos funcionais

A maioria dos venenos contém uma grande variedade de peptídeos, frequentemente ativos biologicamente. Os peptídeos encontrados nos venenos de aracnídeos, principalmente escorpiões e aranhas, estão sendo estudados extensivamente. Esses peptídeos apresentam uma diversidade significativa nas suas estruturas primárias e ligam-se a um grande número de alvos farmacológicos.

Foi mostrado que as toxinas peptídicas animais atuam em canais de sódio (Na_v) (Cestele and Catterall 2000; Catterall et al. 2007), potássio (K_v) (Coetzee et al. 1999; Gutman et al. 2005) e cálcio (Ca_v) (Chuang et al. 1998; Olamendi-Portugal et al. 2002) voltagem-dependentes, assim como em canais de cloreto (Lippens et al. 1995), canais iônicos sensíveis ao ácido (ASICs) (Diochot et al. 2007; Jasti et al. 2007), canais mecano-sensíveis (MSCs) (Lesage and Lazdunski 2000; Ostrow et al. 2003), receptores nicotínicos a acetilcolina (nAChR) e mais recentemente em receptores de capsaicina (Siemens et al. 2006). A maioria das toxinas aracnídicas afetam canais iônicos.

I.2.A. Canais iônicos

I.2.A.i. Função

Os canais iônicos são expressos diferencialmente e de jeito ubíquo na superfície das células excitáveis e não excitáveis, onde eles atuam num grande número de processos fisiológicos. A contração muscular, a percepção do meio ambiente, o processamento da informação no cérebro, a conexão do cérebro aos tecidos periféricos, o metabolismo intracelular, a transdução do sinal, a expressão gênica, a síntese e a degradação das proteínas são representativos dos numerosos contextos envolvendo sinais elétricos conduzidos por membros da superfamília protéica dos canais iônicos (Catterall et al. 2007). Sua disfunção leva a uma atividade anormal num ou mais dos sistemas fisiológicos, o que pode provocar efeitos dramáticos.

I.2.A.ii. Estrutura

A superfamília dos canais iônicos contém mais de 140 proteínas formadoras de poro na membrana estruturalmente relacionadas (Yu and Catterall 2004). Os canais iônicos dependentes de voltagem (Na_v , Ca_v and K_v) são construídos de um motivo estrutural comum contendo seis segmentos trans-membrana (S1-S6) e uma alça de poro, constituindo um domínio D. Os poros dos canais são formados pelos segmentos S5/S6 e a alça de poro entre eles, e retidos pelo dobramento dos segmentos S6 num resíduo glicina ou prolina (Cestele and Catterall 2000; Catterall et al. 2007). O domínio do sensor de voltagem é constituído pelos segmentos S1-S4, os resíduos carregados positivamente do segmento S4 servindo de cargas denominadas de “gating”. A subunidade principal dos canais Na_v e Ca_v (chamada respectivamente subunidade α ou subunidade α_1) é composta de quatro domínios (DI até DIV) e é suficiente para a expressão funcional. A subunidade α do Na_v é associada a até 4 subunidades β (β_1 , β_2 , β_3 e β_4) auxiliares e homólogas, enquanto a subunidade α_1 do Ca_v é assistida por subunidades α_2 , β , γ e δ distintas. Essas subunidades auxiliares modificam a cinética e a dependência de voltagem do *gating* do canal. Os canais K_v são compostos por 4 subunidades α , cada uma sendo análoga de um único domínio das principais subunidades de Na_v e Ca_v e podendo juntar-se em homo- e heterotetrámeros, levando a uma grande diversidade de complexos de canais diferentes. Os

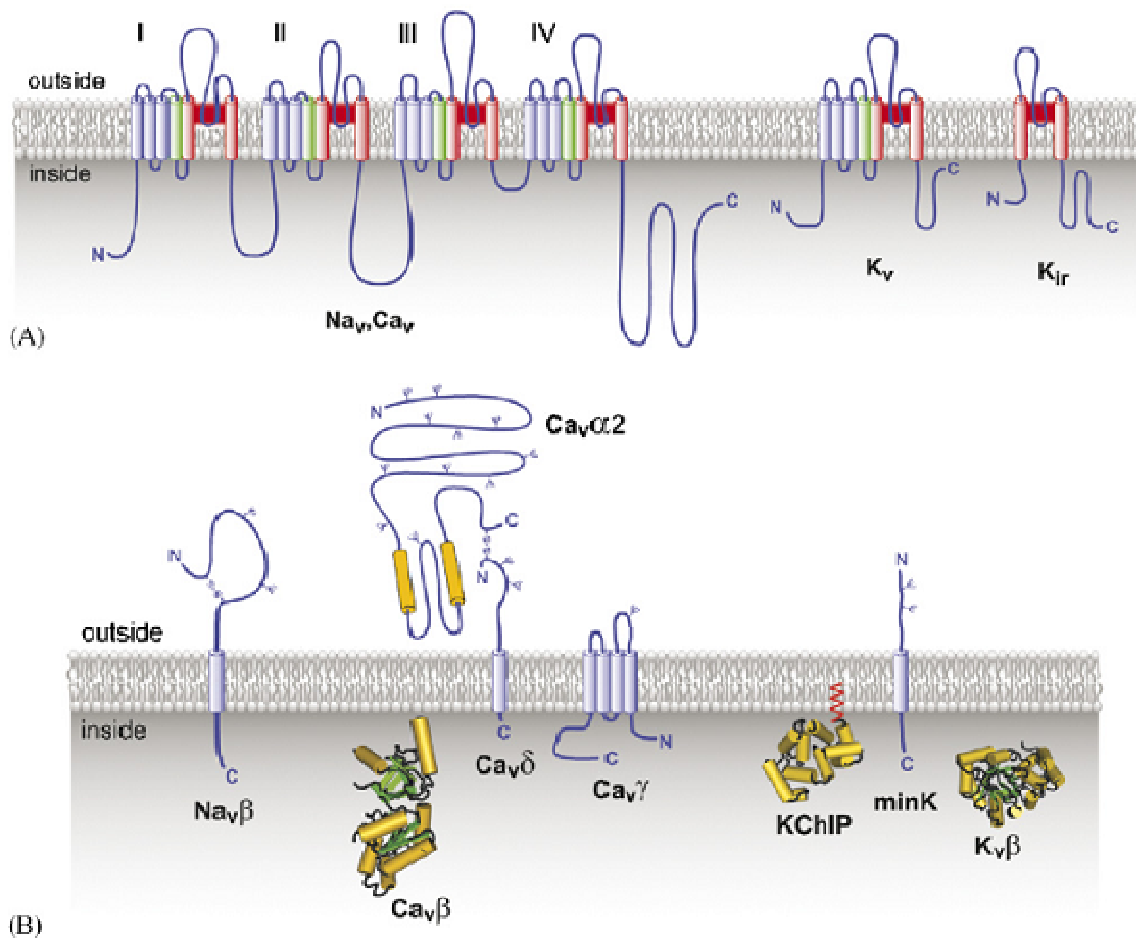


Figura 1. Os canais iônicos dependentes de voltagem : (A) Os diferentes membros da família proteica dos canais iônicos estruturalmente relacionados aos canais iônicos dependentes de voltagem são ilustrados como diagramas nos quais cilindros representam prováveis alfa-hélices transmembranárias. Em vermelho, os segmentos S5-S6 formando um poro; em verde, o sensor de voltagem S4; e em azul, os segmentos transmembranários S1-S3. (B) Subunidades auxiliares dos canais iônicos dependentes de voltagem. As subunidades auxiliares dos canais Nav, Cav, e Kv são ilustradas, com cilindros representando as hélices alfa preditas das subunidades transmembranas. As subunidades auxiliares intracelulares são ilustradas por suas estruturas tridimensionais preditas. (De Catterall et al., 2007)

canais K_v podem ser associados a entre uma e três subunidades $K_v\beta$ intracelulares ($K_v\beta 1-3$) (Catterall et al. 2007) (Figura 1).

I.2.B. Mecanismos de interação toxinas/canais iônicos

As toxinas atuando nos canais ativadas por voltagem operam de dois jeitos: (i) bloqueando o poro de condução dos íons ("*pore-blocking toxins*"), ou (ii) ligando-se numa região do canal envolvida no mecanismo de gating ("*gating-modifier toxins*"). A maioria das toxinas de escorpiões e aranhas afeta os canais Na_v , Ca_v or K_v do segundo jeito (Swartz and MacKinnon 1995; Li-Smerin and Swartz 1998; Bourinet et al. 2001; Middleton et al. 2002; Corzo et al. 2005). Os canais Na_v ilustram a diversidade de ação das toxinas e são os mais estudados, por serem o alvo molecular de vários grupos de toxinas que se ligam em pelo menos seis sítios diferentes nestes canais (Figura 2 e Tabela 2) (Catterall 1980; Cestele and Catterall 2000; Catterall et al. 2007). As interações toxina-canal são caracterizadas por uma alta afinidade e uma alta especificidade.

I.2.B.i. Toxinas de escorpiões

Toxinas peptídicas de escorpiões foram classificadas segundo sua especificidade por mamíferos ou insetos, o tipo de sintomas induzidos, seus mecanismos farmacológicos e suas características estruturais. Assim, as toxinas de escorpiões agindo nos canais Na_v estão divididas em dois grupos maiores : α -toxinas e β -toxinas. As α -toxinas de escorpiões ligam-se no sítio 3 dos canais Na_v (Figura 2) e reduzem a velocidade da inativação rápida da corrente de Na^+ ou a inibem (Bosmans and Tytgat 2007). Essas toxinas são depois divididas em três tipos, segundo sua seletividade farmacológica pelas suas presas : α -toxinas clássicas atuam preferencialmente em mamíferos; α -toxinas anti-insetos atuam preferencialmente em insetos; toxinas ' *α -like*' atuam em ambos mamíferos e insetos (mas são incapazes de se ligar nos sinaptosomas de cérebro de rato) (Gordon et al. 1996; Bosmans and Tytgat 2007; Gordon et al. 2007; Karbat et al. 2007). Para cada tipo de toxina, toxinas modelos são consideradas como referências e usadas em experimentos de competição. As α -toxinas clássicas incluem AaH II do escorpião *Androctonus australis*

Hector e Lqh2 de *Leirus quinquestriatus hebraeus* (Rochat et al. 1972; Chen et al. 2002; Legros et al. 2005). Lqh α IT de *L. quinquestriatus hebraeus* é a α -toxina anti-insetos mais bem caracterizada até agora (Vargas et al. 1987; Eitan et al. 1990; Gordon et al. 2007). Toxinas ' α -like' típicas são representadas por BmK M1 de *Buthus martensi Karsch* e Lqh3 de *L. quinquestriatus hebraeus* (Sautiere et al. 1998; Li and Ji 2000; Goudet et al. 2001; Brone et al. 2003; Bosmans et al. 2005).

A respeito das β -toxinas, elas se ligam no sítio 4 dos canais Na_v e causam um deslocamento para valores negativos na dependência de voltagem da ativação do canal (Cestele and Catterall 2000). Elas estão também divididas em β -toxinas seletivas para mamíferos, β -toxinas seletivas para insetos e toxinas ' β -like' (Bosmans et al. 2007; Gurevitz et al. 2007). Css2 e Css4 de *Centruroides suffusus suffusus* são duas β -toxinas seletivas para mamíferos (Martin et al. 1987), enquanto Ts VII de *Tityus serrulatus* (Barhanin et al. 1982; Bechis et al. 1984; Martin and Rochat 1986; Ceard et al. 1992; Becerril et al. 1997) e Lqh β 1 de *L. quinquestriatus hebraeus* atuam tipicamente em canais Na_v de ambos mamíferos e insetos. Dentro das β -toxinas seletivas para insetos, uma outra dicotomia é observada entre os sintomas induzidos na larva de mosca ("*blow fly*" da família Calliphoridae) : aquelas que provocam uma contração rápida e reversível são chamadas toxinas excitatórias ("*excitatory toxins*") e aquelas que provocam uma contração lenta e transitória seguida por uma curta fase de paralisia flácida são chamadas toxinas depressantes ("*depressant toxins*"). O primeiro grupo inclui AaHIT de *A. australis* Hector e Bj-xtrIT de *Buthotus judaicus* (Froy et al. 1999; Shichor et al. 2002; Karbat et al. 2004), enquanto o segundo grupo é representado por LqqIT2 de *L. quinquestriatus quinquestriatus* e BjIT2 de *B. judaicus* (Zlotkin et al. 1993; Bosmans et al. 2005; Karbat et al. 2007).

I.2.B.ii. Toxinas de aranhas

A maioria das toxinas peptídicas isoladas até agora agem nos processos de sinalização pela modulação dos canais iônicos. Ao contrário das toxinas de escorpião, não há nenhuma classificação funcional das toxinas de aranhas. Várias toxinas peptídicas de aranhas atuando em canais Na_v que foram identificadas ligam-se em diferentes sítios do receptor (Billen et al. 2008). As huwentoxinas IV e hainantoxinas

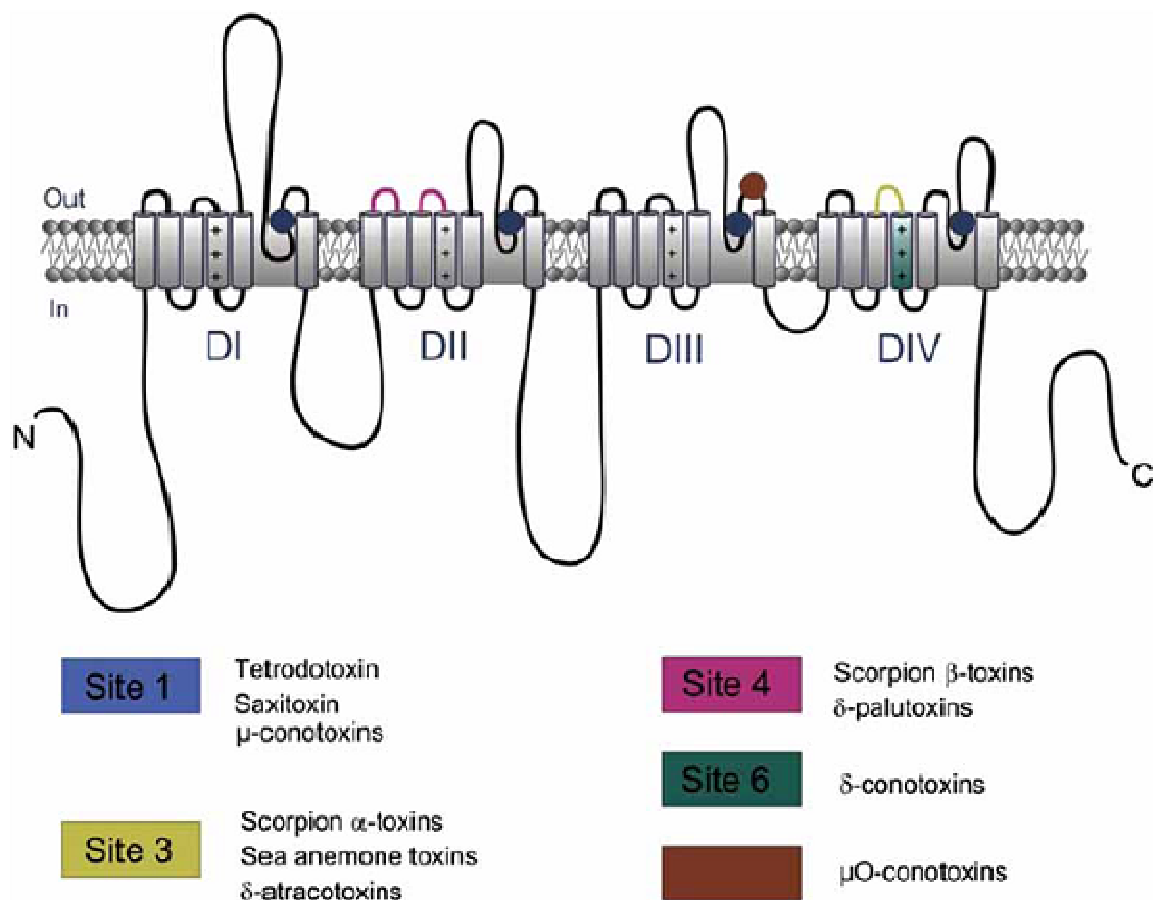


Figura 2. Representação em duas dimensões da organização transmembrana da subunidade alfa do canal de sódio dependente de voltagem. Localização dos sítios receptores de neurotoxinas nos canais de sódio de mamíferos são ilustradas por cores, como indicado. Os sítios receptores onde toxinas polipeptídicas são indicadas em cores. (Adaptado de Billen et al., 2008)

Tabela 2. Sítios receptores de neurotoxinas nos canais de sódio dependentes de voltagem. (De Catterall et al., 2007)

Receptor site	Neurotoxin	Functional effect
Site 1	Tetrodotoxin Saxitoxin μ -Conotoxin	Pore block
Site 2	Batrachotoxin Veratridine Grayanotoxin Aconitine	Persistent activation enhanced activation, and block of activation
Site 3	α -Scorpion toxins Sea anemone toxins Atrachotoxins	Slowed inactivation
Site 4	β -Scorpion toxins	Enhanced activation
Site 5	Brevetoxins Ciguatoxin	Enhanced activation and block of inactivation
Site 6	δ -Conotoxin	Slowed inactivation

III-V de *Selenocosmia huwena* (Chinese bird spider) inibem a conductância do Na⁺ por ligarem-se supostamente ao sítio 1 (Xiao and Liang 2003; Liang 2004), enquanto as δ -atractoxinas das aranhas australianas denominadas “Australian funnel-web” e “eastern mouse”, as toxinas das famílias Tx2 e Tx4(6-1) de *Phoneutria nigrioventer*, Magi 4 e jingzhaotoxina I e III, todas atrasam a inativação do canal Na_v pelo fato de ligarem-se ao sítio 3 do receptor (Matavel et al. 2002; Corzo et al. 2003; Xiao et al. 2004; Xiao et al. 2005). Outras toxinas como as δ -palutoxinas, Magi 5 e as μ -agatoxinas acredita-se que se ligam no sítio 4 uma vez que elas atuam como modificadoras de *gating* e mudam a curva de inativação para potenciais mais negativos (Corzo et al. 2000; Adams 2004; Corzo et al. 2007). Toxinas isoladas de venenos de aranhas tarantulas, incluindo heteropodatoxinas, stromatoxinas, heteroscodratoxinas e hanatoxinas mostraram que são capazes de interagir com canais K_v (Escoubas and Rash 2004). Embora muitas toxinas de aranhas parecem ser seletivas para um único tipo de canal iônico, a capacidade de atuar em mais de um tipo de canal iônico tem sido demonstrada em várias toxinas. Por exemplo, ProTx-I e II da tarantula *Thrixopelma pruriens* inibe a ativação dos canais Na_v, Ca_v and K_v (Middleton et al. 2002), a ω -grammotoxina-SIA apresenta uma atividade em ambos os canais Ca_v e K_v (Li-Smerin and Swartz 1998) e as ω -atractoxinas e ω -agatoxinas atuam nos canais Na_v e Ca_v (Wicher and Penzlin 1998; Escoubas and Rash 2004; Chong et al. 2007).

I.3. Aspectos estruturais

As toxinas peptídicas de escorpiões são ou peptídeos de cadeia comprida contendo entre 58 e 76 resíduos de aminoácidos (atuando em maioria nos canais de Na⁺), ou peptídeos de cadeia curta contendo entre 28 e 46 resíduos (atuando em maioria nos canais de K⁺ ou de Cl⁻) (Possani et al. 1999; Possani et al. 2000; Rodriguez de la Vega and Possani 2004; Rodriguez de la Vega and Possani 2005; de la Vega and Possani 2007). A maioria das toxinas extraídas de venenos de escorpiões adotam o mesmo enovelamento (chamado ‘ α/β scorpion fold’) caracterizado pela presença de dois tipos de estruturas secundárias canônicas (topologia $\alpha\beta\beta$, $\beta\alpha\beta\beta$ ou $\beta\alpha\alpha\beta\beta\alpha$) e o papel central de 3 até 5 pontes dissulfeto que conferem uma estrutura muito compacta ao

peptídeo. Um outro *fold* menos frequente, contendo somente duas hélices α (topologia $\alpha\alpha$) estabilizadas por duas pontes dissulfeto num motivo chamado '*helical hairpin-like*', é representado pela κ -hefutoxina (Srinivasan et al. 2002). Comparações entre sequências de toxinas conhecidas dentro de cada família estrutural mostra uma diversidade significativa de sequências. Além das cisteínas, somente poucos resíduos são conservados (Bontems et al. 1991; Bontems et al. 1991; Bontems et al. 1992; Sollod et al. 2005), indicando uma alta tolerância às mutações residuais nas regiões inter-cisteínas.

A respeito das toxinas de aranhas, a diversidade de sequências é ainda mais notável. Elas são produzidas na forma de uma biblioteca de peptídeos combinatoriais limitados estruturalmente nos quais mutações ocorrem em todos os resíduos das toxinas maduras, com exceção de um punhado de cisteínas rigorosamente conservadas (Sollod et al. 2005). Kozlov e Grishin sugeriram uma classificação das sequências de toxinas peptídicas de aranhas segundo a distribuição conservada dos resíduos de cisteína (Kozlov 2005). De fato, embora alguns peptídeos destituídos de resíduos de cisteína tenham sido descritos (Pimenta and De Lima 2005), a maioria das toxinas de aranhas isoladas até agora são pequenos peptídeos ricos em cisteínas. Essas toxinas contêm entre 28 e 78 resíduos com 3 a 7 pontes dissulfeto. Dois motivos estruturais diferentes foram identificados, correspondendo a conectividades distintas de cisteínas e características estruturais. O motivo mais comum é chamado '*Inhibitory Cysteine Knot*' (ICK) e caracteriza-se por uma folha β anti-paralela tripla fita, estabilizada por um nó de 3 pontes dissulfeto alcançado quando uma ponte atravessa o macrociclo formado pelas outras duas pontes e as porções de cadeia principal interconectadas (Le Nguyen et al. 1990; Narasimhan et al. 1994; Pallaghy et al. 1994; Norton and Pallaghy 1998; Craik et al. 2001; Gelly et al. 2004) (Figura 3). Esse motivo aparece como um dos menores domínios globulares estáveis encontrados em proteínas (Pallaghy et al. 1994) e as proteínas contendo tal nó são chamadas '*knottin*' (Le Nguyen et al. 1990). Aproximadamente 350 toxinas de aranhas e quase 50 toxinas de escorpiões são *knottins*. Em termos de sequência, todas as *knottins* satisfazem o consensus seguinte : '*CX3-7CX3-6CX0-5CX1-4CX4-13C*', onde 'X' pode ser qualquer resíduo de aminoácido. O outro *fold* encontrado em toxinas peptídicas de aranha é

baseado na estrutura (*scaffold*) chamada 'Disulfide-Directed β -Hairpin' (DDH) (Wang et al. 2000).

I.4. Relação estrutura-atividade

I.4.A. Especificidade das toxinas animais

Foi mostrado que toxinas peptídicas são capazes de discriminar os canais iônicos de insetos dos de mamíferos, e até mesmo de distinguir canais iônicos de subtipos celulares diferentes (Zlotkin et al. 1993). Essa preferência diferencial foi ressaltada no caso bem estudado das toxinas de escorpiões que têm como alvo diferentes tipos de canais Na_v (Gordon et al. 2007). Embora a subunidade α do canal Na_v seja bem conservada em insetos e mamíferos na sua organização topológica assim como nas suas propriedades bioquímicas e farmacológicas (Loughney et al. 1989; Cestele et al. 1995), algumas toxinas de cadeia comprida de escorpiões mostram uma especificidade para canais de sódio de insetos (Zlotkin et al. 1993; Gurevitz et al. 2001) e várias α -toxinas são até capazes de discriminar subtipos de canais Na_v de mamíferos (Gilles et al. 1999; Maertens et al. 2006; Gordon et al. 2007). Foi sugerido que o forte potencial inseticida da Lqh α IT deve ser associado à conformação que se projeta do domínio NC e que pode ser observada em todas as α -toxinas de escorpião ativas em insetos (Lqh3, BmK M1, BmK M2, BmK M4, (Guan et al. 2004). Ao contrário, o domínio NC das α -toxinas de escorpião ativas em mamíferos (AaH II, BmK M8) apresenta uma geometria achatada (Fontecilla-Camps et al. 1988; Li et al. 1996; He et al. 1999).

No que diz respeito à preferência por alguns subtipos de canais Na_v de mamíferos, diferenças nas sequências dos canais foram investigadas e foi mostrado que a sequência da junção D4/S3-S4 em diferentes canais de sódio varia consideravelmente comparada com outros segmentos trans-membrana (Goldin 1999). A substituição de um resíduo ácido encontrado em posições homólogas dessa alça externa (Glu1613 em $\text{Na}_v1.2$, Asp1428 em $\text{Na}_v1.4$ e Glu1613 em $\text{Na}_v1.7$) afeta a ligação da toxina ATXII da anêmona do mar *Anemonia sulcata* e das α -toxinas de escorpiões Lqq5 e Lqh2 anti-mamíferos (Rogers et al. 1996; Benzinger et al. 1998; Cestele and Catterall 2000;

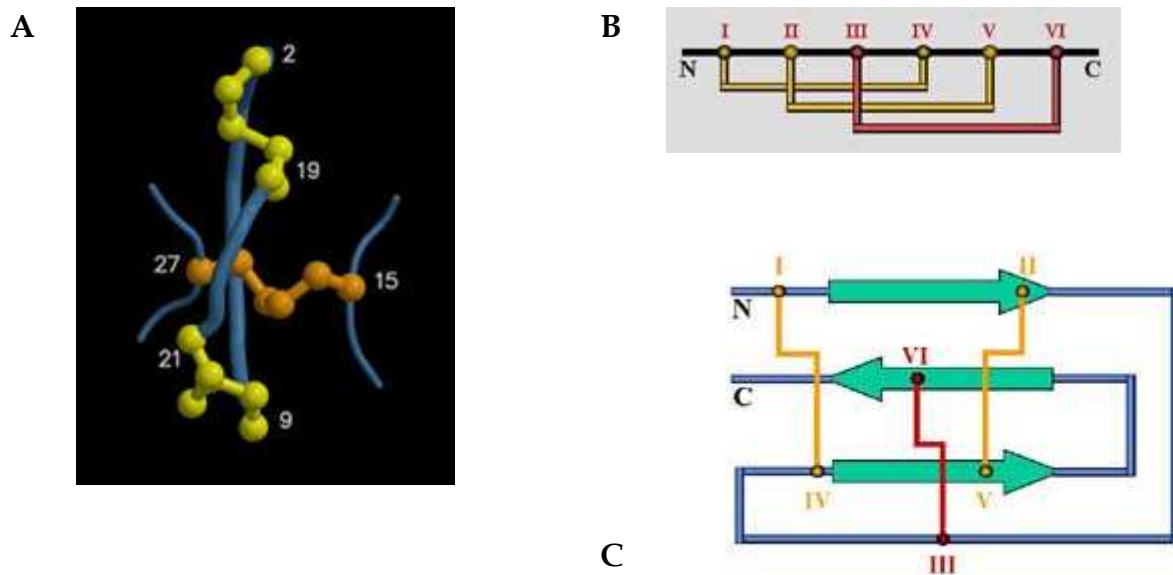


Figura 3. Representação do motivo ICK. **(A)** Vista tridimensional do nó formado pela ponte disulfeto (laranja) que cruza o macrociclo formado pelas duas outras (amarelo). **(B)** Convenção de anotação sequencial dos resíduos de cisteína envolvidos num ICK. Os resíduos respeitam o pareamento seguinte : CI - CIV; CII - CV; e CIII - CVI. **(C)** Representação esquemática do ICK incluindo três folhas beta. (*Adaptado de The Knottin Website <http://knottin.cbs.cnrs.fr/>, Gelly et al., 2004*)

Leipold et al. 2004), e foi sugerido que esse resíduo faça parte do sítio 3 do receptor. O resíduo equivalente Glu277 do canal K_v de drosófila também é considerado crítico para a ligação das toxinas hanatoxina 1 (HaTx1) de *Grammostola spatulata* e ω -grammotoxina SIA (GrTx) de *Grammostola rosea* (Takahashi et al. 2000). Além disso, foi sugerido que o sítio 3 do canal Na_v seja de natureza hidrofóbica (Fontecilla-Camps et al. 1982), o que foi confirmado pelos estudos de mutagênese sítio-dirigida realizadas no canal Nav1.4 de rato por Leipold e colaboradores. Eles mostraram que os resíduos hidrofóbicos Tyr1433, Phe1434 e Val1435 na alça S3-S4 do domínio IV são importantes para a ligação das α -toxinas de escorpiões (Leipold et al. 2004).

I.4.B. Superfícies bioativas das toxinas animais

Experimentos de ligação competitiva mostram que vários peptídeos de estruturas diferentes, encontrados em animais venenosos de phyla diferentes, compartilham sítios comuns ou sobrepondo-se num mesmo canal. Essa observação sugeriu que essas toxinas compartilham características estruturais nas suas superfícies de ligação. A técnica da mutagênese sítio-dirigida permitiu identificar alguns resíduos importantes para as respectivas ligações, e a comparação das superfícies de diferentes toxinas revelou características comuns das chamadas “superfícies bioativas”. Um exemplo bem caracterizado é a alça formada pelos resíduos 37 até 44 da AaH II, que revelou-se importante para a ligação da toxina (Fontecilla-Camps et al. 1988; Kharrat et al. 1989) e que flanqueia uma superfície hidrofóbica conservada dentro das α -toxinas de escorpiões. Segundo os autores, essa “superfície hidrofóbica conservada” (Fontecilla-Camps et al. 1982; el Ayeb et al. 1986) é composta pelos resíduos Tyr5, Tyr35, Trp38 e Tyr/Trp47 (Fontecilla-Camps et al. 1988; Kharrat et al. 1989), embora o papel de Tyr5 e Tyr47 pode ser mais estrutural do que puramente farmacológico (Housset et al. 1994; Li et al. 1996; He et al. 1999; Sun et al. 2003). Os papéis cruciais dos resíduos positivamente carregados Lys2, Lys28 e Lys58 foram também demonstrados (Darbon et al. 1983; Fontecilla-Camps et al. 1988; Kharrat et al. 1989; Kharrat et al. 1990; Loret et al. 1994; Takahashi et al. 2000). O mais recente trabalho de Legros e colaboradores confirmou a importância da Lys58 mas concluiu que o resíduo atua num papel estrutural, mais que participa na interação da toxina

com o canal (Legros et al. 2005). Resíduos funcionalmente e/ou estruturalmente essenciais foram também identificados em BmK M1 (Sun et al. 2003; Wang et al. 2003; Liu et al. 2005). Wang e colaboradores (Sun et al. 2003; Wang et al. 2003; Liu et al. 2005) sugeriram que a volta reversa dos aminoácidos 8 até 12 junta com o resíduo básico Lys62 é crítica para as bioatividades, e poderia interagir diretamente com o sítio receptor no canal de sódio. Eles propuseram que a carga positiva do resíduo possa também favorecer a ligação de BmK M1 em canais de sódio de ambos insetos e mamíferos. A respeito dos resíduos aromáticos importantes, Sun e colaboradores relataram que Trp38 e Tyr42 são envolvidos na atividade da toxina (Sun et al. 2003), enquanto Tyr5, Tyr35 e Trp47 são essenciais para manter a estrutura geral da toxina. Tyr14 tem um papel duplo, já que é essencial para estabilizar a conformação da toxina e é também envolvido na interação com o canal de sódio.

Resíduos básicos das α -toxinas de escorpiões contribuem para a ligação da toxinas no canal (Kharrat et al. 1989; Kharrat et al. 1990; Loret et al. 1990; Loret et al. 1994; Fletcher et al. 1997; Zilberberg et al. 1997; Hassani et al. 1999; Takahashi et al. 2000) mas eles não são os únicos determinantes na interação toxina-canal. A toxina ATX III da anêmona do mar apresenta as mesmas características incluindo Pro5, Tyr5, Trp8, Pro12, Trp13, Arg1 e Lys26. Manoleras e Norton mostraram que esses resíduos poderiam ser funcionalmente importantes (Manoleras and Norton 1994). Características similares de superfície foram também observadas nas toxinas HaTx1 e GrTx, que mostraram que interagem com ambos canais K_v and Ca_v , compartilhando um sítio comum de ligação no K_v (Li-Smerin and Swartz 1998). Infelizmente, a estrutura da GrTx não está disponível ainda. Takahashi e colaboradores fizeram a hipótese que nas toxinas modificadoras de *gating*, a superfície ativa, responsável pela ligação da alça S3-S4 dos canais iônicos voltagem-dependentes, é um arranjo de resíduos hidrofóbicos e básicos (Takahashi et al. 2000). A diversidade de sequências das toxinas modificadoras de *gating* e as variações nas sequências das regiões das alças S3-S4 dos canais iônicos podem explicar as diferenças observadas em afinidade e forma das superfícies de interação.

I.5. Métodos experimentais

A investigação das toxinas depende dos métodos experimentais, o que apresenta várias dificuldades técnicas. Primeiro, a extração dos venenos de animais é uma tarefa precisa e perigosa, que requer habilidade e técnica. As etapas de separação dos componentes do veneno e de purificação das toxinas são geralmente realizadas usando HPLC (*“High Performance Liquid Chromatography”*) e FPLC (*“Fast Protein Liquid Chromatography”*), acompanhadas por ensaios de toxicidade e espectrometria de massa. Os venenos são misturas complexas de proteínas, ácidos nucleicos e sais, e a maioria dos animais liberam somente uma quantidade muito baixa de veneno. Por conseguinte, a quantidade resultante de toxina pura é geralmente muito pequena. Assim, a determinação experimental da estrutura após a extração da proteína do veneno é frequentemente impossível, por requerer quantidades significativas de toxina pura (ler parágrafo *“Determinação da estrutura das proteínas”*). Para determinar a estrutura primária da toxina (i.e. sua sequência), o sequenciamento de Edman tem sido usado, e mais recentemente abordagens de biologia molecular são também usadas. Uma alternativa à purificação de toxina a partir de veneno bruto consiste em produzir os peptídeos em sistemas de expressão como bactérias. Infelizmente, essa técnica tem mostrado um sucesso moderado, uma vez que a maioria das toxinas assim produzidas não são ativas, provavelmente por causa de enovelamento incorreto e de *mismatch* no pareamento dos resíduos de cisteína. Uma outra estratégia tira proveito das técnicas de síntese peptídica acopladas com uma etapa de re-enovelamento. Essa abordagem tem produzido resultados interessantes e toxinas contendo até 3 pontes dissulfeto foram sintetizadas com sucesso. Resíduos de aminoácidos importantes para a interação e/ou a integridade da estrutura da toxina podem ser identificados usando a mutagênese sítio-dirigida. Contudo, o método depende do sucesso da produção de toxinas nativas e mutantes em quantidade suficiente.

Para testar e estudar a atividade da toxina, células isoladas expressando canais iônicos e receptores nas suas superfícies são utilizadas. Técnicas de eletrofisiologia como *patch-clamp* (Prêmio Nobel de medicina para E. Neher e B. Sackmann em 1991)

e *two-electrode voltage clamp* permitem registrar as correntes nos canais iônicos com e sem toxina na preparação. No entanto, essas técnicas são caras, lentas e requerem operadores altamente qualificados. Métodos de ensaios (*screening*) dos canais iônicos como ligação de compostos marcados radioativamente ou sonda fluorescente sensível à voltagem são também usados, embora eles sofram de desvantagens como o alto potencial de falsos negativos ou positivos, a possibilidade de influenciar o mecanismo de funcionamento dos alvos ou a incapacidade de estudar as interações de drogas com ação dependente do estado do canal (Billen et al. 2008).

Entretanto, progressos notáveis foram feitos como, por exemplo, a elucidação da estrutura cristalizada de um canal K_v (Long et al. 2005; Long et al. 2005) (Prêmio Nobel de química para MacKinnon em 2003). Os mecanismos de ação dos canais iônicos, com ou sem a influência de toxinas animais começam a serem esclarecidos e o conhecimento deve progredir nas décadas futuras.

I.6. Recursos existentes sobre as toxinas animais

Várias sequências de toxinas já foram depositadas nos principais bancos de dados generalistas. *GenBank* (<http://www.ncbi.nlm.nih.gov/Genbank/index.html> (Benson et al. 2009)) contém aproximadamente 1300 sequências nucleotídicas de toxinas animais, enquanto *Swiss-Prot* (que é manualmente anotado e revistado) e *TrEMBL* (que é anotado automaticamente e não é revisado) (<http://ca.expasy.org/> (Bairoch et al. 2004)) contêm perto de 2700 e 100 sequências de aminoácidos de toxinas, respetivamente. Em complemento aos bancos generalistas, bancos de dados especializados apareceram que permitem adicionar aos dados genéricos algumas informações específicas sobre as toxinas animais. Uma revisão dos bancos de dados dedicados às toxinas de um ou mais animal(-is) determinado(s) indica algumas coleções pertinentes. *Tox-Prot* ((Jungo and Bairoch 2005) <http://ca.expasy.org/sprot/tox-prot/>) é o projeto de anotação de toxinas do Swiss-Prot. Por agora, *Tox-Prot* indexou 2620 toxinas de animais venenosos. O banco *SCORPION2* (Tan et al. 2006) é dedicado às toxinas de escorpiões e costumava conter perto de 830 sequências, mas não está mais acessível pela Internet. O *ConoServer*

(<http://research1t.imb.uq.edu.au/conoserver/index.php> (Kaas et al. 2008)) contém 3039 entradas de sequências de aminoácidos de conotoxinas e 1529 sequências nucleotídicas, incluindo patentes, fragmentos e precursores. Ferramentas úteis de análise de sequências (alinhamento, predição de pro-peptídeos, digestão de toxinas por agentes químicos, etc.) são disponíveis na página web do mesmo. O *International Venom and Toxin Database* (<http://www.kingsnake.com/toxinology/>) congrega numerosos dados sobre animais venenosos (dos aspectos bioquímicos dos efeitos das toxinas até experiências práticas da caça dos animais venenosos), com enfoque nos aspectos evolutivos e um interesse pronunciado pelos venenos de serpentes e lagartos. O *Animal Toxin Database* (ATDB) ((He et al. 2008) <http://protchem.hunnu.edu.cn/toxin/index.jsp>) lista 3844 sequências de toxinas animais manualmente anotadas de um ponto de vista funcional. ATDB utiliza uma nova ontologia chamada *Toxin Ontology* (TO) para descrever: (i) a atividade biológica da toxina efetiva; (ii) quais organismo, célula, moléculas interagem com a toxina; (iii) os sintomas devidos à toxina. ATDB permite recuperar toxinas com funções biológicas similares, sem considerar suas estruturas ou espécies de origem. Finalmente, o *Knottin Website* ((Gracy et al. 2008) <http://knottin.cbs.cnrs.fr/>) indexa 1250 proteínas incluindo umas 470 toxinas animais (aranha, escorpião e caramujo marinho) que contêm o motivo ICK nas suas estruturas.

No total, esse conjunto de dados abrange a anotação da estrutura primária e as referências das poucas estruturas tridimensionais (3-D) existentes, assim como o possível conhecimento funcional relacionado (sob formato textual).

II. Interações proteína-proteína

Desde o meio dos anos 70 (Chothia 1974; Chothia and Janin 1975), um interesse crescente pelas interações proteína-proteína vem sendo observado na biologia. O termo “*interactome*” foi recentemente introduzido para descrever o conjunto de todas as interações moleculares numa célula, especialmente no contexto das interações proteína-proteína (Plewczynski and Ginalski 2009). De fato, essas interações são cruciais para a maioria dos processos celulares e o entendimento da natureza do

reconhecimento proteína-proteína assim como a caracterização das superfícies de interação abram grandes horizontes para o diagnóstico, a terapêutica e aplicações do desenho de fármacos.

II.1. Caracterização da interface proteína-proteína

A partir do estudo de todos os complexos protéicos conhecidos, foi possível ressaltar características comuns das interfaces dos complexos. Algumas outras propriedades mostraram-se específicas de certos tipos de complexos, como homodímeros, heterodímeros, receptores-ligantes, enzimas-inibidores, enzimas-substratos, anticorpos-antígenos, que sejam complexos permanente ou transitórios.

II.1.A Características topológicas

A área de uma interface é geralmente definida pela diferença entre as superfícies acessíveis ao solvente de ambas proteínas e a superfície acessível ao solvente do complexo. O método mais comum para medir a acessibilidade de uma superfície usa o algoritmo de Lee & Richards que roda uma sonda esférica do tamanho de uma molécula de água na superfície da proteína (Lee and Richards 1971). Admite-se geralmente que o tamanho da interface varia de 400 até 5 000 Å² e representa perto de 20% da superfície de cada proteína (Jones and Thornton 1996).

A arquitetura da interface foi estudada por alguns grupos. Bogan e Thorn mostraram que alguns “*hot spots*” perto do centro da interface são cercados por aminoácidos menos cruciais (Bogan and Thorn 1998). O anel formado por esses resíduos secundários exclui as moléculas de água da interface, criando condições favoráveis para estabelecer interações eletrostáticas e pontes de hidrogênio, e atrapalha a ataque pela água nos resíduos hidrofóbicos do *hot spot*, melhorando assim a afinidade e diminuindo a velocidade da dissociação do complexo. Um outro grupo confirmou o modelo de Bogan e Thorn e especificou que uma interface proteína-proteína tem que ser feita de 3 tipos de átomos (A, B, C): (i) os átomos da interface incluem todos os átomos que perdem a acessibilidade ao solvente quando o complexo é formado (A, B, C); (ii) os átomos de contato são compostos por todos os átomos que fazem contatos

do tipo Van der Waals através a interface e podem conservar alguma acessibilidade ao solvente (B,C); e os átomos enterrados na interface são representados pelos átomos acessíveis na proteína livre que não tem mais nenhuma área acessível no complexo (B) (Lo Conte et al. 1999).

As regiões de interação das proteínas são principalmente descontínuas na sequência. De fato, elas são feitas de 1 até 11 fragmentos da sequência em aminoácidos que ficam perto no espaço por causa do empacotamento estrutural da proteína (Jones and Thornton 1996).

O caráter estrutural de uma interface é aproximadamente intermediário entre a superfície e o *core* de uma proteína. Alguns estudos mostraram que as interfaces têm geralmente mais hélices α (como as regiões *core*) que outras superfícies protéicas, e que a proporção de outras estruturas como espirais (*oils*), voltas (*turns*) e folhas é a mesma que nas superfícies não interagindo (Jones and Thornton 1995; Tsai et al. 1996). Por outro lado, Neurvith e colaboradores mostraram que as interfaces são preferencialmente localizadas em alças compridas, com tamanho médio de 11 resíduos de aminoácidos (Neuvirth et al. 2004). Outros motivos estruturais como o feixe (*bundle*) de 4 hélices, o barril β (β -barrel) (como em complexos anticorpo-antígeno por exemplo) e o sanduíche β são frequentemente encontrados em interfaces proteína-proteína (Jones and Thornton 1996).

II.1.B. Propriedades físico-químicas

II.1.B.i. Hidrofobicidade

Há mais de 15 anos, sabe-se que a maior força de condução do enovelamento das proteínas globulares é a interação das cadeias laterais hidrofóbicas dentro da molécula (Dill 1990; Rose and Wolfenden 1993). Vários estudos mostram que a hidrofobicidade é também uma característica-chave na associação das proteínas (Janin et al. 1988; Tsai et al. 1996). Regiões hidrofóbicas expostas na superfície têm que ser cobertas para estabilizar a interação (Lijnzaad and Argos 1997). A primeira análise estatística demonstrou que as regiões de interação contêm mais resíduos de aminoácidos hidrofóbicos do que o resto da superfície (Chothia 1974; Chothia and

Janin 1975; Argos 1988). Estudos mais recentes confirmaram que as regiões hidrofóbicas mais largas na superfície de uma proteína são também aquelas que são geralmente envolvidas numa interação (Lijnzaad and Argos 1997). Essas regiões de interação são feitas de pequenas áreas (*patches*) hidrofóbicas, correspondendo aos *hot spots* de Bogan and Thorn (Bogan and Thorn 1998), cercadas de resíduos carregados e polares (Larsen et al. 1998; Chakrabarti and Janin 2002). É interessante notar que a hidrofobicidade média pode variar de um tipo de complexo protéico a um outro. Assim, a interface enzima-inibidor tem uma hidrofobicidade média (calculada segundo o índice de Janin (Janin et al. 1988)) intermediária entre aquela de um *core* protéico e uma superfície, e uma interface de complexo anticorpo-antígeno tem uma hidrofobicidade similar àquela de uma superfície protéica comum (Janin et al. 1988). Por comparação, a interface de um heterodímero parece ser menos hidrofóbica que aquela de um homodímero. Resumindo, as interfaces de complexos nos quais os dois parceiros podem existir separadamente são menos hidrofóbicas que aquelas que existem somente na forma complexada (Janin et al. 1988).

II.1.B.ii. Composição em aminoácidos

Segundo a literatura, os resíduos de aminoácidos mais frequentemente encontrados nas superfícies das proteínas são o triptofano, a arginina e a tirosina (Jones and Thornton 1996; Tsai et al. 1997; Bogan and Thorn 1998; Norel et al. 2001). Um outro estudo mostrou que os resíduos de valina, leucina e isoleucina são encontrados no anticorpo nas regiões de alças que se ligam ao antígeno (Padlan 1990). A arginina é um dos resíduos mais frequentes das interfaces (Jin et al. 1992; Young et al. 1994; Bogan and Thorn 1998; Lo Conte et al. 1999; Chakrabarti and Janin 2002), e é mais presente nas largas do que nas pequenas interfaces (Glaser et al. 2001). Sua presença pode ser pelo menos parcialmente esclarecida pela sua capacidade de gerar interações hidrofóbicas graças à sua cadeia carbonada comprida e seu caráter altamente eletropositivo no pH fisiológico. O resíduo de prolina, agindo como quebrador de hélice na superfície de proteínas, é também envolvido nas regiões de interação uma vez que seus átomos polares são capazes de reforçar atrações hidrofóbicas (Tsai et al. 1997; Glaser et al. 2001). Resíduos hidrofóbicos são mais

frequentes nas interfaces mais largas, de homodímeros ou de complexos permanentes, enquanto resíduos polares são encontrados preferencialmente nas menores interfaces, de heterodímeros ou de complexos transitórios.

II.2. Peculiaridades da interação anticorpo-antígeno

Um antígeno (Ag) pode ser qualquer molécula (proteína, ácido nucléico, açúcar, etc.) mas somente o caso do antígeno protéico será descrito aqui. Quando o antígeno é uma proteína, a interação anticorpo-antígeno é somente um caso particular de interação proteína-proteína para qual a maioria das características previamente descritas se aplicam. Porém, algumas diferenças ocorrem e algumas características típicas aparecem.

II.2.A. Sobre o anticorpo

A imunoglobulina (Ig), também chamada anticorpo (Ac), é um dos principais atores do sistema imunológico. É uma glicoproteína encontrada essencialmente na superfície das células que as sintetizam (linfócitos B), e no plasma sanguíneo. Anticorpos produzidos por um mesmo clone de linfócito B são chamados anticorpos monoclonais (mAc). Eles são homogêneos e apresentam a mesma estrutura molecular (Revillard et al. 1998). Além do papel essencial que eles têm na célula, mAc representam uma ótima ferramenta na pesquisa aplicada à saúde, já que podem ser usados para isolar e caracterizar o produto de um gene, identificar regiões de ligação de uma proteína ou desenvolver um diagnóstico viral, por exemplo. Uma variedade de anticorpos é naturalmente produzida por um organismo em resposta à introdução de um antígeno estranho, que eles reconhecem em regiões diversas. O soro resultante é chamado policlonal (por extensão, a expressão “anticorpos policlonais” pode ser usada) e corresponde a uma coleção de mAb.

Uma Ig é composta por duas regiões funcionais. Uma é responsável por reconhecer o antígeno e ligar-se, enquanto a outra é encarregada de desencadear os processos (como a fagocitose) que vão levar a libertar o organismo do intruso.

Três estruturas tridimensionais completas de Ac foram experimentalmente determinadas (Harris et al. 1997; Harris et al. 1998; Saphire et al. 2001), assim como numerosos fragmentos, e duas outras estruturas de Ac foram modeladas computacionalmente (Boehm et al. 1999; Furtado et al. 2004). Portanto, a estrutura 3-D peculiar dessa molécula já foi bem caracterizada.

II.2.B. Epitopo e paratopo

A região de um antígeno que interage especificamente com o anticorpo é chamada “epitopo”. Do mesmo jeito, a região do anticorpo que reconhece especificamente um antígeno é chamada “paratopo”. Distinguem-se dois tipos de epitopos (Atassi and Smith 1978) (Figure 4). Um epitopo “contínuo” (ou “linear”) é um fragmento sequencial do antígeno, enquanto um epitopo “discontínuo” (ou “conformacional”) é composto por várias porções separadas de sequências que são agrupadas no espaço devido ao enovelamento tridimensional do antígeno. A maioria dos epitopos é discontínua e depende altamente da integridade das suas conformações para serem reconhecidos pelo anticorpo (Benjamin et al. 1984; Berzofsky 1984). Ao contrário dos epitopos discontínuos, epitopos contínuos ainda se ligam ao anticorpo mesmo após a denaturação do antígeno.

O tamanho de uma interface Ac/Ag varia de 10 até 20 resíduos de aminoácidos das duas moléculas, estabelecendo entre 15 e 22 contatos (Van Regenmortel 1996). O epitopo pode conter estruturas secundárias como hélices α , dobras β e outros tipos de voltas.

As interfaces Ac/Ag têm uma superfície cerca de 51% não polar (Lo Conte et al. 1999). Os resíduos aromáticos triptofano e tirosina são muito frequentes nos paratopos (Padlan 1990; Lo Conte et al. 1999), porém, as interfaces Ac/Ag incluem menos resíduos valina, isoleucina e leucina (Padlan 1990) e resíduos carregados positivamente, exceto arginina que pelo contrário, predomina.

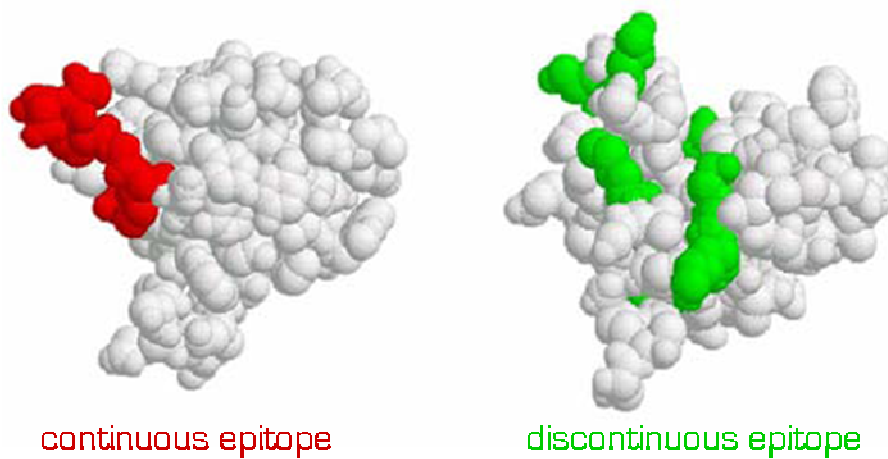


Figura 4. Epitopos contínuos e descontínuos. Os resíduos de aminoácidos envolvidos num epitopo contínuo (vermelho) são sequenciais, enquanto os resíduos de aminoácidos constituindo um epitopo descontínuo (verde) não são necessariamente próximo na sequência mas sim na estrutura tridimensional da proteína. (*Adaptado de Moreau, 2006, tese de doutorado*)

II.2.C. Antigenicidade e imunogenicidade

A antigenicidade é a capacidade de uma molécula de ser reconhecida especificamente por um anticorpo. Estudar a antigenicidade é analisar as reações químicas ocorrendo entre o epítipo e o paratopo e elucidar as bases estruturais da especificidade dessa interação. Devido às suas estruturas, as proteínas são moléculas altamente antigênicas.

Quanto à imunogenicidade, ela é definida como a capacidade de uma molécula de induzir uma resposta imune específica no organismo. Mais do que uma simples ligação, essa propriedade depende de várias interações complexas com um número de atores do sistema imunológico (Revillard et al. 1998; Van Regenmortel 2001). O caráter imunogênico de uma molécula tem que ser definido em relação a dados do hospedeiro e condições experimentais. Uma vez que depende de diversos mecanismos celulares e regulações ocorrendo num dado contexto individual, é uma noção muito relativa. Assim, cada molécula imunogênica é antigênica mas o contrário não é verdadeiro. O melhor entendimento dessas duas propriedades permitirá a produção de vacinas eficientes.

II.2.D. Reação cruzada e poliespecificidade

Apesar do fato que um anticorpo ter fortes afinidade e especificidade para um dado antígeno, a relação anticorpo-antígeno não é exclusiva. Não é raro observar interações desse anticorpo com outras moléculas, que sejam estruturalmente relacionadas ou não. Conseqüentemente, o termo “especificidade” parece às vezes inapropriado (Revillard et al. 1998). Dois termos foram introduzidos para descrever esse fenômeno : “reação cruzada” e “poliespecificidade” (Kramer et al. 1997). Quando é o caso de um epítipo similar em diferentes antígenos reconhecidos pelo mesmo anticorpo, o evento é chamado “reação cruzada”. A poliespecificidade designa associações envolvendo antígenos cujos epítopos não têm nenhuma similaridade um com o outro. Os mimotopos são um caso típico de poliespecificidade (ler o parágrafo “Identificação experimental do epítipo”).

II.3 Métodos de estudo

II.3.A. Métodos experimentais

II.3.A.i. Interações proteína-proteína

O procedimento que permita a caracterização completa de uma interação proteína-proteína é composto por cinco etapas consecutivas : (i) identificar e isolar o complexo; (ii) identificar e separar as proteínas participantes; (iii) testar e validar a interação *in vitro* e *in vivo*; (iv) caracterizar as cinéticas da interação; (v) determinar a estrutura tridimensional do complexo (Moreau, 2006, tese de doutorado).

Para identificar, isolar e analisar complexos protéicos, técnicas como a imunoprecipitação, a espectrometria de massa (Gavin et al. 2002; Ho et al. 2002), “*protein chips*” (Zhu et al. 2001) ou abordagens híbridas (Tong et al. 2002) são geralmente usadas. A cromatografia de exclusão por tamanho (“*Size Exclusion Chromatography*” ou SEC) pode ser usada para descobrir se uma molécula está presente na sua forma monomérica ou complexada, e estimar seu tamanho.

Electrofórese de uma ou duas dimensões, SEC e cromatografia líquida de alta performance (“*High Performance Liquid Chromatography*” ou HPLC) são usadas para separar os parceiros do complexo. Esses são identificados por *Western Blot*, microsequenciamento ou digestão por tripsina e técnicas de espectrometria de massa. Para testar e validar a interação *in vitro*, a técnica de referência é o sistema duplo-híbrido em leveduras (Fields and Song 1989; Luban and Goff 1995). Pode recorrer à cromatografia de afinidade em proteínas purificadas. Em seguida, métodos *in vivo* como o FRET (“*Fluorescence Resonance Energy Transfer*”) podem ser usados (Trakselis et al. 2005). As cinéticas de interação, afinidade e estequiometria são calculadas por filtração de gel, análises calorimétricas, espectrometria de fluorescência SPR (“*Surface Plasmon Resonance*”) (por exemplo usando Biacore®) (Fagerstam et al. 1990; Saunal and Van Regenmortel 1995), ultracentrifugação de densidade ao equilíbrio, microcalorimetria, radiomarcagem ou ELISA. A mutagénesse sítio-dirigida consiste em trocar um resíduo de aminoácido por um outro e testar o efeito da troca na interação. O método permite identificar com precisão os resíduos importantes para a interação.

Finalmente, existem diferentes métodos para determinar a estrutura tridimensional de uma proteína, que serão descritos no próximo capítulo (ler o parágrafo “Determinação da estrutura protéica”).

II.3.A.ii. Identificação experimental do epitopo

A biofísica, a bioquímica, a biologia molecular e a síntese química permitiram o desenvolvimento de vários métodos para identificar epitopos (Meloan et al. 2000).

Um epitopo pode ser identificado usando fragmentos proteolíticos do antígeno sozinho (Cleveland et al. 1977; Jemmerson and Paterson 1986) ou complexado com o anticorpo (Sheshberadaran and Payne 1988), e testando sua reatividade com o anticorpo. Assim, a técnica de síntese múltipla SPOT (Frank 2002) foi desenvolvida para obter rapidamente um grande número de peptídeos diferentes. Esse método permite identificar epitopos contínuos principalmente.

Um outro método muito útil necessita a seleção de mimotopos. Um mimotopo é uma molécula capaz de se ligar ao sítio de associação do antígeno num anticorpo, não necessariamente idêntico ao epitopo reconhecido pelo anticorpo, mas uma mímica aceitável das principais características essenciais do epitopo (Geysen et al. 1986). Um mimotopo representa um epitopo protéico mas pode também mimetizar a interação de um ácido nucléico ou de um polisacárido (Valadon et al. 1996; Wun et al. 2001). Mimotopos são selecionados dentro de um banco de peptídeos testados contra o anticorpo de interesse. Geralmente, tais bancos de peptídeos são construídos usando a técnica do “*Phage Display*” (Smith 1985), que consiste em gerar peptídeos com sequências randômicas que são expostas na superfície de fagos filamentosos e apresentados ao anticorpo. Vários epitopos foram também identificados por comparação das interações do Ac com proteínas homólogas ao Ag cujos epitopos tenham sido previamente caracterizados (Benjamin et al. 1984). Um outro método é baseado na interferência entre vários Ac ligando-se ao mesmo Ag na mesma região ou em regiões sobrepostas.

II.3.B. Métodos preditivos

II.3.B.i. Interações proteína-proteína

A predição das interações proteína-proteína pode ser dividida em três casos de interesse biológico: (i) uma interação é suspeitada mas um dos parceiros não foi ainda identificado; (ii) regiões potenciais de interação de uma proteína, cuja estrutura primária (ou terciária, no melhor caso) é conhecida, precisam ser identificadas, sem conhecimento do(s) potencial(-ais) parceiro(s); (iii) os dois parceiros são estruturalmente bem caracterizados e se quer prever o arranjo espacial do complexo, com ou sem conhecimento das regiões de interação de uma ou das duas moléculas. Vários métodos foram desenvolvidos para resolver o primeiro caso ("*gene neighborhood*" (Dandekar et al. 1998; Overbeek et al. 1999), "*rosetta stone*" (Marcotte et al. 1999; Eisenberg et al. 2000), "*correlated mutations*" (Gobel et al. 1994; Olmea and Valencia 1997)) e foram amplamente passados em revista por Violaine Moreau (Moreau, 2006, tese de doutorado).

O terceiro caso é estudado por métodos de docking. Esses métodos são baseados na hipótese que uma interação favorável tem uma alta afinidade de ligação. Explorando as configurações espaciais, um grande número de interações potenciais são simuladas e avaliadas em termo de energia (Halperin et al. 2002; Smith and Sternberg 2002).

A respeito do segundo caso, a maioria das predições sucedidas são realizadas na estrutura tridimensional da proteína. Assim, vários parâmetros estruturais são usados para prever as regiões de interação de uma dada proteína: propriedades geométricas, habilidade de alguns grupos químicos a formar pontes de hidrogênio, cavidades enzimáticas, taxa de hidrofobicidade e carga relativa das superfícies (Jones and Thornton 1996; Jones and Thornton 1997; Fariselli et al. 2002; Liang et al. 2004).

No entanto, estudos podem ser conduzidos usando a simples sequência, se nenhuma estrutura for disponível. Neste caso, alinhamentos múltiplos de sequências são frequentemente usados para detectar resíduos conservados entre proteínas de uma mesma família funcional (Gotoh 1992; Casari et al. 1995).

II.3.B.ii. Interações Ac/Ag e localização do epitopo

Graças aos bancos de peptídeos, é possível selecionar vários peptídeos capazes de ligar-se à proteína alvo. Desse jeito, um grande número de mimotopos pode ser obtido usando um anticorpo como sonda. A comparação do conjunto de mimotopos com a sequência do antígeno nativo para identificar o epitopo reconhecido pelo anticorpo não é uma tarefa simples. De fato, a literatura relata a dificuldade em localizar epitopos após selecionar mimotopos compartilhando muito pouca similaridade com o antígeno (Stephen et al. 1995; Bonnycastle et al. 1996; Rudolf et al. 1998). Para lidar com essa dificuldade, várias ferramentas foram desenvolvidas. PEPTIDE é o primeiro programa combinando métodos computacionais com resultados experimentais para localizar um epitopo num dado antígeno, usando mimotopos (Pizzi et al. 1995). O método FINDMAP é baseado somente na análise das sequências de mimotopos (Mumey et al. 2003), enquanto dois outros métodos (Mapitope e SiteLight) usam a informação contida na sequência e na estrutura 3-D do antígeno (Enshell-Seijffers et al. 2003; Halperin et al. 2003). Mais recentemente, a ferramenta MIMOP mostrou resultados muito bons, combinando diferentes abordagens para analisar as sequências de mimotopos, considerando a sequência e a estrutura 3-D do antígeno (Moreau et al. 2006).

Uma outra estratégia consiste em prever epitopos potenciais na superfície. Esse tipo de predição pode ser baseada na sequência ou na estrutura 3-D. Na primeira categoria, várias ferramentas foram desenvolvidas e aplicam-se muito bem ao caso da predição de epitopos contínuos. Portanto, 11 ferramentas têm sido listadas que predizem a antigenicidade de uma proteína a partir das suas características de sequência. Até agora, BEPITOPE realizou as predições de epitopos contínuos mais precisas (Odorico and Pellequer 2003). Várias características de sequência são consideradas nesses métodos e podem ser combinadas: (i) acessibilidade dos resíduos de amino ácido ao solvente (Lee and Richards 1971); (ii) escalas de hidrofobicidade ou hidrofiliabilidade (Hopp and Woods 1981; Parker et al. 1986); (iii) escala de antigenicidade (Welling et al. 1985); (iv) flexibilidade (Karplus and Schulz, 1985); (v) elementos de estrutura secundária (Pellequer et al., 1994). Surpreendentemente, dois métodos somente requerem informações da estrutura 3-D

para efetuar as predições de epitopo: o programa EMT (*“Epitope Mapping Tool”*) (Batori et al. 2006) e o servidor CEP (*“Conformational Epitope Prediction”*). É deplorável que suas performances respectivas não tenham sido comparadas às daquelas do BEPITOPE. Usar a informação biológica existindo na estrutura 3-D de uma proteína deve inegavelmente trazer um apoio muito mais valioso na predição de prováveis interações com outras moléculas. Por essa razão, a investigação da estrutura 3-D torna-se necessária.

III. Determinação da estrutura protéica

Uma vez que as interações moleculares dependem muito da conformação espacial, a estrutura tridimensional (3-D) é crítica para o entendimento do modo de ação de uma proteína. Assim, conhecer a estrutura 3-D de uma proteína permite descrever topologicamente as propriedades de superfície e então abordar suas funções biológicas potenciais.

As técnicas de sequenciamento de proteínas e ácidos nucleicos tornaram-se relativamente rápidas, simples e baratas, enquanto a determinação experimental da estrutura 3-D continua difícil. Como resultado, uma lacuna entre o número de sequências experimentalmente conhecidas e o número de estruturas 3-D protéicas determinadas (6.413.124 sequências de proteínas no banco de dados RefSeq *vs.* 52.079 estruturas de proteínas no banco PDB – 10 de março de 2009) cresceu nos últimos anos e deve continuar crescendo com a enorme contribuição dos projetos genoma no mundo inteiro.

III.1. Métodos experimentais

Para determinar as estruturas 2-D e 3-D de uma proteína ou de um complexo, diversos métodos experimentais têm se mostrados eficientes. Assim, a cristalografia de raios X, a espectroscopia por Ressonância Magnética Nuclear (RMN) e a criomicroscopia eletrônica são as técnicas preferencialmente usadas para estudar proteínas sozinhas enquanto a tomografia molecular e a microscopia atômica são

especificamente usadas para analisar grandes complexos moleculares. A técnica do Dicroísmo Circular (DC) representa uma alternativa útil para determinar elementos de estrutura secundária, quando os métodos 3-D tradicionais falham. Cada uma dessas técnicas pode produzir resultados muito bons e é geralmente admitido que a cristalografia de raios X e a espectroscopia RMN levam às estruturas mais precisas, o que faz dessas os dois métodos experimentais de referência em biologia estrutural.

III.1.A. Cristalografia de raios X

A técnica da cristalografia de raios X usa a propriedade dos raios X de poder desviar em diversas direções quando eles batem na treliça de um cristal, de um jeito que depende do arranjo dos átomos dentro do cristal. A partir dos ângulos e das intensidades dos raios desviados, é possível construir um mapa tridimensional da densidade eletrônica do cristal. Então, as posições médias dos átomos, suas ligações químicas, sua desordem e outras informações podem ser reconstruídas usando a expressão matemática da transformada de Fourier.

III.1.B. Espectroscopia RMN

A técnica da espectroscopia RMN explora as propriedades magnéticas de dados núcleos atômicos. Em contraste com a cristalografia de raios X, a RMN é limitada às proteínas pequenas em primeiro lugar (normalmente menor do que 35 kDa), embora os progressos técnicos permitem resolver estruturas sempre maiores. A principal vantagem da RMN é que ela trata proteínas em solução. Conseqüentemente, a dinâmica local (flexibilidade) pode ser abordada de melhor maneira do que a cristalografia de raios X o permite. Contudo, a RMN depende de um conjunto complexo de restrições que o modelo a ser produzido deve satisfazer. A grande complexidade do objeto do nosso estudo não é favorável ao uso dessa técnica.

III.1.C. Dicroísmo circular

O método DC é baseado na capacidade de um material de absorver a luz segundo sua polarização. Essa propriedade é encontrada nas moléculas biológicas por causa

da chiralidade do átomo de carbono alfa ($C\alpha$). Dados tipos de estruturas secundárias em proteínas (α -helix, β -sheet, β -turns...) apresentam espectros típicos de absorção e o espectro DC geral de uma proteína pode ser interpretado em termo de porcentagem de resíduos envolvidos em cada tipo de arranjos estruturados e não estruturados (*coil*).

III.1.D. Limitações e dificuldades

Como se pode perceber, todas essas técnicas apresentam desvantagens e limitações. Primeiro, todas dependem da purificação de uma grande quantidade de proteína e necessitam concentrações importantes, o que constitui um procedimento caro, que exige muito tempo e para qual o sucesso não é garantido.

Depois, a cristalografia de raios X depende da cristalização de uma amostra de proteína, o que representa um evento bastante imprevisível. Um outro problema conhecido como o “problema da fase” é relacionado à reconstrução das posições atômicas dentro do cristal, a partir da densidade eletrônica. Modelos estruturais protéicos de baixa resolução são muito difíceis de refinar até a precisão atômica, e a dificuldade aumenta rapidamente com o tamanho da proteína. Além disso, a maior desvantagem da cristalografia de raios X é que ela produz somente estruturas “estáticas”, apesar da “flexibilidade molecular”, ao contrário da espectroscopia RMN.

As limitações da espectroscopia RMN residem essencialmente na resolução do espectro gerado que é diretamente ligado à intensidade dos campos magnéticos aplicados ao sistema.

III.1.E. Progressos na área

Esforços consideráveis têm sido realizados para melhorar essas técnicas. A técnica de cristalografia de raios X pode hoje ser usada com uma radiação síncrotron capaz de difratar em cristais menores e de baixa qualidade. O problema da fase pode ser resolvido usando métodos computacionais (como por exemplo a modelagem por

homologia ou a predição *ab initio*, ler o parágrafo “Métodos preditivos”) (Qian et al. 2007). Em espectroscopia RMN, cujo uso era tradicionalmente limitado às proteínas relativamente pequenas ou aos domínios protéicos, novas técnicas foram introduzidas para resolver o problema de sinal sobreposto e atenuar a relaxação de magnetização observada em grandes proteínas (Markus et al. 1994; Pervushin et al. 1997). Além disso, novos programas estão sendo desenvolvidos para executar automaticamente diversas etapas do processo, custosas em termo de tempo (Guntert 2004; Liu et al. 2005; Rieping et al. 2007).

Contudo, problemas experimentais ainda sobram e métodos preditivos já bem conhecidos apresentam uma alternativa barata e rápida para resolver estruturas 3-D de proteínas.

III.2. Métodos preditivos

Após mais de 50 anos de pesquisa, uma variedade de abordagens ditas “*in silico*” são disponíveis para a predição das estruturas de proteínas. Eles variam de métodos dependendo de informações de evolução comparativa a métodos puramente estatísticos e métodos baseados somente em princípios físicos (Ginalski et al. 2005).

Três estratégias principais são aplicadas para a predição de estrutura de proteínas : a modelagem comparativa, a predição *ab initio* e o *threading*. Elas podem ser combinadas como no caso dos métodos híbridos ou de méta-predição.

III.2.A. Predição *ab initio*

Anfinsen sugeriu que toda a informação necessária para que uma proteína se enovele está contida na sua estrutura primária (Anfinsen 1973). Baseado nessa teoria amplamente aceita, os métodos *ab initio* foram desenvolvidos para simular o processo de enovelamento. Inicialmente, o termo *ab initio* (sinônimo de “*de novo*”) foi introduzido para qualificar métodos que usavam só princípios físicos para modelar todos os parâmetros energéticos envolvidos no enovelamento e encontrar a estrutura de mais baixa energia livre, segundo a “hipótese termodinâmica” que estabelece que

a estrutura nativa de uma proteína é aquela para qual a energia livre alcança o seu mínimo global. Hoje em dia essa definição tornou-se mais geral e inclui as funções estatísticas e físicas derivadas dos dados experimentais e das outras estruturas protéicas conhecidas.

Houve vários progressos na predição *ab initio* nos últimos 10 anos (Baker and Sali 2001) e o método mais bem sucedido chamado Rosetta (Rohl et al. 2004) produziu recentemente uma predição excepcional numa proteína de 112 resíduos de aminoácido, cujo modelo sobrepôs à estrutura resolvida por cristalografia de raios X com uma resolução de 1,4 Å sobre mais de 90 resíduos (Qian et al. 2007).

Embora progressos consideráveis tenham sido observados na predição da estrutura protéica *ab initio*, o método continua produzindo, na maior parte, modelos estruturais de baixa precisão e confiabilidade. Enquanto a predição *ab initio* é inegavelmente o método mais difícil, os últimos resultados são muito encorajadores para a modelagem de pequenos domínios protéicos e oferecem boas perspectivas futuras (Bradley et al. 2005).

III.2.B. Fold recognition

Os métodos de *fold recognition* são baseados em duas observações. Em primeiro lugar, duas proteínas que não compartilham nenhuma relação evolucionária detectável podem todavia apresentar uma estrutura similar. Em segundo lugar, a análise do arranjo espacial dos resíduos de aminoácidos nas estruturas protéicas permitiu identificar preferências de interação que resultem no desenvolvimento de potenciais contatos entre resíduos que podem ser representados por uma simples matriz de substituição (Ginalski et al. 2005). Os métodos de *fold recognition* em proteínas visam detectar similaridades entre estruturas 3-D de proteínas que não são acompanhadas de nenhuma similaridade significativa de sequências. Existem várias abordagens, mas o objetivo único é tentar encontrar *folds* que são compatíveis com uma dada sequência. Ao contrário da comparação de sequências, esses métodos aproveitam-se da informação extra contida nas estruturas 2-D e 3-D. De fato, esses métodos invertem o problema do enovelamento das proteínas: em vez de predizer como uma

sequência vai enovelar-se, eles predizem quão bem um *fold* vai servir para uma sequência.

III.2.C. Modelagem comparativa

Quando aplicáveis, os métodos de modelagem comparativa são aqueles que fornecem as estruturas mais precisas e confiáveis (Ginalski 2006). A modelagem comparativa é baseada na observação que proteínas relacionadas evolutivamente apresentam estruturas 3D similares (Chothia and Lesk 1986). Conseqüentemente, um modelo 3D de uma proteína alvo pode ser construído a partir de uma(s) proteína(s) de estrutura(s) conhecida(s) chamada(s) "*template(s)*", que compartilha(m) tradicionalmente uma similaridade de sequência significativa com o alvo. O procedimento de modelagem compõe-se das etapas seguintes : (i) achar alguma(s) proteína(s) *template* apropriada(s); (ii) alinhar as sequências alvo e *template(s)*; (iii) identificar regiões conservadas estruturalmente e construir a cadeia principal; (iv) prever as regiões estruturalmente variáveis, incluindo inserções e N e C terminais faltantes; (v) adicionar e otimizar as cadeias laterais dos resíduos de aminoácido; (vi) refinar e avaliar o modelo resultante segundo uma função de pseudo energia. Embora cada etapa possa introduzir erros que afetam a qualidade geral do modelo resultante, o uso ótimo da informação estrutural a partir dos *templates* disponíveis e a precisão do alinhamento sequência-estrutura são os fatores que mais determinam a qualidade final do modelo (Ginalski 2006).

III.2.C.i. Comparação sequência-sequência

A estratégia clássica para selecionar estruturas *templates* correspondendo a homólogos próximos a uma dada proteína consiste em alinhá-la com todas as sequências de proteínas de estruturas conhecidas, usando BLASTp (Altschul et al. 1997). BLASTp permite modular os parâmetros de alinhamento como as penalidades de *gap* (abertura e extensão) e a matriz de substituição. Esse método rápido e simples pode encontrar um *fold* para aproximadamente 30% dos genes nos genomas microbiais (Wang et al. 2000). Porém, o método trata as regiões conservadas e variáveis com o mesmo peso fazendo com que elas tenham o mesmo efeito no

alinhamento final, e sua habilidade em detectar homólogos distantes é limitada (Ginalski et al. 2005).

III.2.C.ii. Comparação perfil-sequência e sequência-perfil

Baseando-se na hipótese que o alinhamento de motivos conservados é mais relevante do que o alinhamento de regiões variáveis, outras estratégias têm aparecido – revistadas em (Ginalski et al. 2005) – que usam matrizes de substituição posição-específicas (Bork and Gibson 1996). A matriz de substituição genérica 20 por 20 é substituída por uma matriz de substituição N por 20, chamada perfil, que define o score para alinhar qualquer um dos 20 aminoácidos com cada um dos resíduos da proteína para qual o perfil é construído. Tais perfis são gerados baseado na variabilidade dos aminoácidos encontrados no alinhamento múltiplo da proteína alvo com seus homólogos próximos. Assim, o perfil define uma família de homólogos em vez de uma sequência só (Ginalski et al. 2005). PSI-BLAST (Schaffer et al. 2001; Altschul et al. 2008) é o programa mais popular que executa tal comparação perfil-sequência. Através a geração iterativa de alinhamentos múltiplos e perfis, o programa é capaz de encontrar homólogos distantes que o BLASTp não poderia detectar para a proteína alvo. Outros métodos baseados nos modelos de Markov (*Hidden Markov Models* – HMMs) têm sido desenvolvidos também (Eddy 1998; Karplus et al. 1999). No lugar de penalidades de *gap*, os HMMs calculam probabilidades de inserções e deleções. Ao contrário de PSI-BLAST, os HMMs não filtram o banco de estruturas. Consequentemente, os HMMs são muito mais lentos que o PSI-BLAST, mas são também mais sensíveis.

III.2.C.iii. Comparação perfil-perfil

A comparação perfil-perfil permite uma comparação simétrica da proteína alvo com os *templates* potenciais. O objetivo dessa estratégia é melhorar a sensibilidade da busca de *templates*. Foi mostrado que esse tipo de método pode detectar similaridades entre duas famílias, não detectáveis usando comparação de um perfil com uma sequência (Rychlewski et al. 2000). Vários métodos perfil-perfil foram desenvolvidas (Ohlson et al. 2004; Sadreyev and Grishin 2004; Wang and Dunbrack 2004).

E interessante notar que estratégias de comparação baseadas no perfil são usadas na modelagem comparativa, assim como nos métodos de *threading*, de tal maneira que esses podem agora ser considerados como dois métodos complementares e cruzados.

III.2.D. Ferramentas de modelagem

Para estudos usuais, servidores de modelagem automática são disponíveis na Internet (Swiss prot, CPH, etc.). Para análises mais focadas, ferramentas de modelagem e meta-servidores são numerosos e têm sido extensivamente revistados por Ginalski e colaboradores (Ginalski et al. 2005). Vale citar o pacote Modeller (Fiser and Sali 2003; Eswar et al. 2008) por este ser o programa de referência para a modelagem comparativa. Modeller tem sido constantemente melhorado desde que foi criado por Andrej Sali em 1990 e está disponível de graça no endereço <http://salilab.org/modeller/>. O programa funciona na maioria das plataformas: G4, G5, and Intel Apple Macintosh (OS X Tiger and Leopard), Intel or AMD Linux and Windows PC, SGI, estações de trabalho Sun. No que diz respeito ao método de *threading*, algumas ferramentas muito práticas como PHYRE (novo 3Dpssm) (Bennett-Lovsey et al. 2008), PredictProtein (Rost et al. 2004) ou GenTHREADER (McGuffin and Jones 2003), podem ser executados via a Internet. Vários meta-servidores são disponíveis, tais como o meta-servidor CBS (<http://abcis.cbs.cnrs.fr/>). Para a modelagem de proteínas *ab initio*, o algoritmo Rosetta desenvolvido pelo grupo do David Baker pode ser usado via um servidor Internet ou instalado localmente (Das et al. 2007).

III.2.E. Métodos de avaliação da modelagem

Desde 1994, o desempenho de todos esses métodos de predição estão sendo avaliados bianualmente pela CASP (*Critical Assessment of techniques for protein Structure Prediction*). Os métodos são avaliados na base da análise de muitas predições “cegas” da estrutura de proteínas. Há 15 anos que a CASP está em vigor e progressos enormes têm sido observados durante esse tempo (Moult et al. 2007). O fato que as predições mais precisas ainda dependem mais do trabalho dos grupos de

especialistas do que do resultado dos algoritmos representa um problema para os experimentos da CASP (Ginalski et al. 2005). De fato, fica difícil para os biólogos saber qual método deveria ser usado para suas próprias previsões. Para julgar os métodos computacionais sem a intervenção dos especialistas, a CAFASP (*Critical Assessment of Fully Automated Structure Prediction*) foi lançada em 1998 e os experimentos estão conduzidos em paralelo com a CASP, de um jeito “cego”.

Estudos sistemáticos comprovaram que os modelos comparativos podem carregar informações novas (chamadas “valor adicionado”) com respeito às estruturas *templates* (Chakravarty and Sanchez 2004; Chakravarty et al. 2005). O valor absoluto do melhoramento sobre o *template* é geralmente pequeno, mas é mesmo assim uma proeza considerável e tem sido um objetivo da CASP há um bom tempo (Moult et al. 2007).

No que diz respeito à aplicabilidade das 3 principais estratégias de modelagem, alguns critérios têm de ser observados. É geralmente admitido que as abordagens de *fold recognition* são mais eficientes para prever o *fold* de uma proteína quando esta é rica em motivos de estrutura secundária, e apresenta dificuldades quando se trata de pequenas proteínas pobres em motivos de estrutura secundária, como é o caso das toxinas peptídicas. Quando a proteína de interesse apresenta mais de 30% de identidade de sequência com uma proteína cuja estrutura 3D já é conhecida, a modelagem comparativa é o método mais óbvio de se aplicar. Recorre-se geralmente à previsão *ab initio* quando nenhuma similaridade pode ser detectada com uma família protéica. Contudo, a estrutura 3D geral de uma proteína pode ser predita por modelagem comparativa e partes dela refinadas por métodos *ab initio*, como é feito para porções ou segmentos da proteína que não existem na estrutura *template*. O uso de um método ou outro depende de 2 critérios principais: (i) a similaridade da sequência alvo com outra(s) sequência(s) protéica(s) para quais uma(s) estrutura(s) determinada(s) experimentalmente está(estão) disponível(-is); (ii) qual precisão é necessária para os estudos a serem conduzidos em cima do modelo 3D, indo de relações estrutura-função que não necessitam um modelo muito preciso, até

caracterização molecular da especificidade de substratos que requer um modelo 3D altamente preciso (Figura 5).

III.2.F. Avaliação do modelo

Diversas ferramentas têm sido desenhados para avaliar se estruturas 3D são corretas, que elas sejam modelos experimentais ou preditivos.

Enquanto ele calcula os modelos, o programa Modeller avalia sua qualidade usando uma função de pseudo energia que avalia quão bem os modelos correspondem à derivada dos *templates*. O resultado é dado pelo *score* de *Modeller Objective Function* (Fiser and Sali 2003). Outros programas como Procheck (Laskowski et al. 1993) verifica a qualidade estereoquímica dos modelos. O método Verify3D (Eisenberg et al. 1997) mede a compatibilidade de um modelo protéico com sua sequência, usando um perfil 3D. Vários outros ferramentas e servidores (ProSA (Wiederstein and Sippl 2007), WHATIF (Vriend 1990), CBS, etc.) oferecem a possibilidade de verificar a coerência dos modelos e de medir a qualidade deles.

No entanto, o único jeito de considerar um modelo 3D como uma verdadeira representação da proteína de interesse é de superpor as coordenatas do modelo com as da estrutura experimentalmente determinada da proteína. Por exemplo, o programa MAMMOTH (Ortiz et al. 2002) é usado para fazer um alinhamento estrutural par-a-par do modelo com a estrutura experimental da proteína. Em outras palavras, quão bom que possa ser, um modelo ainda é um modelo e não pode ser considerado como a estrutura “real” até que experimentos venham confirmar sua precisão.

III.3. Coleções e classificações de estruturas 3D

Existem várias classificações de estruturas protéicas disponíveis na Internet. O *Research Collaboratory for Structural Bioinformatics Protein DataBank* (RCSB PDB) é a referência em termo de bancos de estruturas. Os arquivos do PDB contêm informações sobre estruturas de proteínas, ácidos nucléicos e arranjos complexos

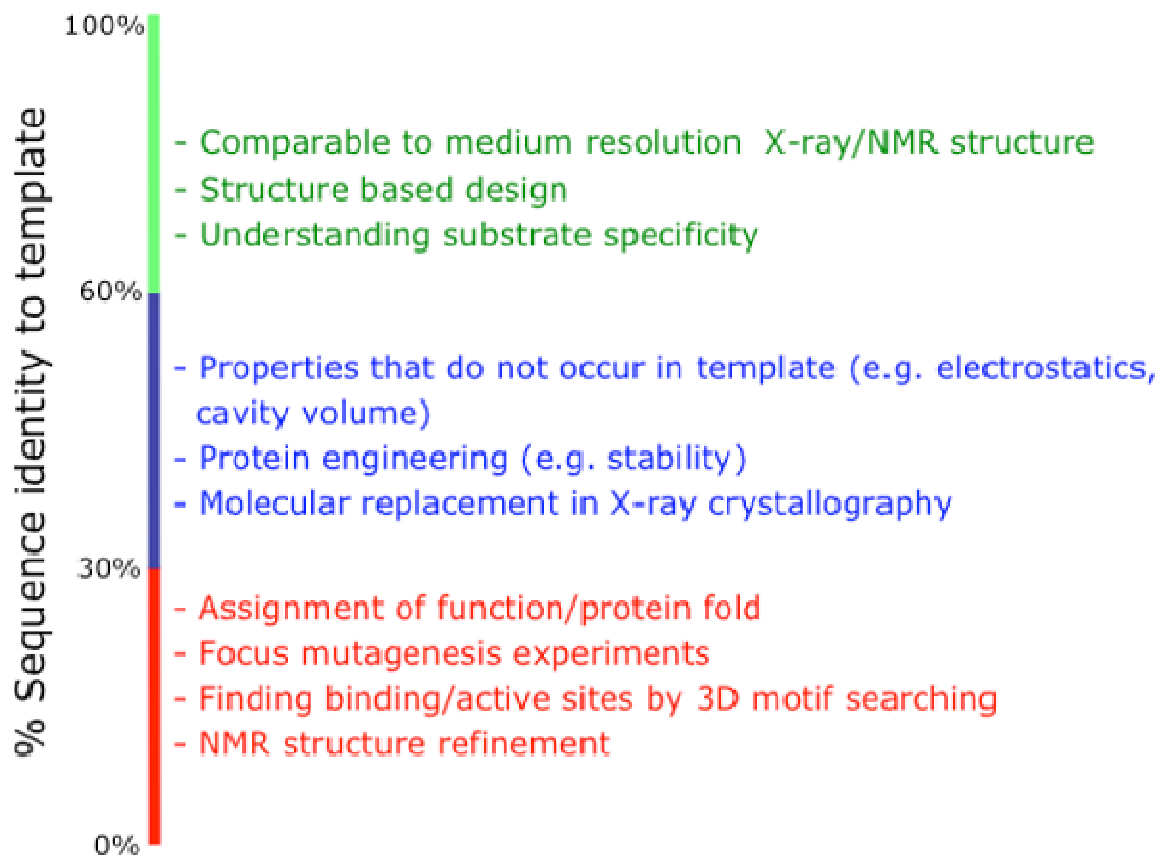


Figura 5. Precisão e aplicações dos modelos obtidos por modelagem por homologia. (De Bourne e Weissig, *Structural Bioinformatics*)

determinados experimentalmente (Berman et al. 2000). O banco SCOP (*Structural Classification Of Protein*), constituído por inspeção manual auxiliada por uma série de métodos automatizados, pretende fornecer uma descrição detalhada e abrangente das relações estruturais e evolucionárias entre todas as proteínas para quais a estrutura 3D é conhecida (Murzin et al. 1995). CATH é uma classificação hierárquica das estruturas dos domínios protéicos, que agrupa as proteínas segundo 4 níveis: Classe (C), Arquitetura (A), Topologia (T) e superfamília de homólogos (H). As fronteiras e avaliações para cada domínio protéico são definidas por uma combinação de procedimentos manuais e automatizados que incluem técnicas computacionais, dados empíricos e estatísticos, revisão da literatura e análise por especialistas (Orengo et al. 1997). O banco de dados Dali é baseado na comparação, todas contra todas, das estruturas 3D protéicas do PDB. Os vizinhos estruturais e alinhamentos são automaticamente inseridos e regularmente atualizados (Holm et al. 2008). O banco 3Dee contém as definições dos domínios estruturais de todas as cadeias do PDB, curados pelo EBI e pelo RCSB, que têm 20 ou mais resíduos e não são modelos teóricos. Estes domínios foram agrupados por similaridades de sequência e de estrutura. As famílias resultando dessa classificação estão estocadas num sistema hierárquico (Siddiqui et al. 2001). HOMSTRAD (*HOMologous STRucture Alignment Database*) é um banco de dados curado de alinhamentos de sequências de famílias de proteínas homólogas baseados na estrutura. Todas as estruturas de proteínas conhecidas são agrupadas em famílias de homólogos (i.e., possuem um ancestral comum), e as sequências de membros representativos de cada família são alinhadas segundo as suas estruturas 3D (Mizuguchi et al. 1998). Esses alinhamentos baseados nas estruturas são anotados por JOY (Mizuguchi et al. 1998) e examinados individualmente.

Vários métodos e ferramentas dedicados ao estudo das estruturas protéicas e das suas interações e à organização dos dados conhecidos, foram desenvolvidos nos últimos anos. A compreensão dos mecanismos de ação das toxinas peptídicas animais conta com a informação disponível sobre as suas estruturas e o jeito com que

elas interagem com seus receptores. Combinando esses métodos que atacam o problema de diversos pontos de vista, desejamos extrair um novo conhecimento sobre as toxinas peptídicas animais, para desenvolver tratamentos clínicos e novas ferramentas biotecnológicas.

OBJETIVO

Como podemos usar a diversidade das toxinas agindo nos canais iônicos para extrair novos conhecimentos sobre estas interações?

ESTRATÉGIA

Cruzar os dados funcionais com as informações estruturais e a predição de superfícies antigênicas.

OBJETIVOS ESPECÍFICOS

- Desenvolver um método para obter modelos tridimensionais confiáveis para as toxinas aracnídicas cujas estruturas experimentais não estão ainda disponíveis.
- Analisar as superfícies das toxinas para evidenciar as características susceptíveis de ter um papel na interação proteína-proteína.
- Estudar as relações estrutura-função-antigenicidade no caso das toxinas Tx2-5 e Tx2-6 da aranha *Phoneutria nigriventer*.
- Organizar os dados num armazém evolutivo dedicado à investigação das toxinas peptídicas animais.

MATERIAL & MÉTODOS

I. Banco de dados : SiliTox

I.1. Dados

Decidimos construir SiliTox, um novo banco de dados que reúne e organiza informações sobre a estrutura 3D, as interações moleculares (incluindo a antigenicidade) e a função das toxinas animais. Para começar este trabalho, enfocamos as toxinas de aranhas e escorpiões após ter excluído as toxinas enzimáticas, as toxinas incompletas (anotadas como fragmentos), as toxinas cuja sequência não inclui nenhum resíduo de cisteína e outras proteínas relacionadas com toxinas (aproximadamente 680 proteínas foram rejeitadas). 221 toxinas de aranhas e 409 toxinas de escorpiões foram finalmente selecionadas e inseridas dentro de SiliTox, totalizando 630 toxinas peptídicas aracnídicas “ricas em cisteínas” (ao menos um resíduo na sequência). Todos os dados estruturais, funcionais e sintômáticos associados à essas toxinas peptídicas foram baixados do banco de dados Swiss-Prot (Boutet et al. 2007) ou extraídos da literatura. Após verificação manual (e correção, caso necessário), os dados foram inseridos nas tabelas apropriadas do banco SiliTox (Figura 6). Para facilitar a verificação dos dados e permitir a inserção automática deles no banco de dados, um formulário de anotação eletrônica escrito em PHP e HTML (*HyperText Markup Language*) foi criado. Novas informações pertinentes para as investigações das relações estrutura-função-antigenicidade foram adicionadas (padrão de distribuição dos resíduos de cisteína, modelos tridimensionais, predição de peptídeos antigênicos, etc.) ou serão adicionadas no futuro (anotação da função segundo o formalismo Bio Ψ (Maziere et al. 2004), aplicações biotecnológicas e farmacêuticas das toxinas peptídicas, etc.), já que SiliTox foi concebido para evoluir.

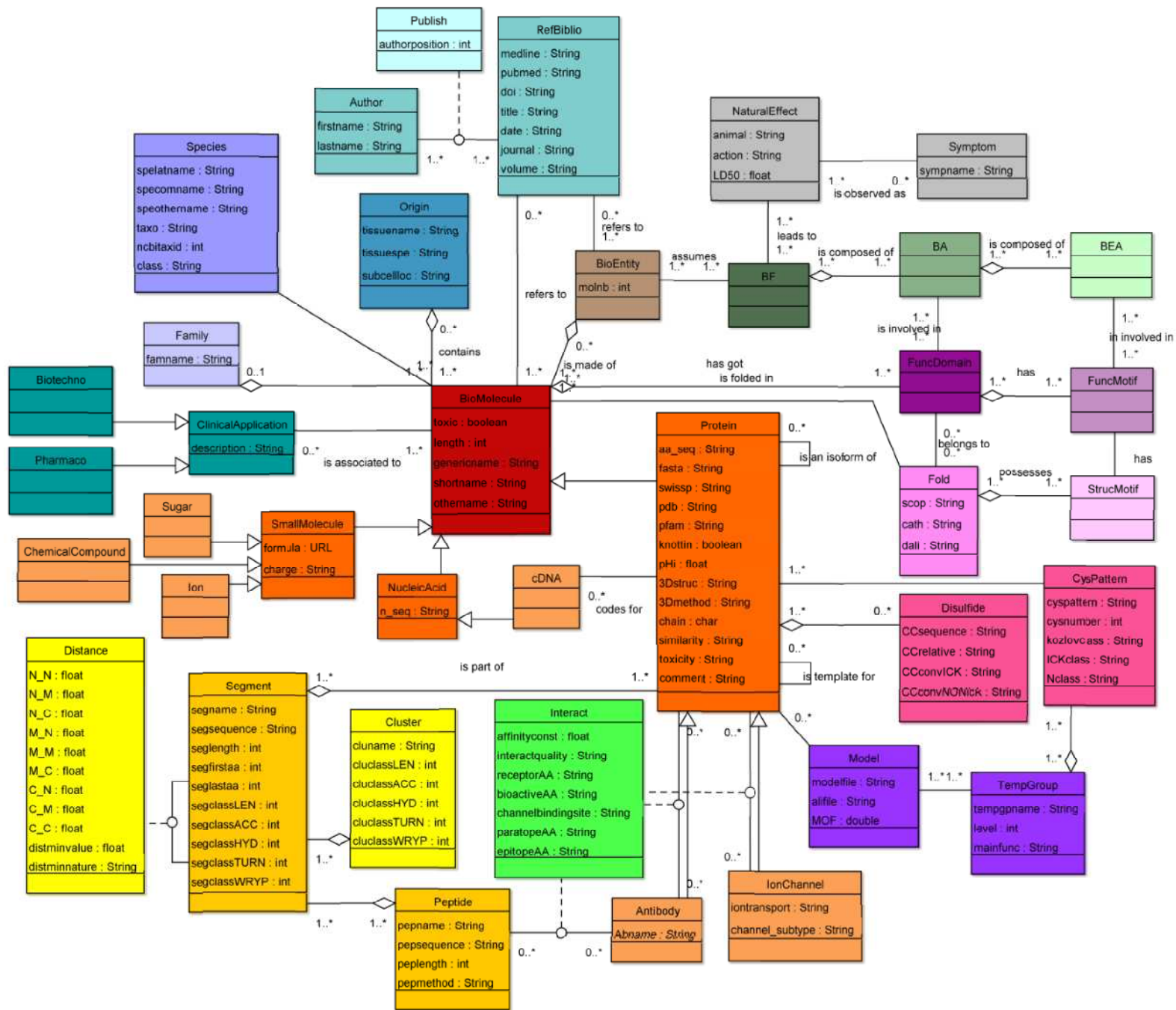


Figura 6. Diagrama de classes do banco SiliTox. O diagrama foi desenhado usando o programa ArgoUML versão 0.18.1. (Veja o texto e os apêndices para a descrição das tabelas.)

I.2. Estrutura do banco de dados

O banco de dados foi concebido e desenhado segundo o método Merise (Tardieu et al. 1983; Tardieu et al. 1985; Rochfeld and Morejon 1989). Primeiro, um esquema conceitual do banco de dados foi desenhado (não mostrado) baseado na teoria das relações (Codd 1970). O diagrama de classes foi então construído usando o program *open-source* ArgoUML versão 0.18.1 (Figura 6). O nível lógico foi lidado pelo uso de phpMyAdmin versão 2.11.9.2 com a linguagem MySQL versão 5.0.45. SiliTox é fisicamente localizado num servidor dedicado ao desenvolvimento no laboratório SysDiag (UMR3145 CNRS / Bio-Rad) em Montpellier na França.

I.2.A. Classes '*BioEntity*'

No SiliTox, os objetos biológicos foram tratados segundo o formalismo multi-escalas Bio Ψ (Maziere et al. 2004). Uma '*BioEntity*' é uma entidade molecular biologicamente ativa, composta de uma ou várias '*BioMolecule*' para realizar uma '*Biological Function*'. No SiliTox, uma *BioMolecule* pode ser uma '*Protein*', um '*NucleicAcid*' ou uma '*SmallMolecule*' ('*Ion*', '*ChemicalCompound*', '*Sugar*'...). Um '*cDNA*' pode ser considerado como um '*NucleicAcid*' cuja sequência deriva do ácido ribonucléico que corresponde a uma toxina peptídica. No SiliTox mais especificamente, uma *Protein* pode ser tóxica (como é o caso da maioria das toxinas peptídicas animais) ou não, pode ser um '*IonChannel*' ou um '*Antibody*'. Essas proteínas podem interagir de um jeito específico. A classe '*Interact*' permite estocar os dados relativos às interações *Protein/IonChannel*, *Protein/Antibody* ou '*Peptide*' antigênico predito por SiliPop/*Antibody*.

I.2.B. Classes Bio Ψ

No que diz respeito aos processos biológicos (função), SiliTox está seguindo o formalismo Bio Ψ (Maziere et al. 2004). Bio Ψ é um esquema multi-escala permitindo a descrição funcional baseada nos elementos básicos de ação. Assim, a função biológica (classe '*BF*') é realizada por uma *BioEntity* e é uma combinação de atividades

biológicas (classe 'BA'). As BAs são efetuadas por um ou mais domínios funcionais (classe 'FuncDom') da(s) BioMolecule(s). As atividades biológicas são compostas de pequenos processos no nível funcional que são chamados de "elementos básicos de ação" (*Basic Elements of Action*, classe 'BEA'). Os BEAs são tijolos teóricos de ação em relação com os motivos funcionais (classe 'FuncMotif'). Ambos 'FuncDom' e 'FuncMotif' correspondem às características estruturais da BioMolecule; as classes 'Fold' e motivo estrutural 'StrucMotif' respectivamente.

I.2.C. Classes 'RefBiblio', 'Species', 'Origin', 'NaturalEffect' e 'Symptom'

As informações relativas às publicações científicas sobre uma determinada BioMolecule, organismo e compartimentos biológicos dos quais essa foi extraída, efeitos induzidos num organismo hospedeiro e os sintomas associados, foram automaticamente lidas no banco de dados Swiss-Prot e inseridas dentro das classes 'RefBiblio', 'Species', 'Origin', 'NaturalEffect' e 'Symptom' do SiliTox, respectivamente.

I.2.D. Classe 'Application'

A classe 'Application' foi concebida para receber os dados sobre aplicações biotecnológicas e farmacêuticas conhecidas que se baseiam no uso de peptídeos ricos em cisteínas. Esse campo de pesquisa está sendo intensivamente investigado. No futuro próximo, SiliTox estará capaz de recuperar e organizar as informações relativas aos biopesticidas, tratamentos da dor ou outras novas drogas derivadas de toxinas peptídicas animais.

I.2.E. Classes 'Disulfide' e 'CysPattern'

Quando conhecida e anotada no Swiss-Prot, a conectividade dos resíduos de cisteína formando pontes dissulfeto numa determinada toxina foram inseridas no SiliTox com uma menção correspondendo ao grau de confiança : 'experimentally determined', 'by similarity', 'probable'. Quando essa informação é ausente mas pode ser inferida por alinhamento com toxinas homólogas, as pontes foram deduzidas e inseridas no

SiliTox com a menção '*SiliTox guess*'. Quando a conectividade dos resíduos de cisteína de uma toxina era duvidosa, nenhuma ponte foi inserida no banco para a dita toxina. Para cada entrada da classe '*Disulfide*', o atributo '*CCsequence*' corresponde às posições sequenciais dos resíduos de cisteína formando uma ponte dissulfeto (exemplo : 'C3 - C17 '). O atributo '*CCrelative*' corresponde às posições relativas do par de cisteínas dentro de todos os resíduos de cisteína da sequência (exemplo : 'C1 - C5').

Para cada toxina, o padrão de cisteínas foi determinado, formalizado segundo a sintaxe PROSITE (Bairoch 1993) e então inserido na classe '*CysPattern*' do SiliTox, junto com a quantidade de resíduos de cisteínas. Como exemplo da sintaxe PROSITE, a sequência em aminoácidos da toxina Tx2-6 da aranha *Phoneutria nigriventer* tem o padrão de cisteína seguinte : 'X(2)-C-X(6)-C-X(3)-C-X(1)-C-C-X(5)-C-X(1)-C-X(3)-C-X(1)-C-X(14)-C-X(2)', com 'C' representando um resíduo de cisteína, 'X' um resíduo de aminoácido qualquer e o número entre parenteses indicando quantos resíduos formam as alças que separam os resíduos de cisteína.

Segundo o conjunto de pontes dissulfeto encontradas em cada toxina, o motivo geral de pareamento das cisteínas pode ser classificado em 3 categorias : '*kozclass*', '*ICKclass*' e '*Nclass*'. Quase todos os padrões descritos na classificação das toxinas de aranhas por Kozlov e Grishin (Kozlov and Grishin 2005) são associados ao motivo ICK. Assim, todas as sequências de toxinas (incluindo a maioria das toxinas de aranhas e algumas toxinas de escorpiões) correspondendo a um desses padrões classificados no artigo de Kozlov, receberam a anotação '*kozclass*' nas suas entradas de *CysPattern*. Todos os outros padrões de cisteínas contendo o motivo ICK, não considerados na classificação de Kozlov mas identificados em *The Knottin Website* (Gelly et al. 2004), foram anotados como '*ICKclass*'. Enfim, os padrões de cisteínas nos quais as pontes dissulfeto não correspondem a um motivo ICK foram anotados como '*Nclass*'. Vale notar que dois padrões de cisteínas que têm a mesma expressão PROSITE mas um apareamento diferente dos resíduos de cisteínas são considerados como duas entradas *CysPattern* diferentes no SiliTox.

As pontes dissulfeto foram anotadas individualmente, segundo a classificação do pareamento geral descrito na entrada *CysPattern* associada. A convenção existente para a numeração dos resíduos de cisteína no motivo ICK (Figura 3) foi usada e adaptada para rotular cada resíduo de cisteína segundo sua posição relativa no motivo formado pelo padrão de pontas dissulfeto. Essa anotação corresponde aos atributos 'CCconvICK' e 'CCconvNONick' na classe *Disulfide* (exemplos: 'CI - CVI'; 'Ca - Cb').

I.2.F. Classes 'TempGroup' e 'Model'

As classes 'TempGroup' e 'Model' são destinadas a estocar os dados de modelagem usadas para / e produzidas pelo programa SiliMod. A classe *TempGroup* contém as informações necessárias para permitir SiliMod de selecionar automaticamente as estruturas *template* pertinentes para a modelagem comparativa das sequências alvos. A classe *Model* contém os dados de saída do programa SiliMod. (Para mais detalhes, leia os capítulos 'Material e Método/Modelagem das estruturas das toxinas por SiliMod' e 'Resultados/Banco de dados de toxinas : SiliTox/Recuperação dos dados'.)

I.2.G. Classes 'Segment', 'Distance', 'Cluster' e 'Peptide'

Quatro classes do banco de dados SiliTox são dedicadas aos resultados do SiliPop da análise das superfícies moleculares e antigênicas das estruturas 3D das toxinas peptídicas. Assim, as classes 'Segment', 'Distance', 'Cluster' e 'Peptide' contêm respectivamente os dados relacionados aos segmentos da *BioMolecule* acessíveis ao solvente, às distâncias calculadas entre os segmentos, ao resultado de *clustering* dos segmentos segundo a matriz de distância e às características do conjunto de peptídeos antigênicos preditos. (Para mais detalhes, leia o capítulo 'Material e Método/Ferramentas de análise das superfícies antigênicas das proteínas : PEPOP & SiliPop'.)

I.3. Implementação do banco de dados

No momento da implementação do banco de dados, cada uma das classes descritas previamente virou uma tabela (*'table'*) e recebeu um novo atributo para identificar cada entrada de um jeito único. Isso corresponde ao atributo chamado chave-primária da tabela. No SiliTox, este identificador único é um número inteiro (*integer*) que é automaticamente incrementado cada vez que uma nova entrada é inserida no banco de dados.

I.3.A. Hierarquia das classes

Para lidar com os níveis hierárquicos vindo da classe *BioEntity* (Figura 6) (*BioMolecule*, *Protein*, *Antibody*, *IonChannel*, *NucleicAcid*, *cDNA*, *SmallMolecule*, *Ion*, *ChemicalCompound*, *Sugar*), as tabelas herdeiras foram “achatas para cima” para a tabela *BioMolecule* (i.e., agrupadas na “tabela mãe” *BioMolecule*), novos atributos foram inseridos para representar o tipo e subtipo de molécula. Todos os atributos das tabelas herdeiras foram inseridos na tabela *BioMolecule* também.

I.3.B. Relações múltiplas

Algumas classes apresentam relações múltiplas entre elas (Figura 6). Para lidar com essas relações, a estratégia adotada é a seguinte : (i) em caso de relação do tipo ‘1-N’ (multiplicidade ‘1- 0...*’ ou ‘1- 1...*’, por exemplo : *BioMolecule* - *Species*), a chave-primária da tabela do lado ‘1’ é inserida na tabela do lado ‘N’ como novo atributo, virando um atributo chamado de “chave estrangeira” (no exemplo, a tabela *BioMolecule* recebe a chave primária ‘*speID*’ da tabela *Species* como chave estrangeira); (ii) no caso de relação ‘N-N’ (multiplicidade ‘0...*- 0...*’, ‘0...*- 1...*’ ou ‘1...*- 1...*’, exemplo : *Segment* - *Peptide*), uma nova tabela é criada e contém as duas chaves primárias das tabelas que ela liga, como atributos de chaves estrangeiras (no segundo exemplo, a nova tabela *SegInPep* foi criada e contém os atributos ‘*segmentID*’ e ‘*pepID*’ como chaves estrangeiras provendo das tabelas *Segment* e *Peptide*).

I.3.C. Tabelas e atributos

Para uma descrição completa e exaustiva de todas as tabelas e todos os atributos do SiliTox, veja o dicionário de dados no apêndice.

II. Modelagem das estruturas 3D das toxinas : SiliMod

II.1. As toxinas Tx2-5 e Tx2-6 da aranha *Phoneutria nigriventer* como caso de estudo

No nosso laboratório, duas toxinas homólogas da aranha brasileira *Phoneutria nigriventer* foram bastante estudadas. Essas toxinas são chamadas toxinas Tx2-5 e Tx2-6 da *Phoneutria nigriventer* ou PnTx2-5 e PnTx2-6 respectivamente. As duas toxinas foram sequenciadas em 1992 (Cordeiro et al., 1992) e são disponíveis no banco de dados Swiss-Prot pelos códigos de acesso TX25A_PHONI (número de acesso O76199) and TX26_PHONI (P29425).

Até hoje, nenhuma estrutura foi determinada experimentalmente para PnTx2-5 e PnTx2-6. Portanto, como alternativa aos métodos experimentais de determinação da estrutura 3D, o método da modelagem comparativa foi aplicado para gerar estruturas 3D teóricas para PnTx2-5 e PnTx2-6 e propor hipóteses baseadas nos resultados. Diversas estratégias de seleção de *templates* foram testadas.

II.1.A. Seleção de *templates*

II.1.A.i. Baseada na similaridade de sequências

Para selecionar estruturas *template*, a abordagem clássica baseada na similaridade de sequências foi inicialmente testada. Usando os servidores NCBI BLASTp e PSI-BLAST (Altschul et al. 1997; Altschul et al. 2008), as sequências das toxinas foram alinhadas às sequências de proteínas cujas estruturas 3D são disponíveis no PDB. Contudo, dado o tamanho e a variabilidade das sequências, ajustamentos manuais foram necessários.

II.1.A.ii. Baseada no reconhecimento estrutural (*fold recognition*)

As toxinas de interesse foram submetidas ao servidor de *fold recognition* mGenThreader (McGuffin et al. 2000), e as predições de estruturas secundárias foram efetuadas nas sequências alvos usando os algoritmos PHD e PROF (Rost et al. 1994), PSIPRED (McGuffin et al. 2000), Jpred (Cuff et al. 1998), e nnPredict (Kneller et al. 1990).

II.1.A.iii. Baseada na composição de aminoácidos

A pesquisa de *templates* foi também feita com o programa PROPSEARCH (Hobohm and Sander 1995). Este programa foi concebido para achar a família protéica putativa quando os métodos clássicos de alinhamento não são capaz de detectar os homólogos estruturais e/ou funcionais de uma proteína alvo. O programa não considera a ordem sequencial dos resíduos de aminoácidos mas a composição geral em aminoácidos, o peso molecular, o conteúdo em pequenos resíduos, o conteúdo em resíduos volumosos, a hidrofobicidade média, a carga média, etc., para um total de 144 propriedades medidas. As sequências de PnTx2-5 e PnTx2-6 foram assim submetidas à análise por PROPSEARCH via o servidor *on-line*.

II.1.A.iv. Baseada na similaridade de padrões de cisteínas

Desenvolvemos nossa própria estratégia para a seleção de *templates*. Esse método é baseado na observação que a maioria das toxinas peptídicas de aranha são muito ricas em resíduos de cisteína (~15-20% dos resíduos) e que as posições relativas dos resíduos de cisteínas estão relativamente conservadas, apesar de uma diversidade significativa entre as sequências das toxinas. Portanto, o padrão de conectividade das cisteínas de PnTx2-5 e PnTx2-6 foi deduzido por alinhamento de sequências com todas as toxinas peptídicas de aranha contendo pelo menos 8 resíduos de cisteína e cujas pontes dissulfeto tenham sido identificadas experimentalmente (Figura 10A). As análises do *fold* foram conduzidas usando os bancos SCOP (Murzin et al. 1995) , CATH (Orengo et al. 1997) e DBAli (Marti-Renom et al. 2007).

II.1.A.v. Seleção final

Dentro do grupo de toxinas de aranhas de estruturas conhecidas, estruturas *templates* foram selecionadas segundo os critérios seguintes: (i) similaridade dos padrões de cisteínas (segundo a classificação das neurotoxinas de aranhas proposta por Kozlov e Grishin (Kozlov and Grishin 2005)); (ii) similaridade das sequências (usando o algoritmo de alinhamento local BLAST); (iii) similaridade das estruturas secundárias (por alinhamento das estruturas secundárias preditas usando o algoritmo SOV (Rost et al. 1994) de todas as toxinas peptídicas de aranhas); (iv) similaridade do alvo de ligação, quando conhecido (toxinas modificadoras de *gating* do canal de sódio). As estruturas *templates* selecionadas são aquelas da δ -atracotoxina-Hv1 (Fletcher et al. 1997), da δ -palutoxina IT2 (Ferrat et al. 2005), da toxina AcTx-Hi:OB4219 (Omečinsky et al. 1996) e da μ -agatoxina-1 (Rosengren et al. 2002). As coordenadas dos seus átomos são depositados no PDB (Berman et al. 2000) sob os códigos respectivos 1VTX, 1V91, 1EIT e 1KQH.

II.1.B. Alinhamento das estruturas *templates*

Usando uma rotina adaptada do código do script '*salign.py*' do Modeller (ver apêndice), um alinhamento estrutural dos quatro *templates* foi efetuado pelo Modeller 9v4 (Fiser and Sali 2003).

II.1.C. Alinhamento das sequências alvo-*template* baseado no alinhamento das estruturas dos *templates*

Os alinhamentos de cada sequência alvo com as sequências dos *templates* foram feitos usando a rotina '*multalign.py*' (ver apêndice) do Modeller, baseado no resultado de *fitting* 3D e do alinhamento das sequências baseado no alinhamento das estruturas que foram calculados por '*salign.py*'.

II.1.D. Modelagem e restrição das pontes dissulfeto

O padrão de conectividade dos resíduos de cisteína foi deduzido por alinhamento da sequência do alvo com as dos *templates*. O pareamento resultante é o seguinte : Cys3

- Cys17, Cys10 - Cys23, Cys14 - Cys46, Cys16 - Cys31, and Cys25 - Cys29 para ambas as sequências das toxinas. A rotina *model.py* do Modeller contém os parâmetros de modelagem, os caminhos para os arquivos PDB dos *templates* e a quantidade de modelos a serem construídos. O programa também permite obrigar virtualmente a formação das pontes dissulfeto durante a modelagem. Assim, as posições das cisteínas na sequência alvo foram especificadas nos parâmetros de modelagem do arquivo *model.py* segundo o pareamento previsto das cisteínas (ver apêndice). Para cada toxina, 100 modelos estruturais foram calculados.

II.1.E. Avaliação dos modelos e seleção

Uma rotina foi implementada para classificar todos os modelos calculados segundo seu valor de *Modeller Objective Function* (MOF) e retornar o melhor deles para cada toxina. Em seguida, eles foram avaliados usando *WHAT IF Check Report* (Vriend 1990), a análise do *plot* de Ramachandran por *Procheck* (Laskowski et al. 1993), o *score* de *Verify3D* (Eisenberg et al. 1997), e o *z-score* de *ProSA* (Wiederstein and Sippl 2007). O modelo que satisfaz o melhor os critérios de multi-avaliação foi selecionado para cada toxina. Os potenciais eletrostáticos foram calculados usando *GRASP2* (Petrey and Honig 2003) e as representações das estruturas 3D foram desenhadas usando o programa *PyMol* (DeLano 2002).

II.1.F. Fluorescência do triptofano e dicroísmo circular

Registros de fluorescência da PnTx2-6 foram obtidos com um fluorômetro *ABL2 Spectronic* (Applied Biosystems, Foster City, CA) em cubetas de quartzo de um centímetro de largura. A amostra foi excitada no comprimento de onda de 280 nm e o espectro de emissão foi coletado em 300 e 400 nm à temperatura de 20°C. O espectro de dicroísmo circular (DC) da PnTx2-6 foi obtido com um espectropolarímetro (Jasco J-810) numa cubeta de quartzo de um centímetro de largura com luz UV em 190-260 nm, 23°C. A toxina foi diluída na água. Os dados foram obtidos cada 0.5 nm de intervalo a 50 nm/min. Todos os espectros de DC são médias de 20 varreduras espectrais. Em cada condição, a linha de base foi subtraída do espectro com a toxina.

II.2. Constituição dos grupos estruturais baseados no padrão de cisteínas para a modelagem automática das toxinas

Considerando-se que : (i) a maioria das toxinas peptídicas de aracnídeos são muito ricas em resíduos de cisteína; (ii) as posições desses resíduos nessas sequências são altamente conservadas dentro dos membros de subfamílias; (iii) a variabilidade observada nas sequências reside nas regiões inter-cisteínas; (iv) essas toxinas parecem adotar somente poucos *folders*, fizemos a hipótese geral de que *templates* apropriados para uma determinada sequência de toxina seriam estruturas de toxinas que apresentam padrões de distribuição e pareamento dos resíduos de cisteína similares àqueles da toxina alvo. Esta hipótese foi testada com as toxinas PnTx2-5 e PnTx2-6 e a estratégia de modelagem foi adaptada e estendida a todas as toxinas peptídicas de aracnídeos ricas em cisteínas previamente inseridas no banco de dados SiliTox.

II.2.A. *Templates* disponíveis

Dentro das 409 toxinas de escorpiões presentes no SiliTox, as estruturas determinadas experimentalmente de 72 delas são disponíveis no PDB (57 foram determinadas por RMN, 14 por espectrometria de raios X e 1 é um modelo teórico). Dentro das 221 sequências de toxinas de aranhas, 39 correspondem a estruturas disponíveis no PDB (36 RMN e 3 modelos teóricos). Os modelos teóricos estando deixados de lado, um total de 111 estruturas 3D de toxinas de aracnídeos são possíveis *templates* para as outras 519 toxinas presentes no SiliTox.

II.2.B. Composição dos grupos

Todos os padrões de cisteína existentes dentro das toxinas do SiliTox foram classificados e grupos foram formados segundo a similaridade dos padrões, a quantia de resíduos de cisteína e a conectividade em pontes dissulfeto. Os grupos estruturais formados podem ou não conter algumas toxinas com estrutura 3D conhecida. Escolhemos limitar a cinco a quantia de estruturas *templates* potenciais. Nos grupos grandes que contêm mais de cinco toxinas cujas estruturas 3D são

conhecidas, as sequências de todas as toxinas do grupo foram alinhadas pelo ClustalW (Thompson et al. 2002) e agrupadas segundo a similaridade das suas sequências, usando o algoritmo *Neighbor-Joining* (Saitou and Nei 1987). Assim, novos subgrupos foram formados e inseridos no SiliTox. Num grupo, cada toxina é anotada como '*template*' ou '*target*' segundo se sua estrutura é disponível no PDB ou não. Uma exceção foi feita para as estruturas teóricas do PDB que foram anotadas como '*model*' e não puderam ser usadas como *templates* para a modelagem.

II.3. O algoritmo do SiliMod

A estratégia de modelagem das estruturas 3D usando a seleção de *templates* baseada na similaridade de padrões de cisteínas foi automatizada num novo programa chamado SiliMod, dedicado às toxinas peptídicas ricas em cisteínas. SiliMod usa rotinas escritas em Perl e Python para rodar o programa de modelagem por homologia Modeller na sua versão 9v4, pegando recursivamente as sequências das toxinas de uma lista dada como arquivo de entrada (Figura 7).

II.3.A. Conjunto de alvos

Todas as sequências de toxinas presentes no SiliTox foram extraídas do banco juntas aos seus nomes de acesso do Swiss-Prot (Bairoch et al. 2004) e formatadas num arquivo multi-fasta. SiliMod lê o nome e a sequência de cada toxina do arquivo multi-fasta de entrada e chama uma rotina chamada *ali.pl* que foi escrita para converter a sequência para o formato PIR e escrevê-la no arquivo *proteinname.ali* que será depois lido pelo programa Modeller.

II.3.B. Seleção de *templates*

Para cada sequência do arquivo multi-fasta, SiliMod conecta-se ao banco SiliTox e recupera o grupo estrutural no qual a toxina alvo pertence. Além disso, o padrão de pontes dissulfeto e todos os *templates* disponíveis do grupo são retornados. Se não tiver *template* disponível no grupo estrutural da toxina alvo, o programa retorna uma mensagem de aviso e pula para a próxima toxina do arquivo multi-fasta.

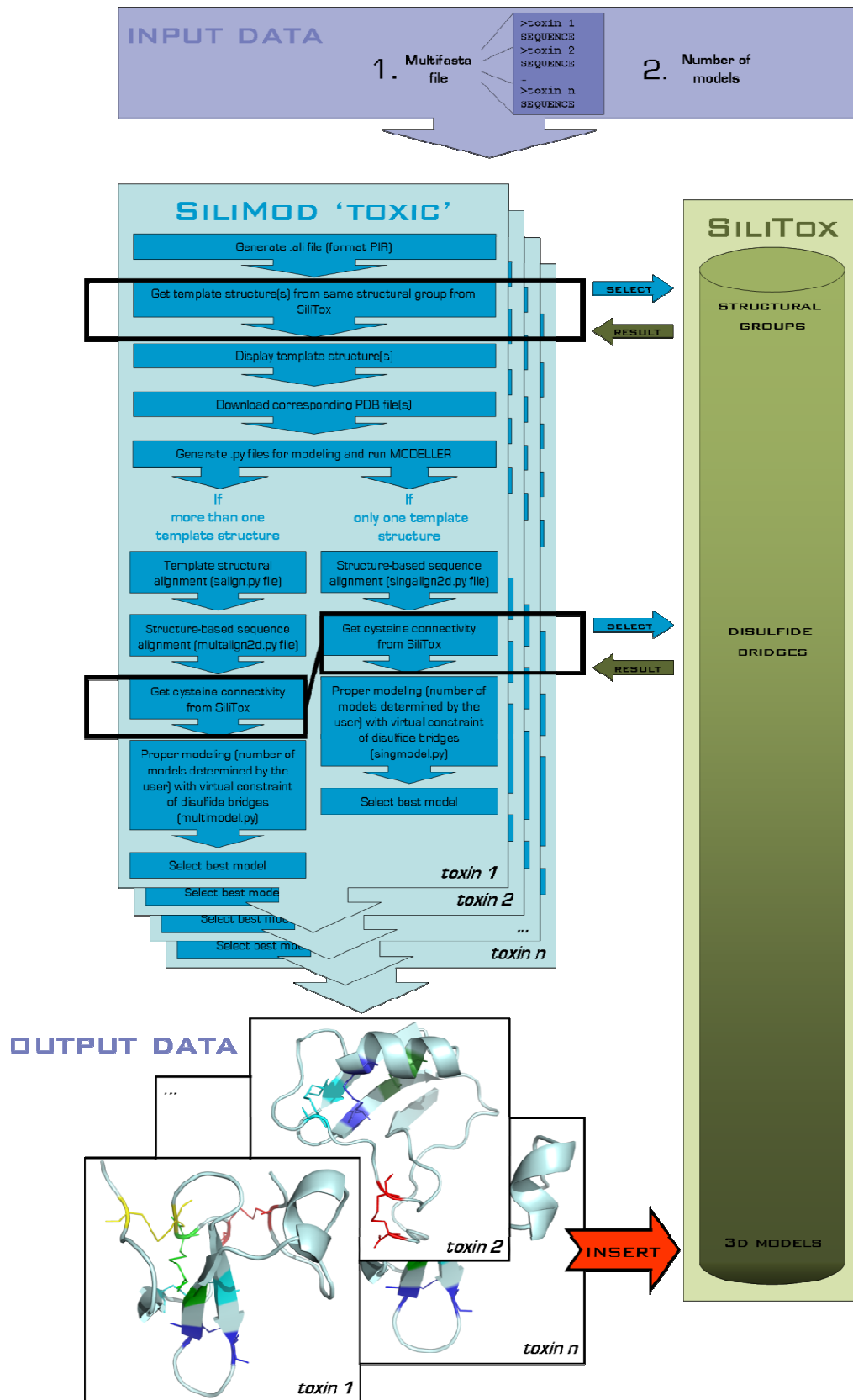


Figura 7. Esquematização do algoritmo do SiliMod.

II.3.C. Alinhamento estrutural dos *templates* (optativo)

Se tiver mais de um *template* para uma determinada sequência de toxina alvo, SiliMod chama uma rotina chamada *salign.pl* e manda para ela o nome da proteína alvo, o caminho para o arquivo *proteinname.ali* contendo sua sequência, os códigos PDB dos *templates* estruturais e os parâmetros para o alinhamento estrutural. Todos esses dados são formatados e escritos num novo arquivo *proteinname_salign.py* pela rotina *salign.pl*. Então, SiliMod chama o programa Modeller para rodar o alinhamento estrutural usando *proteinname_salign.py* como arquivo de entrada. Em saída do *proteinname_salign.py*, vários arquivos são criados, dentro de quais *proteinname_salign.ali* que será requerido por *proteinname_multalign.py* para a próxima etapa do procedimento de modelagem.

II.3.D. Alinhamento das sequências *template(s)*-alvo baseado na(s) estrutura(s)

Se mais de uma estrutura *template* foi detectada para uma toxina alvo, depois do SiliMod ter rodado a rotina *salign.py*, o programa chama a rotina *multalign.pl* para gerar *proteinname_multalign.py*. Do mesmo jeito que *salign.pl* faz, *multalign.pl* e *singalign.pl* requerem o nome da proteína, o(s) código(s) PDB do(s) *template(s)*, o caminho do(s) arquivo(s) PDB e o caminho para *proteinname_salign.ali* ou *proteinname.ali* respectivamente. SiliMod chama então o Modeller para rodar o arquivo previamente gerado (*proteinname_multalign.py* ou *proteinname_singalign.py*). O arquivo de saída será chamado *proteinname_multalign.ali* ou *proteinname_singalign.ali*, dependente de qual rotina foi usada.

II.3.E. Obrigação das pontes dissulfeto e construção dos modelos

Depois, SiliMod conecta-se ao SiliTox para recuperar a quantia de resíduos de cisteína e suas conectividades nas pontes dissulfeto da molécula. Em seguida, o programa chama a rotina *model.py* que formatará as pontes a serem forçadas durante a modelagem. O nome da proteína, o(s) código(s) PDB do(s) *template(s)*, o(s) caminho(s) do(s) arquivo(s) PDB e do arquivo de alinhamento final gerado na etapa

precedente (*proteinname_multalign.ali* ou *proteinname_singalign.ali*) são também escritos no arquivo *proteinname_model.py*, junto ao número de modelos a serem calculados para a sequência da toxina alvo. Finalmente, SiliMod chama Modeller mais uma vez para rodar a modelagem da estrutura 3D correspondente à toxina alvo usando o arquivo *proteinname_model.py* que contém os parâmetros e as restrições de modelagem.

II.3.F. Seleção do melhor modelo

Finalmente, SiliMod abre todos os arquivos PDB dos modelos gerados, lê cada valor de MOF e move o arquivo do modelo de mais baixo score para um outro diretório, que conterá o melhor modelo 3D para cada toxina para qual um ou mais *template(s)* pôde(/puderam) ser identificado(s), no final do procedimento recursivo.

II.4. Validação do método

Para testar o desempenho do SiliMod, modelagens “cegas” foram realizadas com sequências de toxinas cujas estruturas já foram determinadas experimentalmente ou modeladas por outros autores. Todas essas sequências para quais um ou mais *template(s)* puderam ser detectado(s) no mesmo grupo estrutural no SiliTox foram modeladas por SiliMod com a mesma metodologia do que aquela que foi descrita acima. Usando o servidor MAMMOTH (Ortiz et al. 2002), o melhor modelo para cada toxina foi alinhado com a estrutura original do PDB e dois parâmetros foram considerados: (i) o valor de *Root Mean Square Deviation* (RMSD) corresponde à distância em angströms que separa os átomos C α dos esqueletos das duas estruturas protéicas; (ii) o *score* de probabilidade $-\ln(P\text{-value})$ do MAMMOTH representa a similaridade estrutural esperada entre o modelo e a estrutura experimental comparado ao valor esperado quando tira-se ao acaso qualquer par de estruturas. Como *cutoffs* de referência, os *scores* de MAMMOTH de 4.5 e 2.99 correspondem à níveis de confiança de 99% e 95% respectivamente (Ortiz et al. 2002).

III. Ferramentas de análise da superfície antigênica das proteínas: PEPOP & SiliPop

Enquanto a maioria dos métodos existentes para a predição de propriedades antigênicas de uma proteína são baseados na sequência, necessita-se métodos usando a informação 3D para poder prever epítopos descontínuos para derivar peptídeos imunogênicos eficientes. No contexto de um vasto trabalho colaborativo, uma ferramenta dedicada à predição de peptídeos antigênicos foi desenvolvida e validada experimentalmente. O programa é chamado PEPOP. Neste documento, explicações completas serão limitadas ao trabalho que foi realizado no contexto desta tese (implementação do algoritmo PEPOP a partir da etapa de notação dos segmentos até o fim, assim como a validação bibliográfica). No entanto, alguns outros pontos serão detalhados quando necessário para uma boa compreensão do projeto em geral. Para qualquer informação complementar (sob a validação experimental por exemplo), ver o artigo anexado intitulado *PEPOP: Computational design of immunogenic peptides* (Moreau et al. 2008) que foi recentemente publicado no jornal *BMC Bioinformatics*. A segunda ferramenta chamada SiliPop foi desenvolvida depois do PEPOP e faz o objeto de uma segunda publicação (Fleury *et al.*, em preparação).

III.1. Algoritmo do PEPOP

III.1.A. Agrupamento (*Clustering*) dos segmentos acessíveis na superfície

A partir da estrutura 3D da proteína alvo, a região da superfície que fica acessível ao solvente é calculada (Figura 8A, etapa 1). A acessibilidade de cada resíduo de aminoácido é determinada usando DSSP (Kabsch and Sander 1983) com os parâmetros por *default*. Os aminoácidos contíguos e acessíveis são listados como segmentos (um segmento pode ser composto de um único resíduo) (Figura 8A, etapa 2). A posição espacial de cada segmento é aproximada por um segmento geométrico representado por três pontos: os dois primeiros pontos são os C α dos resíduos das extremidades N- e C-terminal do segmento; o terceiro ponto é calculado como o ponto médio entre os dois outros. Assim, um segmento é representado pelas

coordenadas 3D (x, y e z) dos três pontos característicos. Depois disso, uma matriz de distância é calculada dentro da qual a comparação entre dois segmentos produz nove valores (3 x 3). A partir dessa matriz, uma matriz de menor distância é derivada, na qual a distância entre dois segmentos é representada pelo menor dos nove valores previamente calculados. Usando essa matriz, o agrupamento dos segmentos é realizado pelo programa Kitsch do pacote PHYLIP v3.6 (Retief 2000) (Figure 8A, step 3). Nessa altura, um conjunto de segmentos acessíveis à superfície da proteína é arrumado em grupos segundo as distâncias espaciais relativas dos segmentos entre si.

III.1.B. Notação dos segmentos

Segundo diferentes análises de interações Ac/Ag e/ou outras proteína-proteína (Bogan and Thorn 1998; Chakrabarti and Janin 2002), cinco parâmetros foram escolhidos para a influência que foi mostrado que eles têm no caráter antigênico, e até imunogênico, de uma proteína, ou porque eles são encontrados nas interfaces de complexos proteicos. No programa PEPOP, essas cinco características podem ser selecionadas individualmente ou combinadas entre si. Portanto, uma nota é atribuída aos segmentos de uma dada lista (seja a lista dos segmentos de um grupo, ou todos os segmentos existentes, segundo a escolha do usuário) para cada propriedade selecionada pelo usuário : (i) comprimento do segmento; (ii) acessibilidade do segmento; (iii) hidrofobicidade do segmento; (iv) ocorrência dos resíduos de aminoácido especiais que são : triptofano (W), arginina (R), tirosina (Y) e prolina (P); (v) presença de resíduos envolvidos em β -turns. O “*length score*” é o número de resíduos de um segmento. O “*accessibility score*” é a acessibilidade média dos resíduos que compõem um segmento (segundo os valores calculados por DSSP). O “*hydrophobicity score*” é o número de resíduos de aminoácidos hidrofóbicos (consideramos Tyr, Trp, Phe, Leu, Val, Ile, Cys, Pro, Met) presentes no segmento. O número de W, R, Y e P no segmento representem a nota para a propriedade de “*particular amino acid presence*”. Finalmente, o “ *β -turn score*” é dado pelo número de resíduos envolvidos numa estrutura de tipo β -turn, segundo o DSSP.

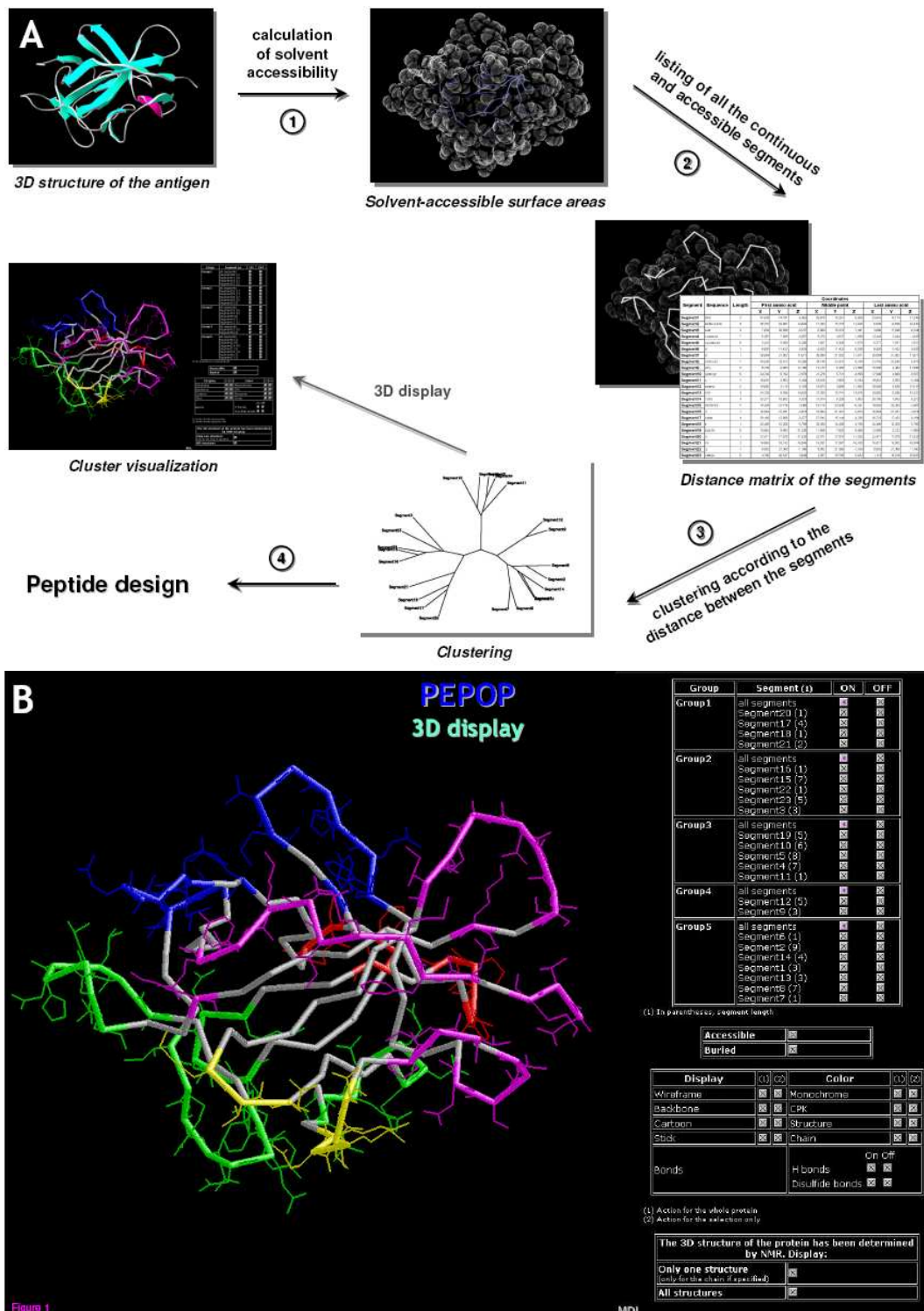


Figura 8. (A) Descrição do algoritmo do PEPPOP. Etapa 1 : A superfície acessível ao solvente é calculada a partir da estrutura 3D de uma proteína. Etapa 2 : Segmentos de aminoácidos acessíveis e contíguos são listados, e a matriz das menores distâncias Euclidianas entre os segmentos é calculada. Etapa 3 : Os segmentos são agrupados em *clusters* segundo a matriz de distância (os *clusters* podem ser mostrados na estrutura 3D da proteína). Etapa 4 : Peptídeos são desenhados segundo as análises do *clustering*. (B) Interface do PEPPOP. Os segmentos e os *clusters* podem ser mostrados numa vista 3D da proteína de uma maneira interativa. (De Moreau et al., 2008)

III.1.C. Notação dos *clusters*

Cada *cluster* recebe uma nota para as cinco propriedades descritas acima, segundo os *scores* dos segmentos que o compoñham. O “*length score*” de um *cluster* é dado pelo comprimento do maior segmento deste *cluster*. Num *cluster*, o número de segmentos que fazem parte do primeiro quartil dos segmentos mais acessíveis representem o “*accessibility score*” do *cluster*. O “*hydrophobity score*” corresponde ao número de resíduos de aminoácidos hidrofóbicos de todos os segmentos contidos no *cluster*. Do mesmo jeito, o “*particular amino acid presence score*” é dado pelo número de resíduos W, R, Y e P dos segmentos pertencendo ao *cluster*. O “*β-turn score*” é representado pelo número de resíduos de aminoácidos envolvidos numa estrutura β-turn dentro de todos os resíduos dos segmentos do *cluster*.

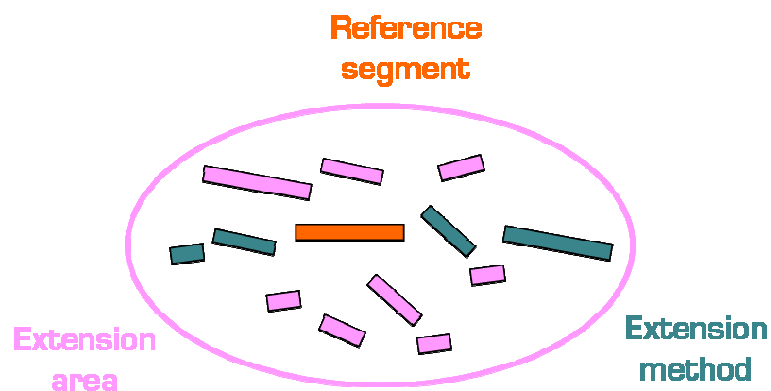
III.1.D. Classificação dos segmentos e dos *clusters*

Para cada uma das cinco propriedades, os segmentos são classificados segundo os *scores* previamente assinados e as cinco posições obtidas por um mesmo segmento na classificação são somadas. Segundo este valor multi-critério final, os segmentos são ordenados mais uma vez. O mesmo processo é aplicado para ordenar os *clusters*.

III.1.E. Desenho dos peptídeos antigênicos e/ou imunogênicos

Para desenhar um peptídeo com potencial antigênico e possivelmente imunogênico também, um “segmento de referência” (*reference segment*) é escolhido pelo usuário a partir do conjunto de segmentos acessíveis calculados e sua sequência é estendida usando um dado “método de extensão” (*method of extension*) numa “região de extensão” (*area of extension*) especificada. O segmento de referência pode ser selecionado manualmente dentro de todos os segmentos, segundo a escolha do usuário, se se quer visar uma determinada região da proteína, em cujo caso o peptídeo predito resultate deve ser capaz de mimetizar essa região. De outra maneira, o algoritmo do PEPOP foi concebido para realizar a seleção automática do “melhor segmento” assim chamado, i.e. o segmento classificado como primeiro

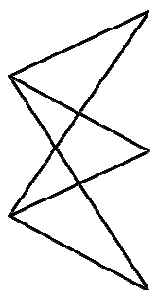
segundo as critérios selecionados dentro das cinco propriedades fisicoquímicas e estruturais dos resíduos compondo os segmentos, e que foram descritas previamente. Assim, após ter ordenado os segmentos segundo as características escolhidas, o segmento de melhor *score* é automaticamente selecionado como segmento de referência. Num segundo passo, o segmento de referência é alongado num peptídeo que atinge um tamanho molecular apropriado. Três métodos de extensão foram implementados. O primeiro, chamado “FPS” (*Flanking Protein Sequence*), consiste em estender os dois lados da sequência do segmento de referência com a sequência que a flanqueia na proteína alvo. Num segundo método, chamado “NN” (*Nearest Neighbors*), o algoritmo procura os segmentos mais próximos do lado C-terminal do segmento de referência no espaço, usando a matriz de distância mínima previamente calculada. A sequência do segmento mais próximo é adicionada àquela do segmento de referência na sua extremidade C-terminal (Figura 9). Se for necessário, a extensão continua do mesmo jeito com o último segmento adicionado, até que o comprimento total (definido pelo usuário) do peptídeo seja alcançado. O terceiro método, chamado “ONN” (*Optimized Nearest Neighbor path*) consiste em achar o caminho mais curto entre os segmentos que constituem o peptídeo. A partir do conjunto de segmentos encontrados usando o método NN, todas as combinações de segmentos são exploradas para selecionar aquela para qual a distância global entre todos os segmentos é a menor possível. O último passo do desenho de peptídeos por PEPOP é a escolha da zona de extensão. O método de extensão de peptídeos pode considerar o conjunto inteiro de segmentos acessíveis da proteína ou somente um dos predeterminados *clusters* de segmentos acessíveis. O usuário pode selecionar este *cluster* dentro de todos os demais ou escolher a opção de seleção automática, em qual caso PEPOP selecionará o “melhor *cluster*” como descrito previamente para a seleção do segmento de referência. De fato, os *clusters* estão classificados segundo as mesmas cinco características que podem ser combinadas juntas: acessibilidade, comprimento, hidrofobicidade, número de resíduos W, R, Y, P e número de resíduos em β -turns. O *cluster* que recebe a melhor nota segundo os parâmetros selecionados é aquele que servirá de zona de extensão.

A**B**

Selection of the reference segment

Automatic selection

Arbitrary selection



Peptide extension method

sequential

nearest neighbors

nearest neighbors with optimized path

Peptide extension area

one cluster

whole set of segments



Figura 9. Descrição esquemática dos métodos de construção dos peptídeos empregados por PEPOP e SiliPop.

III.2. O algoritmo do SiliPop

SiliPop é uma versão totalmente automatizada do PEPOP que permite ao usuário analisar um conjunto de proteínas sob as mesmas condições e todas de uma vez só, em vez de ter que submeter cada alvo um por um e fixar os parâmetros cada vez. Através de uma interface simples, o programa requer do usuário o caminho do diretório contendo todas as estruturas (no formato PDB) a serem analisadas. Do mesmo jeito que PEPOP faz, os segmentos acessíveis são gerados, as matrizes de distância são calculadas e *clusters* de segmentos acessíveis espacialmente próximos são formados. Por default, SiliPop usa os cinco critérios descritos no PEPOP (ver o parágrafo “Notação dos segmentos”) para ordenar os segmentos da proteína. Para adaptar o programa às moléculas pequenas (que contêm muitos segmentos com o mesmo *score* final), o segmento mais comprido é selecionado como referência, no caso de segmentos classificados como primeiros *ex aequo*. A seleção de segmentos de referência para construir os peptídeos é feita de acordo com o resultado da avaliação multi-critérios dentro de dois conjuntos : (1) a lista inteira dos segmentos; (2) a lista de segmentos de cada *cluster*. Depois disso, SiliPop gera três tipos de peptídeos para cada alvo proteico : um peptídeo desenhado pelo método FPS usando o melhor segmento dentro do conjunto (1); um peptídeo desenhado pelo método ONN com o melhor segmento do conjunto (1); e para cada *cluster*, um peptídeo desenhado pelo método ONN a partir do melhor segmento do conjunto (2). A zona de extensão é preferencialmente o *cluster* ao qual o segmento de referência pertence, mas se este último estiver sozinho no seu *cluster*, SiliPop usa o conjunto inteiro de segmentos para estender o peptídeo. Se a combinatória for alta demais (oito segmentos ou mais), o peptídeo será feito de um segmento só. Por *default*, o comprimento de extensão é fixado a 20 resíduos de aminoácidos.

III.3. Avaliação do desempenho da ferramenta

Como existe uma relação documentada entre superfície acessível e antigenicidade (Novotny et al. 1986; Thornton et al. 1986), a correspondência entre os *clusters* de superfície identificados por PEPOP em determinadas proteínas e seus epitopos

conhecidos foi avaliada. Neste objetivo, 13 complexos Ac/Ag em quais os resíduos de aminoácidos contribuindo ao epitopo descontínuo foram identificados, foram selecionados na literatura e extraídos do PDB (Berman et al. 2000). Uma análise detalhada foi primeiro realizada na lisozima de ovo de galinha (*Hen Egg Lysozyme*, HEL) já que vários epitopos bem definidos foram identificados por cristalografia de raios X (resoluções de 1.5 até 3.22 Å). O algoritmo do PEPOP foi rodado na estrutura 3D do HEL (disponível sob o código 1SFG no PDB). Sete complexos adicionais foram também analisados para estender essas observações a um conjunto de 13 epitopos cristalográficos. Análises do PEPOP foram efetuadas nas estruturas 3D do Ag sozinho, i.e. não complexado com o mAc específico do epitopo estudado, porque este é o caso mais frequentemente encontrado. A distribuição dos resíduos formando o epitopo dentro dos *clusters* identificados por PEPOP foi analisada, e o número de resíduos do epitopo no maior *cluster* (i.e. aquele que contém mais resíduos do epitopo) foi calculado.

A validação experimental em animais do potencial antigênico e imunogênico dos peptídeos preditos por PEPOP foi também realizada, e o desempenho do PEPOP foi comparado ao de outras ferramentas de predição de epitopos. A descrição dos métodos pode ser consultada no artigo publicado anexado (Moreau et al. 2008).

III.4. Análise da superfície antigênica das toxinas

Num primeiro tempo, os modelos estruturais das toxinas PnTx2-5 e PnTx2-6 foram submetidas à análise por PEPOP, via *website*. O segmento de referência foi selecionado automaticamente, de acordo com os critérios descritos previamente, dentro de todos os segmentos acessíveis calculados. O peptídeo foi estendido usando o método ONN com a lista dos segmentos pertencendo ao mesmo *cluster* que o segmento de referência. O comprimento da sequência do peptídeo foi fixada a 20 resíduos de aminoácidos.

Num segundo tempo, todas as 461 estruturas 3D de toxinas aracnídicas, provendo ou do PDB ou das predições por SiliMod e estocadas no banco SiliTox, foram submetidas à análise automática por SiliPop. As predições de peptídeos foram

realizadas usando três métodos : (i) o melhor segmento segundo os cinco critérios pre-citados usando o método FPS; (ii) o melhor segmento estendido pelo método ONN dentro do seu *cluster*; (iii) o melhor segmento de cada *cluster* estendido pelo método ONN dentro do mesmo *cluster*. As listas de segmentos, as listas de *clusters*, as matrizes de distâncias, as árvores de agrupamento (*clustering*) e todas as sequências do peptídeos antigênicos preditos resultando foram inseridas no SiliTox na tabelas apropriadas. Os parâmetros correspondendo aos métodos de construção dos peptídeos mimetizando as superfícies potencialmente antigênicas foram também armazenados no banco SiliTox.

Todos os peptídeos preditos por SiliPop para todas as estruturas 3D do banco SiliTox foram comparados em termo de similaridade de sequência (usando BlastP e ClustalW). Os alinhamentos múltiplos resultando foram transformados numa matriz de distância usando o programa Prodist do pacote PHYLIP (Retief 2000). A matriz de distância foi então usada para agrupar os peptídeos das 461 estruturas disponíveis de toxinas (estruturas experimentais e teóricas de toxinas de aranhas e escorpiões). Depois disso, o grafo de saída foi transformado num dendrograma ou numa árvore sem raíz.

IV. Implementação das ferramentas

Para permitir um acesso fácil ao banco de dados usando as tecnologia *web*, a interface do SiliTox foi implementada em HTML, PHP (versão 5.2.4-3) e MySQL (mysql-max-5.0.45). Um conjunto de *queries* foi implementado para permitir a recuperação dos dados segundo vários critérios que serão descritos no capítulo “Resultados”. O banco SiliTox está temporariamente localizado no endereço <http://192.168.8.4/~cfleury/silitox/> durante o seu desenvolvimento.

Os algoritmos de SiliMod, PEPOP e SiliPop foram implementados em PHP e Perl. SiliMod e SiliPop foram desenvolvidos num servidor Linux (*Linux Mandriva corporate kernel 2.6.22.19 SMP, i686 Intel(R) Quad Celeron(R) CPU 4x2.53GHz, RAM 4Go, 250Gb HD*), com uma arquitetura de programa do tipo LAMP (*Linux Apache MySQL*

Php), rodando PhpMyAdmin versão 2.11.92 e um servidor *openssh* (4.7 patch1). O servidor é hospedado e suportado pelo serviço ARIT ("*Administration Réseau Informatique et Télécommunication*") do laboratório SysDiag.

PEPOP roda num servidor Linux (servidor Dell PE2250 virtualizado com a distribuição *Mandriva 2007 OS*) com o servidor web Apache versão 2.0. PEPOP é acessível no endereço <http://diagtools.sysdiag.cnrs.fr/PEPOP>.

Os segmentos e clusters gerados por PEPOP podem ser visualizados diretamente na estrutura 3D do antígeno graças ao uso do *plug-in* Chime (<http://www.mdl.com/>). SiliTox usa o *plug-in* Jmol (<http://jmol.sourceforge.net/>) para mostrar as estruturas 3D das toxinas.

RESULTADOS

I. Modelagem da estrutura 3D das toxinas

I.1. PnTx2-5 e PnTx2-6

Os efeitos de PnTx2-5 e PnTx2-6 podem ser responsáveis dos maiores sintomas do envenenamento pela aranha *Phoneutria nigriventer*, que incluem hiperexcitação, salivação, lacrimação e priapismo (Matavel et al. 2009). Ao nível celular, foi mostrado que essas toxinas causam efeitos complexos na cinética do canal de sódio, inibindo sua inativação e mudando a ativação voltagem-dependente para potenciais mais negativos (Araujo et al. 1993; Matavel et al. 2002), efeitos que correspondem às ações das toxinas α e β , respetivamente. As duas toxinas PnTx2-5 e PnTx2-6 são similares nas suas sequências primárias, já que elas diferem por somente 5 resíduos de aminoácidos: Thr/Pro9; Val/Glu12; Asn/Tyr35; Leu/Trp37; Ser/Asn45 (Figura 10A). Os dados produzidos por nosso grupo mostram que elas têm diferenças significativas nas suas interações com o canal: PnTx2-6 tem uma afinidade 6 vezes mais alta que a PnTx2-5, seus efeitos não são reversíveis após 10-15 minutos de lavagem. PnTx2-6 compete parcialmente (59%) com a α -toxina AaH II de escorpião, mas não com a β -toxina CssIV de escorpião, sugerindo um modo de ação similar aquele das toxinas de sítio 3. No entanto, PnTx2-6 não é removida por pulsos fortemente depolarizantes, ao contrário das toxinas de sítio 3 conhecidas (Matavel et al. 2009).

Investigações das suas estruturas 3D poderiam fornecer idéias sobre seus mecanismos de ação, explicar suas diferenças a respeito das suas afinidades pelos canais Na_v e clarificar porque PnTx2-6 é capaz competir com a α -toxin AaH II.

I.2. Estruturas teóricas de PnTx2-5 e PnTx2-6

Segundo os bancos SCOP, CATH e DBAli, todas as toxinas com 8 resíduos cisteína apresentam o mesmo fold, incluindo o motivo estrutural chamado “*Inhibitory Cystine Knot*” (ICK) (Craik et al. 2001; Gelly et al. 2004). As estruturas das toxinas δ -

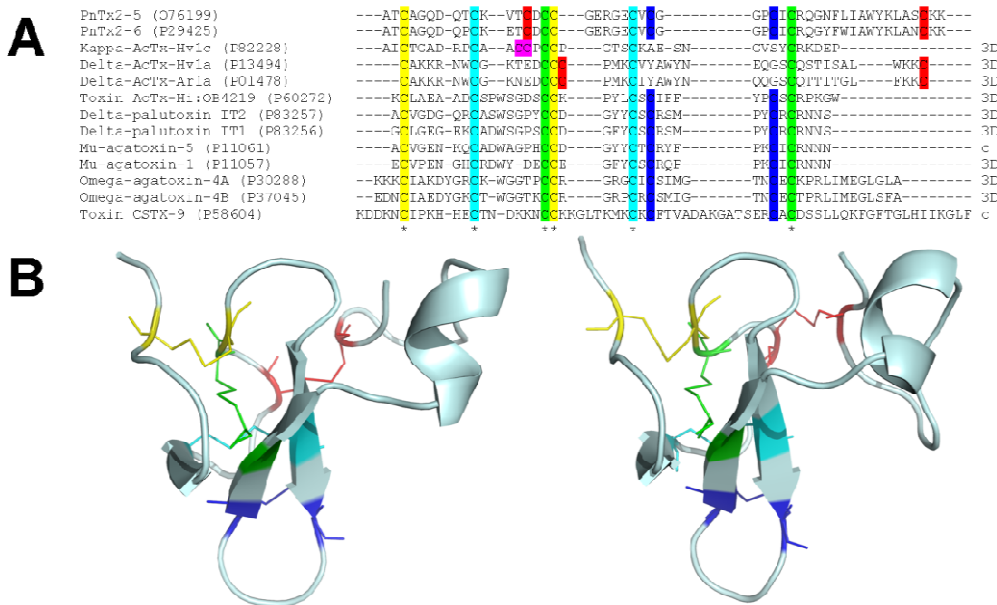


Figura 10. Modelagem molecular de PnTx2-5 e Pn2x2-6. (A) Sequências e padrões de conectividade dos resíduos de cisteínas das toxinas de aranhas ricas em cisteínas. As sequências de PnTx2-5 e PnTx2-6 foram alinhadas com aquelas de todas as toxinas peptídicas ricas em cisteínas contendo pelo menos 8 resíduos de cisteína e para quais as pontes dissulfeto tenham sido identificadas experimentalmente, por resolução da estrutura 3D (κ -atratoxin-Hv1c, δ -atratoxin-Hv1a (versutoxin), δ -atratoxin-Ar1a (robustoxin), AcTx-Hi:OB4219, δ -palutoxins IT2, δ -palutoxin IT1, μ -agatoxin-1, ω -agatoxin-4A e ω -agatoxin-4B) ou por outros métodos químicos (μ -agatoxin-5 e CSTX-9). Os resíduos de cisteína que formam pontes dissulfeto juntas são evidenciadas com a mesma cor. (B) Representação *cartoon* dos modelos tridimensionais teóricos da PnTx2-5 (esquerda) e da PnTx2-6 (direita). As pontes dissulfeto Cys3-Cys17, Cys10-Cys23, Cys14-Cys46, Cys16-Cys31 e Cys25-Cys29 são mostradas em amarelo, turquesa, vermelho, verde e azul escuro, respectivamente. As extremidades amino- e carboxi-terminal são localizadas respectivamente à esquerda e à direita de cada molécula.

atractoxin-Hv1 (1VTX) (Fletcher et al. 1997), δ -palutoxin IT2 (1V91) (Ferrat et al. 2005), toxin AcTx-Hi:OB4219 (1EIT) (Omečinsky et al. 1996) e μ -agatoxin-1 (1KQH) (Rosengren et al. 2002) foram selecionadas como estruturas *template*.

I.2.A. Descrição das estruturas

A figura 10A mostra o alinhamento das sequências de PnTx2-5 e PnTx2-6 com aquelas de 11 toxinas ricas em cisteínas cujas conectividades tenham sido determinadas experimentalmente. O alinhamento mostra que podemos deduzir que as toxinas da *Phoneutria* contêm um motivo ICK conservado. Uma quarta ponte dissulfeto (azul escuro) foi deduzida por similaridade e induz uma hipótese razoável para o quinto pareamento (vermelho). Pareamentos alternativos não mostram nenhuma semelhança com as conectividades de cisteínas conhecidas em toxinas. A figura 10B mostra uma representação *cartoon* dos modelos moleculares das PnTx2-5 e PnTx2-6, com as pontes dissulfeto ressaltadas. Os modelos predizem que cada uma das estruturas de PnTx2-5 e PnTx2-6 contém uma curta folha β formada pelas 3 fitas Gln8–Cys10, Gly21–Cys25 e Cys29–Gln33, e uma curta hélice α formada pelos resíduos 35 (Asn na Tx2-5, Tyr na Tx2-6) até 39 (Ala nas duas toxinas).

I.2.B. Avaliação dos modelos

As estruturas teóricas de PnTx2-5 e PnTx2-6 foram avaliadas por ProSA. Os gráficos de energia de PnTx2-5 e PnTx2-6 são similares áqueles das estruturas conhecidas (Figura 11). Os *scores* de avaliação de PnTx2-5 e PnTx2-6 por Verify3D são classificados como “relativamente bons” e são similares áqueles das toxinas usadas como *templates* (não mostrado). Apesar do fato que algumas características estruturais incomuns podem ser encontradas nas estruturas *templates* e modelos, as análises por WHAT IF informam que os modelos não apresentam erros e têm as propriedades de estruturas corretas (não mostrado). A avaliação da qualidade geométrica das estruturas por Procheck é mostrada na Tabela 3. Os resultados da análise segundo Ramachandran mostram que nenhum resíduo de aminoácido foi encontrado numa região não autorizada, e que a maioria dos resíduos não-terminais

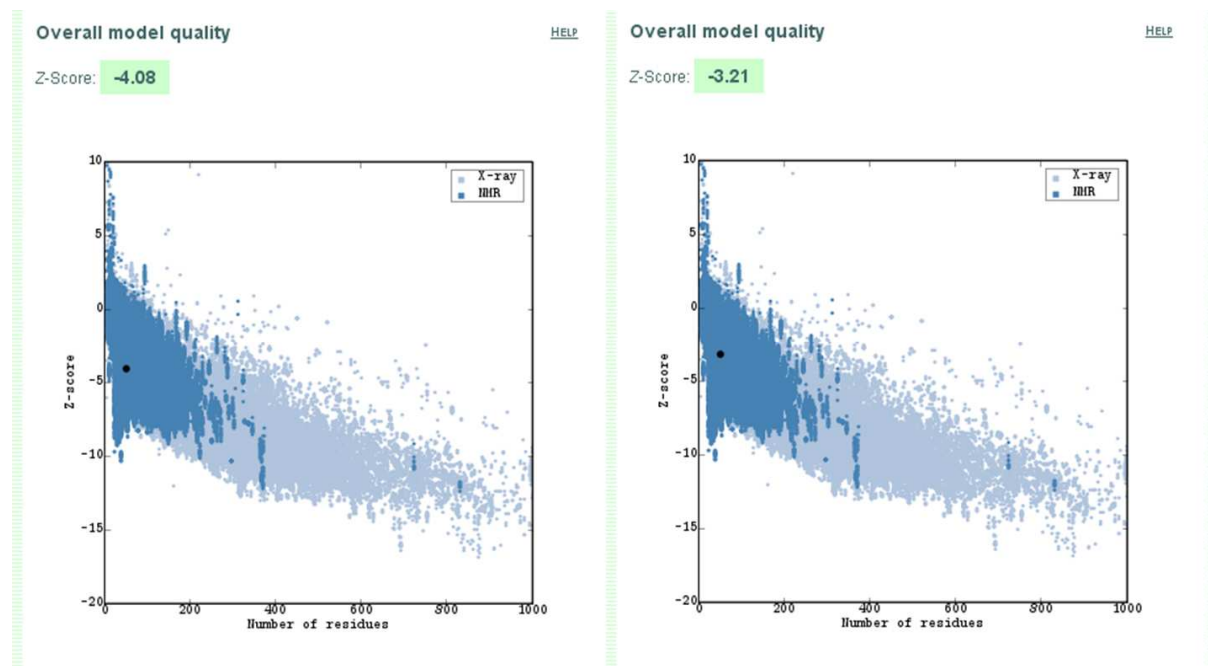


Figura 11. Resultados da avaliação das estruturas teóricas de PnTx2-5 (esquerda) e PnTx2-6 (direita) por ProSA-web. A qualidade dos dois modelos é comparável àquela de estruturas experimentais resolvidas por RMN.

Tabela 3. Resultados estatísticos da avaliação pelo diagrama de Ramachandran das estruturas dos *templates* e dos modelos. A tabela mostra para cada estrutura o número e a percentagem de resíduos de aminoácidos em cada região diedrica do diagrama de Ramachandran (excluindo glicina, prolina e os resíduos terminais). Os dados foram obtidos usando o programa Procheck.

Structures	1VTX Residues (%)	1V91 Residues (%)	1EIT Residues (%)	1KQH Residues (%)	PnTx2-5 Residues (%)	PnTx2-6 Residues (%)
Most favored regions	24 (64.9)	15 (51.7)	18 (62.1)	24 (80.0)	30 (74.4)	32 (84.2)
Additional allowed regions	9 (24.3)	13 (44.8)	10 (34.5)	6 (20.0)	7 (20.5)	4 (10.5)
Generously allowed regions	3 (8.1)	0 (0.0)	1 (3.4)	0 (0.0)	2 (5.1)	2 (5.3)
Disallowed regions	1 (2.7)	1 (3.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Non-Gly non-Pro non-terminal residues	37 (100.0)	29 (100.0)	29 (100.0)	30 (100.0)	39 (100.0)	38 (100.0)

não-Gly e não-Pro (74%) ocupa a região mais favorável (Tabela 3). Notamos que a qualidade geométrica das estruturas modelos é até melhor que a maioria das estruturas *templates*, exceto 1KQH. Os programas de predição de estruturas secundárias produzem resultados compatíveis com os dois modelos. Esses resultados sugerem que os modelos teóricos propostos podem ser considerados como umas das estruturas prováveis das toxinas PnTx2-5 e PnTx2-6 em solução. Em contraste com as α -toxinas de escorpiões, nossos modelos predizem um conteúdo muito baixo em α -hélices. Essa predição está em acordo com as análises de DC da PnTx2-6, como mostra a figura 12A. O espectro observado suporta o modelo proposto. De fato, a proporção de α -hélices estimada pela fórmula usando a elipticidade à 222 nm (Chen et al. 1972) é 7.7% (correspondendo a 3.7 resíduos de aminoácidos), e corresponde bem ao modelo proposto. Para comparar, mostramos também o espectro DC experimental de uma toxina de estrutura similar, que foi determinada experimentalmente (ω -atractoxin Hv1a PDB code 1AXH) (Tedford et al. 2004). Além disso, nossos modelos predizem que os resíduos de triptofano são expostos na superfície das toxinas. Isso é confirmado pelo experimento de fluorescência mostrado na figura 12B. Além de serem coerentes, esses modelos predizem características que podem ser verificadas experimentalmente, como discutido.

I.3. Grupos estruturais

A análise das 630 sequências de toxinas de aracnídeos presentes em SiliTox permitiu identificar 279 motivos de cisteínas diferentes. Segundo as suas similaridades e as conectividades de cisteínas correspondentes nas sequências, esses 279 motivos foram divididos em 30 grupos. Dentro desses grupos, 20 contêm toxinas com uma estrutura 3D conhecida no PDB e também toxinas cujas estruturas 3D não são conhecidas. As toxinas que pertencem a esses grupos foram listadas e alinhadas, e a Figura 13 mostra os grupos 'Kozclass1403', 'Kozclass1502', '8C_ICK2.10.2', '6C_N3.x.x.5.4' como exemplos (para mais clareza, os outros grupos foram anexados no final desse documento). Os grupos são chamados segundo seu conteúdo: 'Kozclass1.1.0.2' significa que o grupo contém toxinas já classificadas por Kozlov e Grishin, no qual

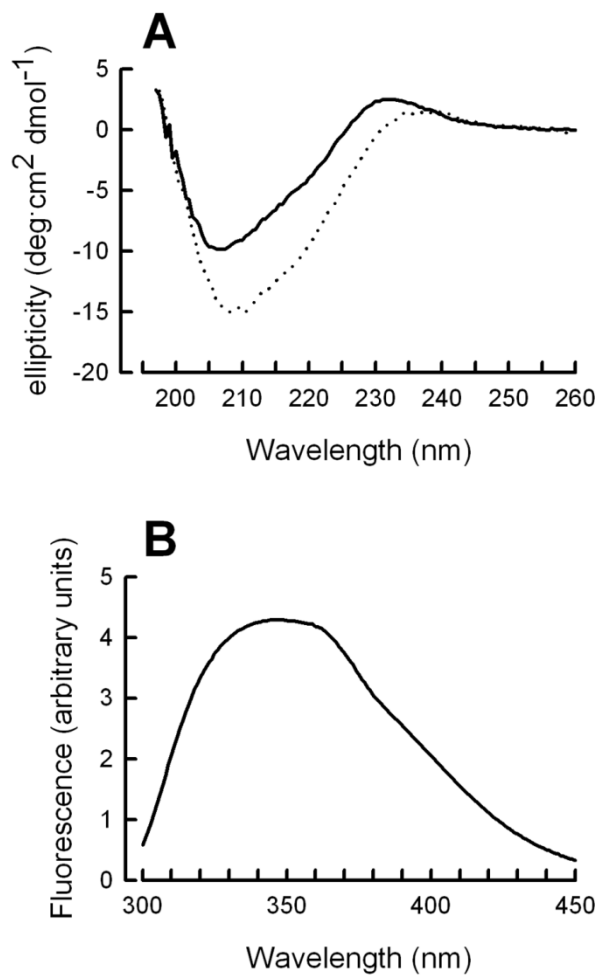


Figura 12. Testes experimentais do modelo de PnTx2-6 proposto. **(A)** O espectro de dicroísmo circular da PnTx2-6 (traço contínuo) é comparado ao espectro da ω-atracotoxin Hv1a (traço discontinuo). Em **(B)**, o espectro de fluorescência a 346 nm, de acordo com os resíduos de triptofano sendo expostos ao médio hidrofílico.

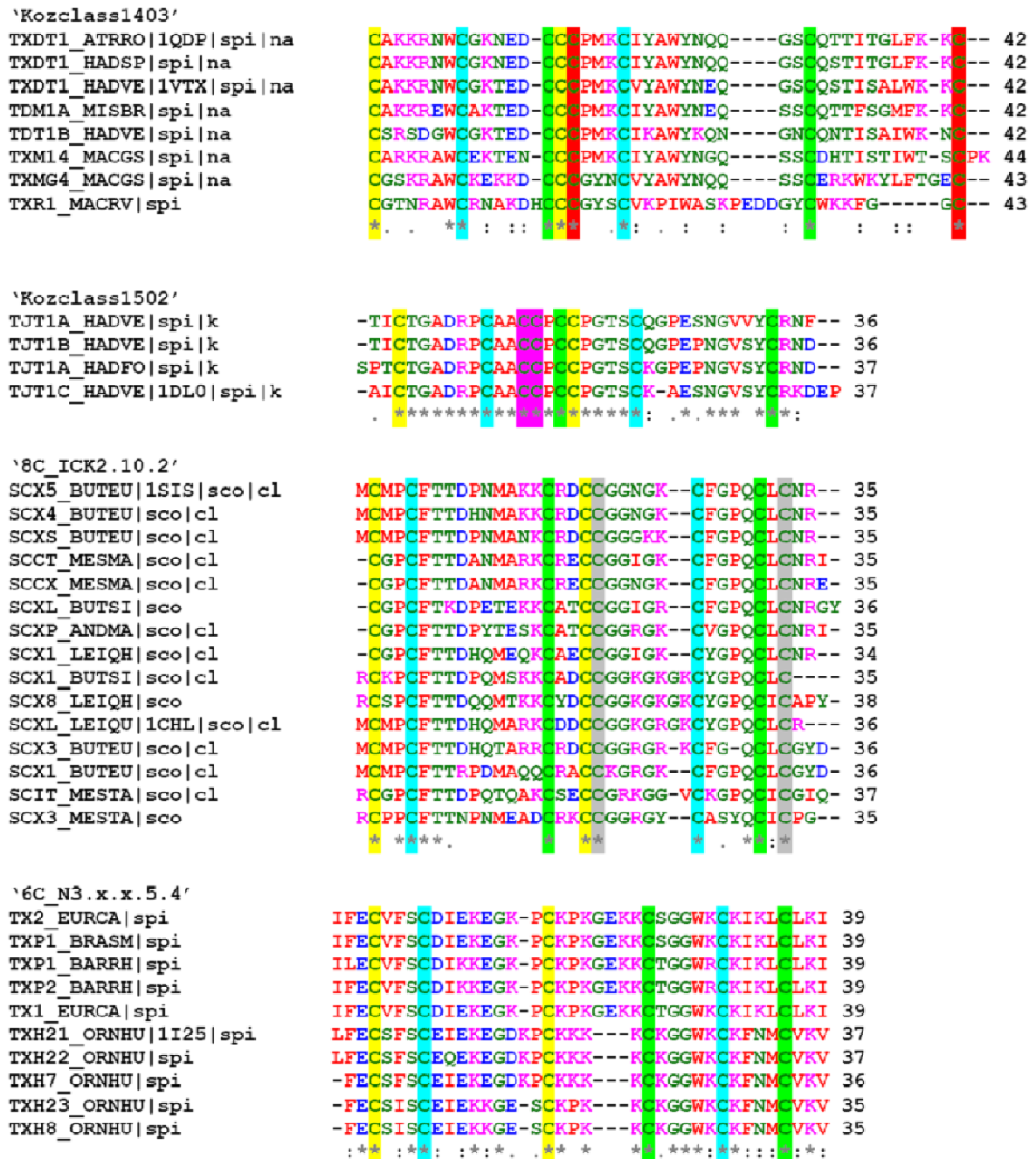


Figura 13. Alinhamento múltiplo das sequências das toxinas pertencendo aos grupos estruturais “Kozclass1403”, “Kozclass1502”, “8C_ICK2.10.2”, “6C_N3.x.x.5.4” baseados nas similaridades dos padrões de conectividade das cisteínas. Dentro de uma determinada sequência, os resíduos de cisteína evidenciados na mesma cor formam juntos uma ponte disulfeto.

caso, a nomenclatura foi respeitada (Kozlov and Grishin 2005); '6C_Nx.3.x.4.1' significa toxinas contendo 6 resíduos de cisteína, nenhum motivo ICK e uma distribuição das cisteínas que segue o motivo PROSITE 'C-N(x)-C-N(3)-C-N(x)-C-N(4)-C-N(1)-C'; '8C_ICK2.10.2' é o grupo composto pelas toxinas contendo 8 resíduos de cisteína, envolvendo um motivo ICK nas suas conectividades. No caso dos grupos Kozclass1.1.0.2, 6C_Nx.3.x.4.1 e 8C_N3.5.3, mais de 5 estruturas *templates* eram disponíveis para a modelagem das outras toxinas pertencendo aos grupos. Então, esses grupos foram subdivididos segundo a similaridade das sequências, usando o programa ClustalW de alinhamento para produzir árvores segundo o método Neighbor-Joining e formar grupos menores como mostrado na Figura 14 com o exemplo do grupo 6C_Nx.3.x.4.1. Isso significa que a classificação final das toxinas pertencendo a tais grupos estruturais subdivididos ocorre em vários níveis. SiliMod foi concebido para selecionar *templates* disponíveis no nível mais baixo primeiro (i.e., dentro do menor subgrupo), e se nenhuma estrutura conhecida pertence ao subgrupo, o algoritmo SiliMod repete a seleção de *templates* no próximo nível, e assim por diante até que ele encontra *templates* disponíveis. Assim, as toxinas são modeladas usando como *templates* toxinas relacionadas em termo de sequência se tais toxinas são disponíveis, e *templates* mais distantemente relacionadas no caso contrário.

Os 20 grupos resultantes e seus possíveis subgrupos foram usados como ponto inicial da seleção de estruturas *templates* para a modelagem comparativa por SiliMod.

I.4. Resultados SiliMod

I.4.A. Modelos tridimensionais

O programa SiliMod gerou modelos 3D para 350 toxinas de escorpiões e aranhas. Junto às 111 estruturas que já tinham sido determinadas experimentalmente, um total de 461 estruturas tridimensionais de toxinas de aracnídeos são agora disponíveis e estocadas no banco de dados SiliTox.

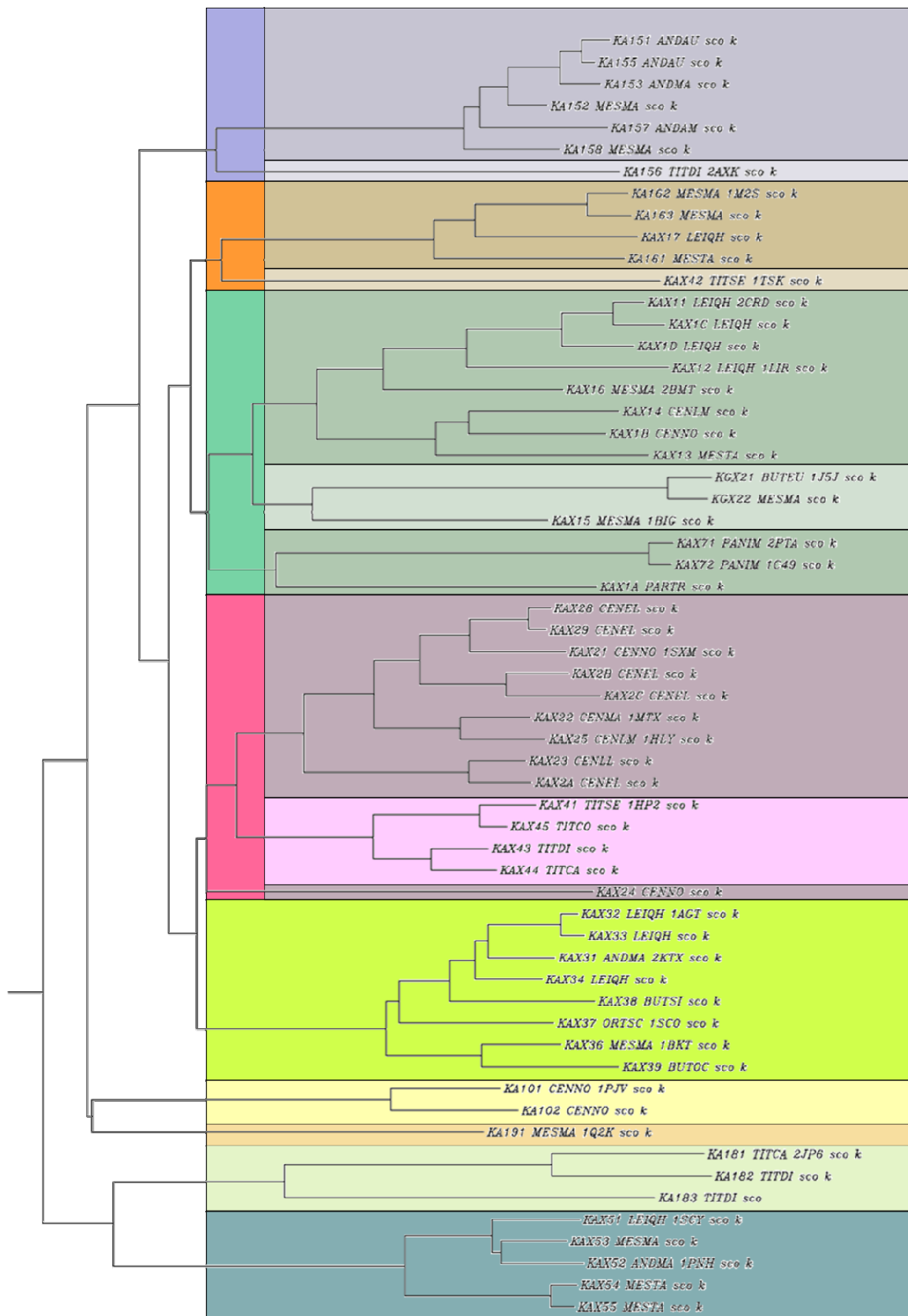


Figura 14. Árvore de Neighbor-Joining das sequências de toxinas aracnídicas membros do grupo de motivo cisteína 6C_Nx.3.x.4.1, calculado por ClustalW. O comprimento dos ramos é relacionado com a distância entre as folhas. O sub-clustering foi feito para dividir o grupo em sub-grupos que contêm cada um menos de 5 estruturas *templates*. Em frente ao nome de cada toxina, “sco” significa “scorpion” e “spi” significa “spider”, e o código PDB é dado quando disponível. Do mesmo jeito, “na”, “k”, “ca”, e “cl” significam “sódio”, “potássio”, “cálcio” e “cloreto” respectivamente, e indicam os canais iônicos conhecidos por serem alvos da toxina.

Tabela 4. Resultados da comparação entre os modelos teóricos produzidos pelo SiliMod e as estruturas experimentais (ou teóricas*) correspondentes no PDB. O alinhamento estrutural foi obtido pelo programa MAMMOTH.

model	PDB	RMSD (Å)	-ln(P-value)	model	PDB	RMSD (Å)	-ln(P-value)
IPTXA_PANIM	1IE6	3,74	4,41	SCX3_CENSC	2SN3	0,46	9,91
JZT11_CHIJI	2A2V	1,35	4,84	SCX3_LEIQU	1LQQ	1,59	9,77
JZTX3_CHIJI	2I1T	3,24	4,48	SCX4_MESMA	1SN4	0,7	9,77
JZTX7_CHIJI	1ZJQ	2,26	4,84	SCX5_BUTEU	1SIS	1,69	4,78
KA101_CENNO	1PJV	1,53	4,47	SCX7_MESMA	1KV0	1,75	10,06
KA156_TITDI	2AXK	1,45	5,39	SCX8_MESMA	1SNB	0,95	9,77
KA162_MESMA	1M2S	1,51	5,21	SCXA_LEIQH	2ASC	2,25	9,65
KA181_TITCA	2JP6	2,86	5,03	SCXC_SCOMA	1C6W	3,65	4,41
KA191_MESMA	1Q2K	1,52	4,28	SCXI_CENSC	2B3C	1,47	9,77
KAX11_LEIQH	2CRD	1,39	5,21	SCXL_LEIQU	1CHL	1,75	4,97
KAX12_LEIQH	1LIR	1,57	5,21	SIX1_MESMA	1WWN	1,87	10,50
KAX15_MESMA	1BIG	1,55	5,21	SIX2_LEIQH	2I61	2,65	9,32
KAX16_MESMA	2BMT	1,24	5,21	SIXP_MESMA	1T0Z	1,87	10,33
KAX21_CENNO	1SXM	1,01	5,75	TOG4A_AGEAP	1IVA	2,27	5,27
KAX22_CENMA	1MTX	1,45	5,75	TOG4B_AGEAP	1AGG	2,31	5,27
KAX25_CENLM	1HLY	1,62	5,75	TOT1A_HADVE	1AXH	3,79	0,64
KAX31_ANDMA	2KTX	1,51	5,57	TOT1F_HADVE	1ZJV	0,59*	5,39
KAX32_LEIQH	1AGT	1,61	5,57	TOT2A_HADVE	1G9P	2,6	3,82
KAX36_MESMA	1BKT	1,83	5,39	TX1_SCOGR	1LA4	1,33	4,84
KAX37_ORTSC	1SCO	1,57	5,57	TXAG_AGEOP	2E2S	3,31	5,03
KAX41_TITSE	1HP2	1,4	5,39	TXDP1_PARLU	1V90	3,23	5,15
KAX42_TITSE	1TSK	1,45	5,03	TXDP2_PARLU	1V91	2,43	4,42
KAX51_LEIQH	1SCY	1,35	4,28	TXDT1_ATRRO	1QDP	2,43	6,27
KAX52_ANDMA	1PNH	1,5	4,28	TXDT1_HADVE	1VTX	2,53	6,27
KAX54_MESTA	2CF1*	1,34	4,28	TXFK1_PSACA	1X5V	3,29	1,96
KAX71_PANIM	2PTA	1,02	4,54	TXH1_ORNHU	1QK6	1,23	4,66
KAX72_PANIM	1C49	1,14	5,03	TXH10_ORNHU	1Y29	2,78	3,70
KAX81_ANDMA	1ACW	3,3	3,89	TXH4_ORNHU	1MB6	1,16	5,03
KAX82_MESMA	1WM7	3,18	3,63	TXHA1_SELHA	1NIX	1,11	4,66
KAX91_MESMA	1DU9	2,59	3,43	TXHA3_SELHA	2JTB	0,97	4,66
KAX92_MESMA	1WM8	2,51	3,43	TXHA4_SELHA	1NIY	1,12	5,03
KGX21_BUTEU	1J5J	1,48	5,21	TXHA5_SELHA	1ZJU	0,78*	5,03
KURT_PARTR	1T1T	3,16	9,62	TXHN1_GRARO	1D1H	1,5	5,03
MTX2_GRARO	1LUP	2,46	4,28	TXHP2_HETVE	1EMX	2,79	4,09
MTX4_GRARO	1TYK	2,45	4,84	TXMG1_AGEAP	1EIT	2,48	5,21
SC12_MESMA	1OMY	0,81	9,77	TXMG4_AGEAP	1EIU*	2,21	5,39
SCL3_LEIQH	1FH3	2,62	8,99	TXMG5_MACGS	2GX1	2,58	3,63
SCM2_MESMA	1CHZ	0,35	9,77	TXMO_HADIN	1KQH	2,91	5,57
SCX1_CENSC	1VNB	1,41	9,91	TXP1_PARSR	1V7F	2,01	3,89
SCX1_MESMA	1DJT	0,3	9,77	TXP1_PSACA	1LMM	3,17	5,93
SCX1_MESTA	2A7T	0,8	9,77	TXP7_APTSC	1WQB	3,5	3,97
SCX2_CENNO	1CN2	1,86	10,06	VSTX1_GRARO	1S6X	2,08	4,84
SCX2_CENSC	1JZA	0,57	10,06	WGRTX_GRARO	1KOZ	2,48	5,21

I.4.B. Validação do método

Dentro das 111 estruturas 3D de toxinas aracnídicas extraídas do PDB (107 experimentais e 4 teóricas), 86 têm sido classificadas num grupo estrutural baseado no motivo de cisteínas que inclui mais de uma estrutura *template*. Assim, as sequências correspondentes a essas 86 toxinas puderam ser modeladas por SiliMod com um bom grau de confiança. As estruturas teóricas resultantes puderam ser alinhadas com as estruturas experimentais (ou modelos do PDB). Os resultados dos alinhamentos estruturais modelo versus experimento pelo programa MAMMOTH são mostrados na Tabela 4. Os resultados mostram que 100%, 73% e 37% dos pares de estruturas apresentam um valor de RMSD inferior a 4Å, 2.5 Å e 1.5 Å, respectivamente. Do mesmo jeito, o $-\ln(P\text{-value})$ para 74% e 98% do conjunto de pares de estruturas alinhadas obtém *scores* superiores a 4.5 e 2.99 respectivamente.

II. Análise da superfície antigênica das toxinas : *PEPOP* & *SiliPop*

II.1. PEPOP

II.1.A. Saída do programa PEPOP

Uma vez que a estrutura 3D de uma proteína foi submetida através da interface (Figura 15), a análise é dividida e apresentada em 6 etapas : (i) cálculo da acessibilidade dos resíduos da proteína ao solvente (Figura 16A); (ii) determinação dos segmentos sequenciais de resíduos acessíveis e das suas respectivas coordenadas (como descrito no capítulo “Material & Método”) (Figura 16B); (iii) cálculo das distâncias espaciais entre os segmentos e determinação das menores distâncias (Figura 16C); (iv) agrupamento dos segmentos segundo as distâncias mínimas (Figura 16D); (v) visualização 3D dos segmentos na superfície da proteína (usando o *applet* Chime) (Figuras 16E e 16F); (iv) desenho de peptídeos (Figuras 16F e 17). Ao clicar no *link* “*peptide design*”, o usuário acessa uma nova página na qual ele pode seleccionar os parâmetros de desenho do peptídeo da sua escolha. A Figura 17 mostra as diferentes etapas do desenho. No primeiro passo, o usuário selecciona o segmento

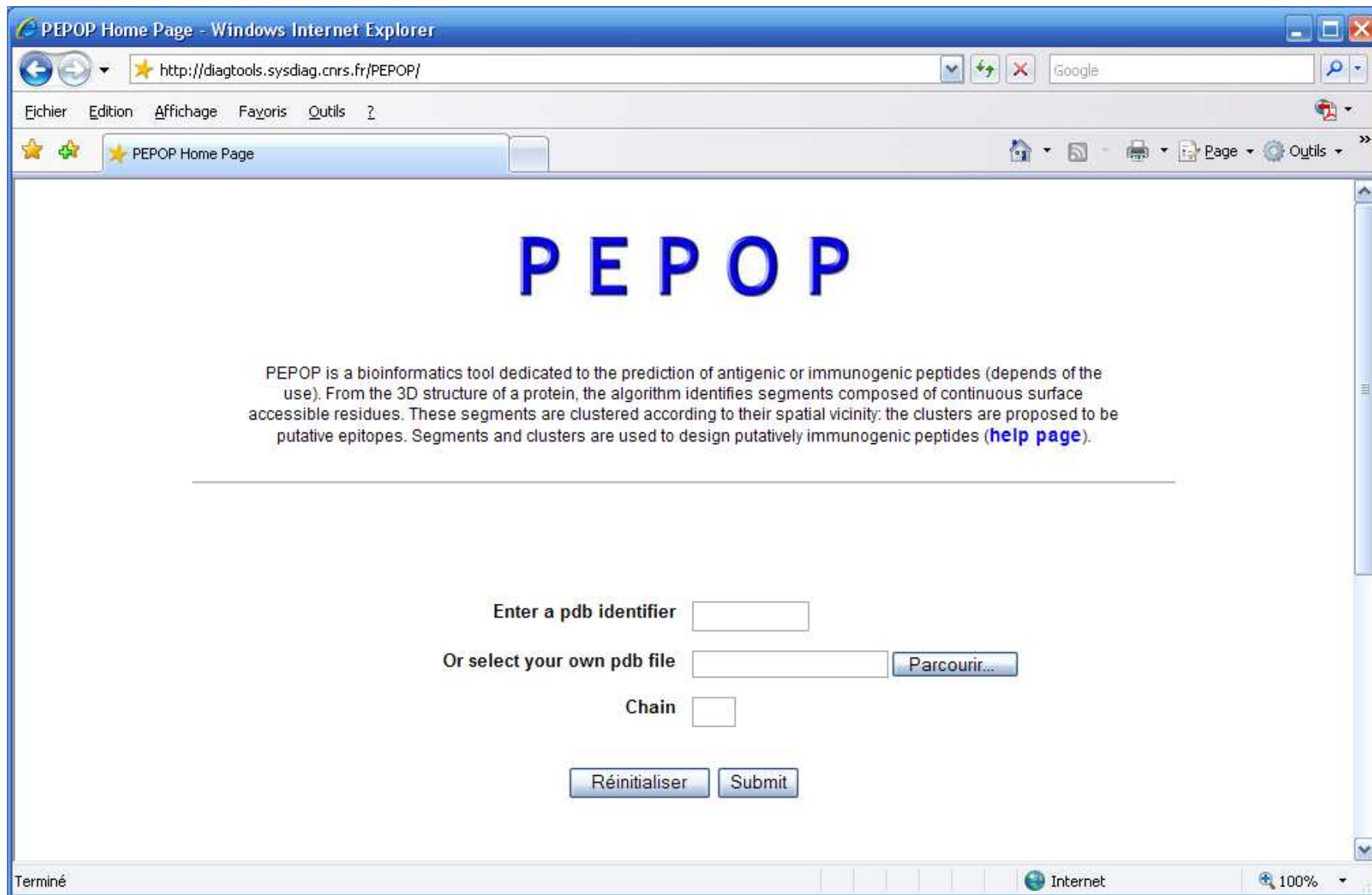


Figura 15. Página principal do PEPOP. O usuário pode submeter seja um código PDB ou uma estrutura protéica original a ser analisada por PEPOP. O programa está disponível ao endereço : <http://diagtools.sysdiag.cnrs.fr/PEPOP/>.

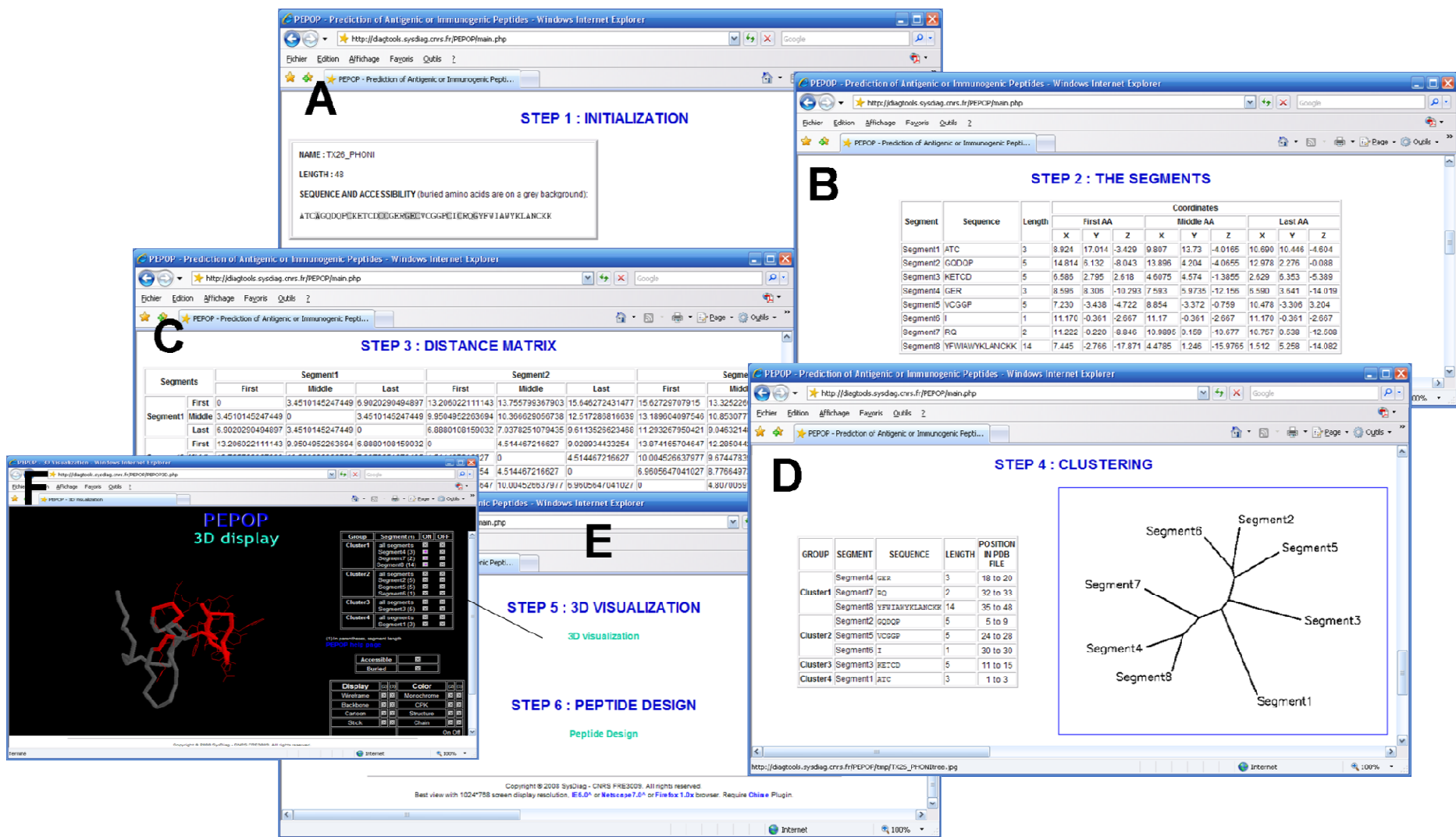


Figura 16. Saída da análise da toxina de aranha PnTx2-6 por PEPPOP.

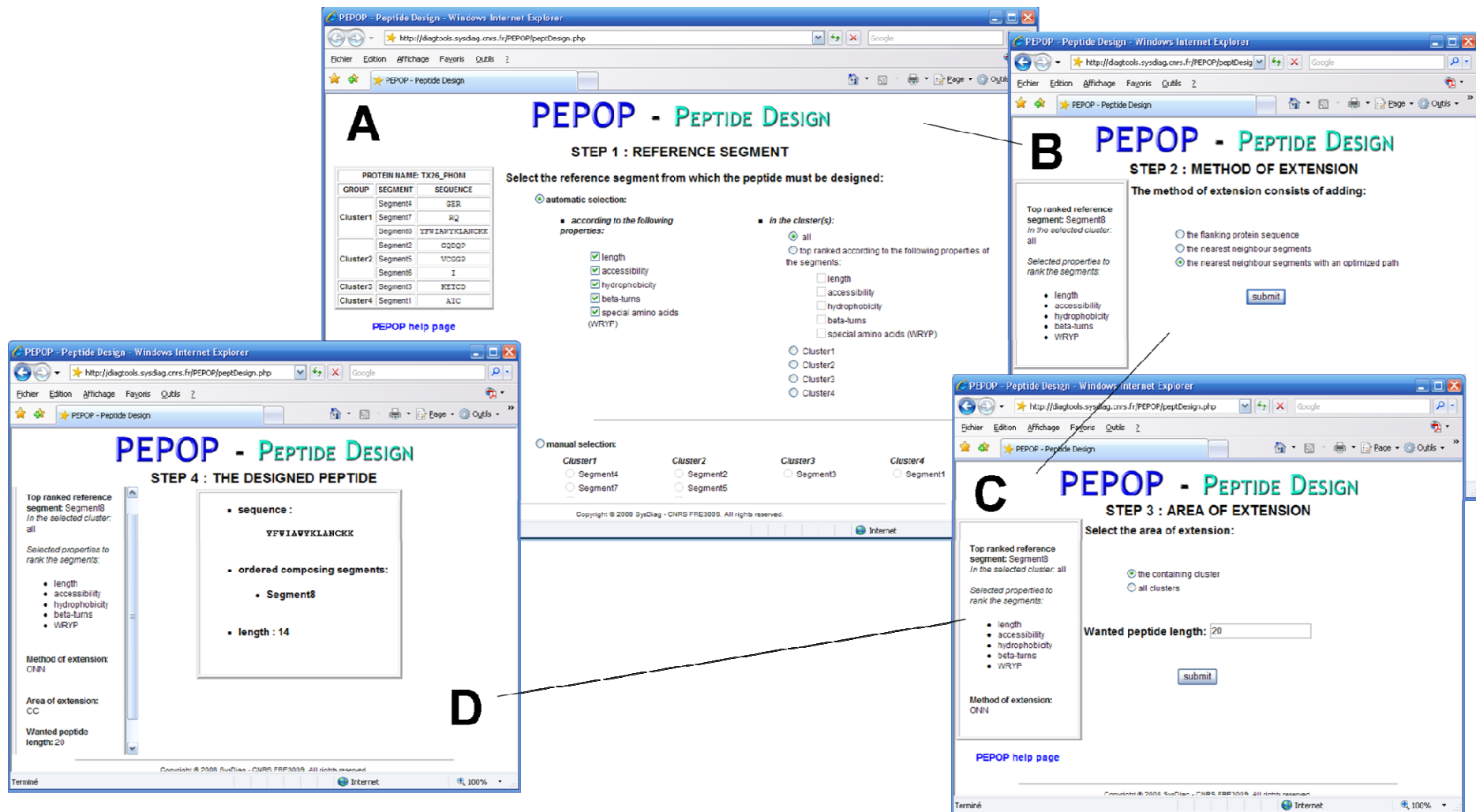


Figura 17. Desenho de peptídeo antigênico por PEPOP. Em (A), ao usuário é pedido escolher o segmento de referência. O modo automático permite selecionar o segmento segundo diversos critérios (veja o texto para os detalhes), dentro do conjunto total de segmentos ou dentro de um *cluster*, que também pode ser classificado segundo os mesmos critérios, ou pode ser selecionado arbitrariamente pelo usuário. Em (B), o método de extensão do segmento tem que ser selecionado (método “FPS”, “NN” ou “ONN”, veja o texto). Em (C), a região de extensão pode ser o *cluster* que contém o segmento de referência ou o conjunto inteiro de segmentos, e o comprimento final do peptídeo tem que ser especificado. Enfim, o peptídeo predito é retornado com as suas características principais, em (D).

de referência (Figura 17A). A seleção pode ser automática, no qual caso o usuário decide quais propriedades têm de ser consideradas pelo programa para classificar e selecionar o melhor segmento, e quais grupos de segmentos têm de ser classificados. A seleção do segmento de referência pode também ser feita manualmente (Figura 17A). Num segundo passo, o usuário tem que selecionar o método de extensão a ser usado sobre o segmento de referência (Figura 17B). Pode-se escolher entre os métodos FPS, NN e ONN, clicando no botão correspondente. O último passo é a seleção da área de extensão, que pode ser o conjunto de todos os segmentos, ou o grupo contendo o segmento de referência (Figura 17C). No caso do método FPS, a área de extensão é a sequência que flanqueia o segmento de referência. O comprimento da sequência do peptídeo final também tem que ser especificado. Em função das restrições impostas, PEPOP vai retornar um peptídeo que satisfaz o mais possível as características selecionadas e um curto resumo do procedimento de desenho do peptídeo será apresentado (Figura 17D).

II.1.B Desempenho do PEPOP sobre epitopos conhecidos

Para avaliar o desempenho do PEPOP, primeiramente, a estrutura 3D da lisozima HEL foi submetida ao programa, e os resultados foram comparados às informações disponíveis na literatura com respeito aos diferentes epitopos que foram identificados na sua superfície até hoje. Numa segunda etapa, as estruturas 3D dos antígenos pertencendo a 13 complexos Ac/Ag (descritos no capítulo “Material & Método”) foram submetidas individualmente à análise por PEPOP.

A análise por PEPOP da superfície acessível do HEL levou à identificação de 23 segmentos de resíduos acessíveis e os segmentos foram automaticamente divididos em 4 grupos espaciais (Tabela 5). Ao comparar com a literatura, o grupo 1 revelou conter resíduos do antígeno que pertencem a três epitopos (reconhecidos pelos anticorpos D11.15, HyHEL-10, e HyHEL-63). O grupo 2 contém o epitopo inteiro reconhecido pelo anticorpo D1.3 e 9 dos 10 resíduos do epitopo reconhecido por D11.15. Ele inclui também alguns dos resíduos dos epitopos reconhecidos por HyHEL-10 e HyHEL-63. O grupo 3 contém o epitopo inteiro do D44.1 e o epitopo do

Tabela 5. Distribuição dos resíduos do epitopo HEL nos *clusters* identificados por PEPOP. A proporção entre o número de aminoácidos do epitopo incluídos no *cluster* em relação ao número total de aminoácidos do epitopo está indicada. As posições dos aminoácidos do epitopo incluídos no *cluster* estão indicadas entre parênteses. As proporções para o maior *cluster* estão em tipo negrito.

PEPOP cluster	PEPOP segment	Position in PDB file (1SFG chain A)	Monoclonal antibody and corresponding publication						
Cluster1	Segment17	93-94	D1.3 [54]	D11.15 [55]	D44.1 [56]	HyHEL-5 [57]	HyHEL-10 [58]	HyHEL-63 [59]	
	Segment18	96-97							
	Segment19	100-103							
Cluster2	Segment2	10-16	10/10 (13, 14, 19, 21, 22, 24, 117, 119, 121, 125)	9/10 (21, 23, 106, 112, 113, 116, 117, 118, 119)			5/16 (15, 16, 19, 20, 21)	8/23 (13, 14, 15, 16, 18, 19, 20, 21)	
	Segment20	106-109							
	Segment21	111-114							
	Segment22	116-119							
	Segment23	121-129							
	Segment3	18-24							
	Segment4	27-28							
Cluster3	Segment10	56-57			11/11 (41, 45, 46, 47, 49, 53, 67, 68, 70, 81, 84)	HyHEL-5 [57]	13/14 (41, 43, 44, 45, 46, 47, 63, 73, 75, 48, 49, 89 53, 67, 68, 70, 84)	4/16 (63, 73, 75, 89)	6/23 (62, 63, 73, 75, 77, 89)
	Segment11	59							
	Segment12	61-63							
	Segment13	65-68							
	Segment14	70-79							
	Segment15	81-82							
	Segment16	84-90							
	Segment8	41-49							
	Segment9	53							
	Cluster4	Segment1							
Segment5		33-35							
Segment6		37							
Segment7		39							
		inaccessible amino acid residues							
			98				1/14 (69)	2/23 (98, 99)	

Tabela 6. Avaliação do desempenho do PEPOP sobre 13 complexos Ac/Ag determinados experimentalmente. (Veja o texto para as referências.)

Ab - Ag complex and corresponding bibliographic reference	PDB code and chain	Number of epitope-forming residues	PEPOP prediction	
			Number of clusters containing epitopic residues/total cluster number	Number of predicted antigenic residues among epitope residues
D1.3 - hen egg lysozyme	1SFG_A	10	1/4	10
D44.1 - hen egg lysozyme	1T6V_M	11	1/5	11
HyHEL-5 - hen egg lysozyme	1VDP_B	14	1/5	13
Jel42 - HPr	1POH	16	1/4	13
D11.15 - hen egg lysozyme	1HEL	10	2/5	8
5G9 - tissue factor	1WV7_T	18	1/3	14
BH151 - hemagglutinin	5HMG_C	19	3/7	12
Bo2C11 - FVIII C2 domain	1D7P_M	15	3/5	8
NC41 - neuraminidase	1NMC_N	22	3/4	11
F9.13.7 - guinea fowl lysozyme	1HHL	10	3/5	5
HyHEL-10 - hen egg lysozyme	1UC0_A	16	3/5	8
HyHEL-63 - hen egg lysozyme	1VFB_C	23	3/4	9
N10 - staphylococcal nuclease	1EYO_A	18	1/2	6

HyHEL-5 quase inteiro (13 resíduos em 14). Ele contém também parte dos epitopos do HyHEL-10 e HyHEL-63. Com respeito ao grupo 4, ele não corresponde a nenhum epitopo já descrito. Os resultados apresentados na Tabela 6 mostram que os resíduos de aminoácidos identificados experimentalmente para cada epitopo respectivo pertencem a um só grupo no caso de D1.3, D44.1, e HyHEL-5 no HEL e 5G9 no fator de tecido. Eles pertencem a dois grupos no caso de D11.15 no HEL e Je142 na proteína HPr, e são divididos em 3 grupos no caso dos complexos BH151 - hemagglutinin, Bo2C11 - domínio C2 do FVIII, NC41 - neuraminidase, F9.13.7 - guinea fowl lysozyme (GEL), HyHEL-10 - HEL, HyHEL-63 - HEL, e N10 - staphylococcal nuclease (SN).

Uma validação experimental da habilidade do PEPOP em predizer peptídeos imunogênicos em modelos animais e uma comparação do desempenho do PEPOP comparado com outras ferramentas existentes foram realizadas no laboratório SysDiag, e os resultados são descritos no artigo publicado sobre PEPOP (Moreau et al. 2008).

II.2. Predição de peptídeos antigênicos nas toxinas PnTx2-5 e PnTx2-6

Os resultados da análise por PEPOP das estruturas teóricas previamente calculadas para PnTx2-5 e PnTx2-6 são apresentados na Figura 18. Para cada toxina, a lista dos segmentos acessíveis é fornecida (Figura 18A). Os segmentos expostos que compõem o peptídeo antigênico predito são evidenciados em ambas estruturas das toxinas (Figura 18B), e o segmento de referência é mostrado sozinho na Figura 18C. Emfim, a sequência de cada peptídeo antigênico predito é indicada na Figura 18D, junto às suas características principais. É bom notar que os peptídeos preditos para as duas toxinas PnTx2-5 e PnTx2-6 correspondem à mesma região das suas sequências respectivas, ou seja a extremidade C-terminal, que é também a região que envolve as diferenças mais significativas nas suas sequências de aminoácidos (Gln vs. Tyr35, Ile vs. Trp37, Ser vs. Asn45 para PnTx2-5 e PnTx2-6 respectivamente), evidenciadas por um traço vermelho na Figura 18D.

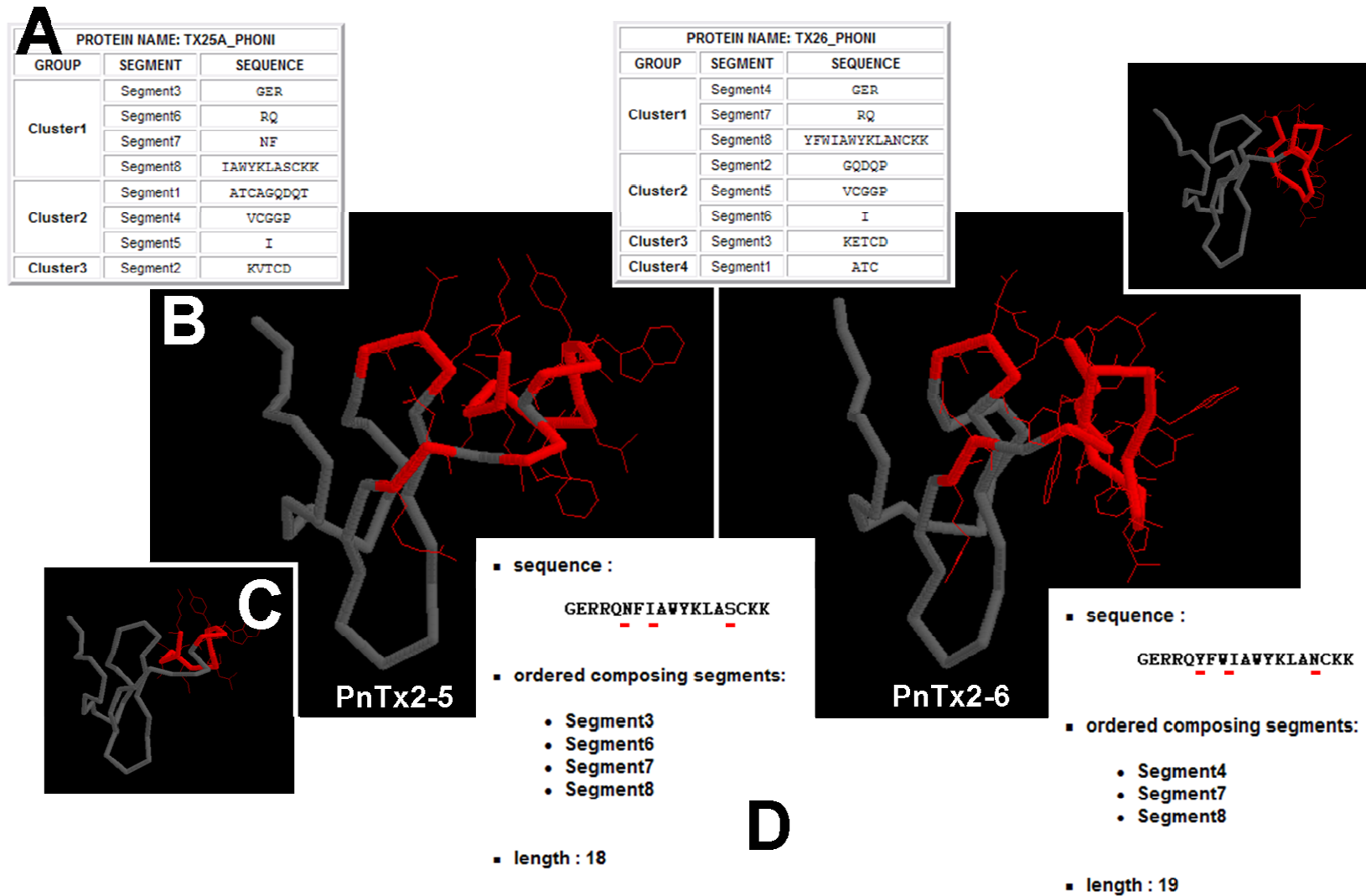


Figura 18. Resultado da análise por PEPOP das estruturas modeladas de PnTx2-5 e PnTx2-6. A lista dos segmentos expostos é mostrada em (A). Os segmentos compoendo o peptídeo final predito são mostrados em vermelho na estrutura inteira de PnTx2-5 (B). O segmento de referência é evidenciado sozinho em (C). A sequência do peptídeo potencialmente antigênico resultante é mostrada em (D), junta à lista dos segmentos que o compõem e ao comprimento do peptídeo construído. Em relação a PnTx2-6, as mesmas características são mostradas no lado direito da figura.

II.3. SiliPop

Para cada uma das 461 estruturas 3D de toxinas aracnídicas estocadas no SiliTox, SiliPop predisse entre 3 e 9 peptídeos antigénicos, lidando a um total de 2863 peptídeos mimetizando regiões de potencial interação proteína-proteína nas toxinas aracnídicas, cobrindo suas superfícies inteiras. As sequências dos peptídeos, os segmentos, as matrizes e os grupos resultando da análise automática por SiliPop são disponíveis no SiliTox e chamados segundo os métodos usados para construir cada peptídeo predito.

III. Banco de dados de toxinas : SiliTox

O banco de dados SiliTox contém todas as toxinas peptídicas ricas em resíduos de cisteínas que provêm de venenos de aranhas e escorpiões e foram identificadas e indexadas em Swiss-Prot até hoje. Os dados estruturais e funcionais disponíveis para essas toxinas foram classificados e organizados no SiliTox segundo a pertinência que eles apresentam em relação às investigações futuras sobre suas relações estrutura-função-antigenicidade. Novas informações que foram geradas pela modelagem 3D (SiliMod) e as análises sistemáticas das superfícies antigénicas das toxinas (SiliPop) foram inseridas no SiliTox de uma maneira que permite que elas sejam facilmente recuperadas através da interface do SiliTox. Essa última foi concebida para facilitar a pesquisa do usuário no banco de dados através de *queries* pré-programadas.

III.1. Estocagem dos dados

Para ilustrar como os dados são estocados nas 47 tabelas do SiliTox, as Figuras 19 e 20 mostram a organização das tabelas *BioMolecule*, *Model*, *Peptide* e *SegInPep*.

III.2. Recuperação dos dados

A Figura 21A mostra a página principal de pesquisa do SiliTox, na qual a expressão “*phoneu*” é sugerida como palavra-chave a ser procurada nas categorias “*molecule name*”, “*species*”, “*Swiss-Prot access number*” e “*PDB code*” dentro dos dados do

The image displays a web browser window showing the BioMolecule database interface. The browser address bar indicates the URL: http://192.168.8.4/phyadmin/index.php?db=toxdb&token=88355e49e72567e05e5d317ed20. The interface is divided into two main sections: a list of molecules and a detailed view of a selected molecule.

Top Section: List of Molecules

molid	type	subtype	iontransport	channelsubtype	family	sequence
1	protein	toxin	NULL	NULL	1	ATCAGQDQPCKECTDCCGERGEVCVGGSPICRQGYFWIAWYKLANCKK
2	protein	channel	sodium	Nav1.2	2	MARSYLVYPPGPDSEFRFFFTRESLAAGIEAIEAKKPKQERKDEDDENGP...
3	protein	toxin	NULL	NULL	1	ATCAGQDQTCVATCCCGERGEVCVGGSPICRQGNFLAWYKLANCKK

Bottom Section: Detailed View of a Molecule

pdb	pdb2	pdb3	3Dmethod	otheracnumb	knottin	2Dstruc	3Dstruc	chain	cyspattern	fold	similarity	toxicity	otheracnumb	knottin	2Dstruc	3Dstruc	chain	cyspattern	fold	similarity	toxicity	comment	accession	cluster		
NULL	NULL	NULL	myrmodel	NULL	yes	NULL	TXQ6_PHONI.pdb	A		160	1	Belongs to the spider toxin Tx7 family.	death at dose levels of 0.75 mg/mouse	del	NULL	yes	NULL	TXQ6_PHONI.pdb	A	150	1	Belongs to the spider toxin Tx7 family.	death at dose levels of 0.79 mg/mouse	Blocks voltage-gated sodium channels. Causes scrut.	TXQ6_PHONI	tree.jpg
NULL	NULL	NULL	myrmodel	NULL	yes	NULL	TXQ5A_PHONI.pdb	A		160	1	Belongs to the spider toxin Tx2 family.	death at dose levels of 0.24 mg/mouse	del	NULL	yes	NULL	TXQ5A_PHONI.pdb	A	150	1	Belongs to the spider toxin Tx2 family.	death at dose levels of 0.24 mg/mouse	Blocks voltage-gated sodium channels. Causes scrut.	TXQ5A_PHONI	tree.jpg
2A2V, NMR, ; A=51-54			NMR	NULL	yes	NULL	2A2V.pdb	A		1	NULL	Belongs to the huwentoxin-1 superfamily.			NULL	yes	NULL	2A2V.pdb	A	1	NULL	Belongs to the huwentoxin-1 superfamily.	Inhibits Kv2.1 (KCNB1), Kv4.1 (KCND1), and Kv4.2 (KCNK2).	2A2V	tree.jpg	
21T, NMR, ; A=28-53			NMR	NULL	yes	NULL	21T.pdb	A		3	NULL	Belongs to the huwentoxin-1 superfamily. Jingzhaot...			NULL	yes	NULL	21T.pdb	A	3	NULL	Belongs to the huwentoxin-1 superfamily. Jingzhaot...	Selectively inhibits activation of voltage-gated s...	21T	tree.jpg	
1ZJQ, NMR, ; A=1-34			NMR	NULL	yes	NULL	1ZJQ.pdb	A		4	NULL	Belongs to the huwentoxin-1 superfamily. Huwentoxi...			NULL	yes	NULL	1ZJQ.pdb	A	4	NULL	Belongs to the huwentoxin-1 superfamily. Huwentoxi...	Inhibits voltage-gated sodium channels (by similar...	1ZJQ	tree.jpg	
1LUP, NMR, ; A=1-31			NMR	NULL	yes	NULL	1LUP.pdb	A		5	NULL	Belongs to the shrtoxotoxin			NULL	yes	NULL	1LUP.pdb	A	5	NULL	Belongs to the shrtoxotoxin	Blocks mechanosensitive ion channels.	1LUP	tree.jpg	

Figura 19. Organização dos dados na tabela *BioMolecule* do banco de dados SiliTox.

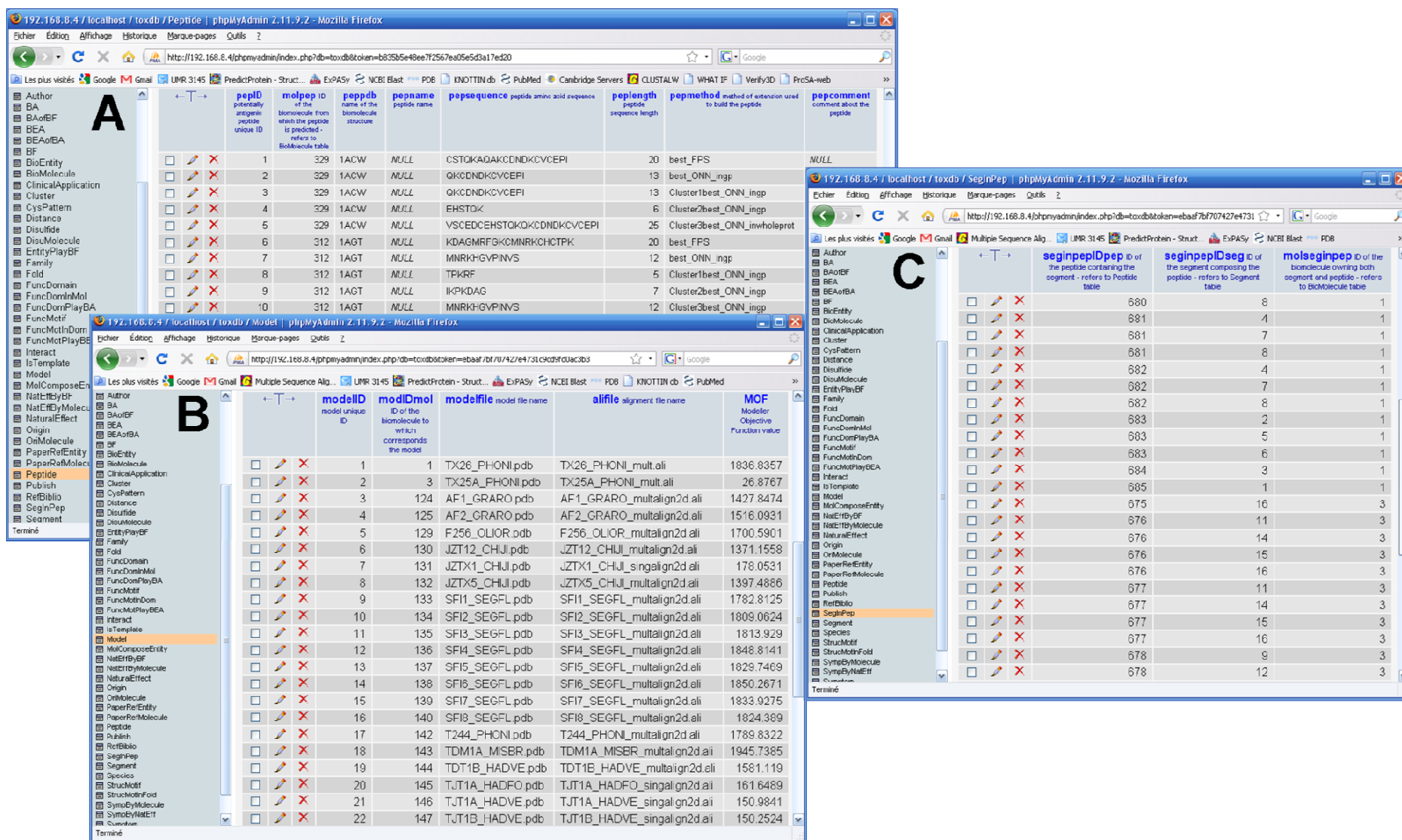


Figura 20. Organização dos dados nas tabelas (A) Model, (B) Peptide, e (C) SegInPep do banco SiliTox.

A

Search

Key-word:

Search by:

- Molecule name
- Species
- Swiss-Prot accession number
- PDB code
- Function description
- Natural effect
- Application

Or search among:

Terminé

B

41 entry(-ies) of SiliTox database match query *phoneu*

BIOMOLECULE GENERIC NAME	SHORT NAME	3D STRUCTURE	SPECIES	CLASS
Neurotoxin PnTx2-6	PnTx2-6	TX26_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin PnTx2-5	PnTx2-5	TX25A_PHONI.pdb	Phoneutria nigriventer	spider
Venom protein PNTx22C3	-	-	Phoneutria nigriventer	spider
Neurotoxin PNTx24A0C4	-	T244_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin PNTx13C3	-	TX13_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin Pn2-1A	-	-	Phoneutria nigriventer	spider
Neurotoxin Tx2-1	-	-	Phoneutria nigriventer	spider
Neurotoxin PNTx22A0C1	-	TX22_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin PNTx24A0C3	-	TX24_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin Tx2-5	-	-	Phoneutria nigriventer	spider
Neurotoxin PNTx27C4	-	TX27_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin Tx2-9	-	TX29_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin Tx3-1	-	TX31_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin Tx3-2	-	TX32_PHONI.pdb	Phoneutria nigriventer	spider
Neurotoxin Pn3-3A	-	TX33A_PHONI.pdb	Phoneutria nigriventer	spider

Terminé

Figura 21. Páginas de pesquisa do banco SiliTox e de resultado de uma procura. **(A)** A expressão “*phoneu*” é inserida no campo “key-word” e é procurada nas categorias tidadas (note que as três opções escritas em cinza estarão funcionais em breve), e então o botão “Find in SiliTox” é acionado. **(B)** Conseqüentemente, a página de resultados é retornada. A neurotoxina PnTx2-6 é selecionada.

SiliTox. É possível selecionar entre uma e todas as categorias para realizar a pesquisa. Ao acionar o botão “*Find in SiliTox*”, a *query* é mandada ao banco de dados e o resultado aparece numa tabela (Figura 21B). A primeira molécula da lista é selecionada acionando o nome dela, e a Figura 22B mostra a descrição individual da toxina Tx2-6 da aranha *Phoneutria nigriventer*. O topo da tabela dá um resumo da identidade da PnTx2-6. Se acionar o nome da espécie, todas as moléculas vindo da mesma espécie serão listadas numa nova aba do navegador *web* (Figura 22B). Do mesmo jeito, é possível recuperar todas as moléculas classificadas na mesma família protéica (Figura 22C). O número de acesso Swiss-Prot da molécula é vinculado à entrada correspondente no banco Swiss-Prot pela abertura de uma nova aba (Figura 22D). Vínculos *web* para as principais publicações envolvendo a molécula selecionada conectam diretamente o usuário ao banco de dados PubMed, e apresentam os resumos correspondentes (Figura 22E).

Na segunda parte da tabela, diversos aspectos da estrutura da molécula selecionada são apresentados. A toxina SF18 da aranha *Segestria florentina* foi escolhida como segundo exemplo para ilustrar a interface do SiliTox. Portanto, (i) sua sequência; (ii) seu número de cisteínas; (iii) seu motivo de cisteínas; (iv) sua classificação de motivo de cisteínas; (v) suas pontes dissulfeto; (vi) sua estrutura 3D e o método que foi usado para obtê-la, são fornecidos (Figura 23). Ao acionar nos itens (ii), (iii) ou (iv), todas as outras moléculas que compartilham as mesmas propriedades selecionadas serão retornadas numa nova aba (Figura 23B, 23C e 23D). Vínculos para o artigo do Kozlov sobre a classificação das toxinas peptídicas de aranha e para o Knottin Website são também propostos como ajudas simbolizadas por um “(?)” (Figura 23E e 23F). Se uma estrutura 3D pode ser modelada pelo SiliMod, os parâmetros da modelagem são apresentados e ambos arquivos de alinhamento e de coordenadas podem ser baixados (Figuras 23G e 23H). A(s) estrutura(s) PDB usada(s) como *template(s)* no procedimento da modelagem é também fornecida e cada uma é vinculada à entrada do SiliTox correspondente. A Figura 24 mostra os dados estruturais no caso da δ -palutoxina IT2 da aranha *Paracoelotes luctuosus*, para qual a estrutura 3D foi determinada experimentalmente e indexada no Protein Data Bank. A entrada PDB correspondente pode ser acessada ao acionar o código PDB

A SiliTox database entry for *Neurotoxin PnTx2-6*

BIOMOLECULE IDENTIFICATION

Complete name: Neurotoxin PnTx2-6
 Short name: PnTx2-6
 Type: protein
 Species: [Phoneutria nigriventer](#)
 Family: [spider toxin Tx2 family](#)
 Origin: venom gland
 Swiss-Prot accession number: [P29425](#) (Q95UF2)
 Bibliographic reference(s): [Penaforte et al., Toxicon, 2000](#)
[Cordeiro et al., FEBS Lett., 1992](#)

STRUCTURE

Mature sequence: ATCAGDQPKETDCCGGERGECVGGGPCICRQGYFWIAWYKLANCKK
 Length: 48
 Cysteine residue number: 10
 Cysteine pattern: X(2)-C-X(6)-C-X(3)-C-X(1)-C-C-X(5)-C-X(1)-C-X(3)-C-X(1)-C-X(14)-C-X(2)

B 25 entry(-ies) of SiliTox database match query species = *Phoneutria nigriventer*

BIOMOLECULE GENERIC NAME	SHORT NAME	SPECIES	FAMILY	CLASS
Neurotoxin PnTx2-6	PnTx2-6	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin PnTx2-5	PnTx2-5	Phoneutria nigriventer	spider toxin Tx2 family	spider
Venom protein PNTx22C3	-	Phoneutria nigriventer	spider toxin Tx3-6 family	spider
Neurotoxin PNTx24A0C4	-	Phoneutria nigriventer	spider neurotoxin 21C2 family	spider
Neurotoxin PNTx13C3	-	Phoneutria nigriventer	spider neurotoxin 21C2 family	spider
Neurotoxin PnTx2-1A	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin Tx2-1	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin PNTx22A0C1	-	Phoneutria nigriventer	protease inhibitor 17 family	spider
Neurotoxin PNTx24A0C3	-	Phoneutria nigriventer	spider neurotoxin 21C2 family	spider
Neurotoxin Tx2-5	-	Phoneutria nigriventer	spider toxin Tx2 family	spider

C 14 entry(-ies) of SiliTox database match query family = spider toxin Tx2 family

BIOMOLECULE GENERIC NAME	SHORT NAME	SPECIES	FAMILY	CLASS
Neurotoxin PnTx2-6	PnTx2-6	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin PnTx2-5	PnTx2-5	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin PnTx2-1A	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin Tx2-1	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin Tx2-5	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin PnTx4-3	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Insecticidal neurotoxin Tx4(5-5)	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Insecticidal neurotoxin Tx4(6-1)	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Toxin Pn4B	-	Phoneutria nigriventer	spider toxin Tx2 family	spider
Neurotoxin ANC45C1	-	Anaxylometes sp.	spider toxin Tx2 family	spider
Neurotoxin PRTx32C1	-	Phoneutria reidyi	spider toxin Tx2 family	spider
Neurotoxin PKTx20C2	-	Phoneutria keyserlingi	spider toxin Tx2 family	spider
Neurotoxin PKTx20C4	-	Phoneutria keyserlingi	spider toxin Tx2 family	spider
Neurotoxin PKTx36C1	-	Phoneutria keyserlingi	spider toxin Tx2 family	spider

D UniProtKB/Swiss-Prot entry P29425

Notice: This page will be replaced with the UniProtKB/Swiss-Prot entry page for P29425.

E The purification and amino acid sequences of four Tx2 neurotoxins from the venom of the Brazilian 'armed' spider *Phoneutria nigriventer* (Keys).

Cordeiro, Mda N, Diniz, CR, Valentim, Ade C, van Eickelst, VS, Gilroe, J, Rickardson, M. *Centro de Pesquisa e Desenvolvimento, Universidade Federal do Rio de Janeiro (COPPE), Brazil.*

Four neurotoxic polypeptides (Tx2-1, Tx2-5, Tx2-6 and Tx2-9) were purified from the venom of the South American 'armed' spider *Phoneutria nigriventer* (Keys) by gel filtration and reverse phase HPLC. These systems rich polypeptides exhibited different levels of neurotoxicity in mice after intracerebroventricular injection. Tx2-1, Tx2-5 and Tx2-6 caused spastic paralysis and death, but the less toxic Tx2-9 produced only tail erection and scratching. The molecular weights of the polypeptides as determined by desorption mass spectrometry were 9038.8 for Tx2-1, 5116.6 (Tx2-5), 5291.3 (Tx2-6) and 3742.1 (Tx2-9). The complete amino acid sequences of the neurotoxins were determined by automated Edman degradation and by manual OABTC-PTC microsequence analysis of peptides obtained after digestion with various enzymes. The amino acid sequences of Tx2-1 (153 residues), Tx2-5 (149 residues) and Tx2-6 (148 residues) were homologous, but had only limited similarities to the less toxic Tx2-9 (132 residues). All four polypeptides had varying sequence identities with other neurotoxins from different spider species and biologically active peptides from scorpions, a sea anemal and seeds of *Mimulus jalapa*.

PMID: 1307265 [PubMed - indexed for MEDLINE]

Figura 22. Entrada do SiliTox para a biomolécula PnTx2-6 previamente selecionada, dados de identificação. (A) Ao acionar nos diferentes vínculos, páginas de resultados se abrem. Elas correspondem respectivamente a : (B) todas as biomoléculas da espécie *Phoneutria nigriventer*; (C) todas as biomoléculas da família Tx2 de toxinas de aranhas; (D) a entrada do banco Swiss-Prot para a biomolécula PnTx2-6; (E) a entrada do PubMed de uma publicação importante relacionada à biomolécula selecionada PnTx2-6.

A

Cysteine residue number: 8

Cysteine pattern: $X(2)-C-X(6)-C-X(7)-C-C-X(2)-C-X(1)-C-X(15)-C-X(4)$

Kozlov class: 2.1.1.0

Sequence position	Relative position	ICK convention	Disulfide bridge quality
C3 - C19	C1 - C4	CI - CIV	my guess
C10 - C22	C2 - C5	CII - CV	my guess
C18 - C42	C3 - C8	CIII - CVI	my guess
C24 - C40	C6 - C7	Ga - Cb	my guess

B

314 entry(-ies) of SiliTox database match query **cysteine number = 8**

BIOMOLECULE GENERIC NAME	SHORT NAME	SPECIES	CYSTEINE NUMBER	CYSTEINE PATTERN
Kappa-actinotoxin IIc	Kappa-ACTN-IIc	Laelaspis scorpions	8	$X(2)-C-X(6)-C-X(7)-C-C-X(2)-C-X(1)-C-X(15)-C-X(4)-C-X(4)$

C

8 entry(-ies) of SiliTox database match query **cysteine pattern = $X(2)-C-X(6)-C-X(7)-C-C-X(2)-C-X(1)-C-X(15)-C-X(1)-C-X(4)$**

BIOMOLECULE GENERIC NAME	SHORT NAME	SPECIES	CYSTEINE NUMBER	CYSTEINE PATTERN
Troxin SFI1	-	Scutigera mucronata	8	$X(2)-C-X(6)-C-X(7)-C-C-X(2)-C-X(1)-C-X(15)-C-X(1)-C-X(4)$

D

34 entry(-ies) of SiliTox database match query **Kozlov class = 2.1.1.0**

BIOMOLECULE GENERIC NAME	SPECIES	KOZLOV CLASS	CYSTEINE NUMBER	CYSTEINE PATTERN
Delta-palutodin II	Paraclosetes brevis	2.1.1.0	8	$X(1)-C-X(6)-C-X(7)-C-C-X(2)-C-X(1)-C-X(15)-C-X(1)-C-X(4)$

E

Classification of spider toxin...
Kozlov S, Grishin I.

F

What are knottins

Knottins are small disulfide-rich proteins characterized by a very special 'disulfide through disulfide knot'

G

theoretical model

Template structure PDB code(s):

- 1V90
- 1E11
- 1K0H

Alignment file: [SFI8_SEGFL_multialign2.d.ali](#)

Model file: [SFI8_SEGFL.pdb](#)

H

Ouverture de SFI8_SEGFL.pdb

Vous avez choisi d'ouvrir SFI8_SEGFL.pdb qui est un fichier de type Brookhaven Protein Databank File à partir de http://192.100.0.4

Que doit faire FireFox avec ce fichier ?

Ouvrir avec Internet Explorer (IE6 ou 7)

Télécharger le fichier

Toujours effectuer cette action pour ce type de fichier.

Figura 23. Informações estruturais do SiliTox sobre a biomolécula SFI8. Vínculos na página (A) abrem páginas de resultados correspondendo respectivamente a : (B) todas as biomoléculas contendo 8 resíduos de cisteína; (C) todas as biomoléculas contendo o padrão de cisteínas selecionado; (D) todas as biomoléculas cujos padrões de cisteínas são classificados como 2.1.1.0 segundo Kozlov; (E) a referência PubMed da publicação do Kozlov sobre a classificação das toxinas de aranhas; (F) a descrição do pareamento das cisteínas nas *knottins* pelo *Knottin Website* ; os dados de modelagem da estrutura da SFI8 por SiliMod : (G) o arquivo de alinhamento e (H) o arquivo das coordenadas atômicas podem ser ambos baixados do SiliTox.

A

Cysteine connectivity

Sequence position	Relative position	ICK convention (?)	Disulfide bridge quality
C9 - C23	C2 - C5	CII - CV	experimentally determined
C2 - C18	C1 - C4	CI - CIV	experimentally determined
C17 - C33	C3 - C8	CIII - CV	experimentally determined
C25 - C31	C6 - C7	Ca - Cb	experimentally determined

Tertiary structure

STRUCTURE

Collection of 20 models
model 120
Configurations
Select (10 400)
View
Style
Color
Surfaces
Symmetry
Zoom
Spin
Vibration
Animation
Measurement
Set picking

METHOD

experimental model

3D determination method : NMR

PDB code : [1V91](#)

[TRP]12:A.H|1 #151 4.5360003 8.721001 2.123

B

RCSB PDB - Structure Explorer - Mozilla Firefox

http://www.rcsb.org/pdb/explore/structure.do?structureId=1V91

RCSB PDB - Structure Explorer

AN INFORMATION PORTAL TO BIOLOGICAL MACROMOLECULAR STRUCTURES

As of Tuesday Feb 24, 2008 there are 56066 Structures | PDB Statistics

CONTACT US | FEEDBACK | HELP | PRINT

1V91

DOI: 10.2210/pdb1-91/pdb

Red - Derived Information

Title	Solution structure of insecticidal toxin delta-palut2-NH2
Authors	Ferrat, G., Bosmans, F., Tytgat, J., Pimentel, C., Chagot, B., Nakajima, T., Darbon, H., Corzo, G.
Primary Citation	Ferrat, G., Bosmans, F., Tytgat, J., Pimentel, C., Chagot, B., Gilles, N., Nakajima, T., Darbon, H., Corzo, G. (2008) Solution structure of two insect-specific spider toxins and their pharmacological interaction with the insect voltage-gated Na(+) channel. <i>Proteins</i> 59: 368-370 [Abstract] PubMed
History	Deposition: 2004-01-19, Release: 2005-03-29, Last Modified: (PDB-STAT) 2005-04-05
Experimental Method	Type: NMR, 20 STRUCTURES Data: N/A
NMR Ensemble	Conformers Calculated: N/A, Conformers Submitted: N/A, Selection Criteria: N/A
NMR Refine	Method: NMR, 20 STRUCTURES

Quick Tips:

To view the 3D structure click on one of the viewers under the image.

Images and Visualization

Asymmetric Unit

Display Options

EMSG
Jmol
WebMol
Mol Protein workshop
QuickPDB
All Images

Figura 24. Informações do SiliTox sobre a estrutura determinada experimentalmente da biomolécula δ -paluIT2. **(A)** As pontes disulfeto são anotadas como “experimentally determined”. Ao clicar na estrutura, as funcionalidades do Jmol podem ser acessadas. **(B)** A estrutura da δ -paluIT2 foi determinada por RMN e está disponível no banco PDB sob o código 1V91.

indicado (Figura 24B). Para cada molécula do SiliTox que tem uma estrutura 3D disponível (que seja experimental ou teórica), um *applet* Java do programa Jmol apresenta a estrutura e permite ao usuário modificar as opções de visualização, usando o botão direito do seu *mouse* (Figura 24A).

Quando uma estrutura 3D puder ser determinada ou prevista para uma determinada molécula, a análise por SiliPop é realizada e os resultados são apresentados numa terceira parte da tabela de descrição da molécula da interface do SiliTox (Figura 25). Os resultados são reunidos numa tabela que contém a lista dos segmentos acessíveis, suas sequências, seus comprimentos e seus agrupamentos segundo as distâncias espaciais (Figura 25A). Um *zoom in* na árvore de *clustering* pode ser aberto numa nova aba (Figura 25D). A lista dos peptídeos antigênicos preditos é apresentada numa tabela contendo as sequências dos peptídeos e o método usado para os obter (Figura 25A). Se o usuário quiser desenhar peptídeos antigênicos segundo seus próprios critérios, ele pode acessar o servidor do PEPOP via um vínculo HTML (Figura 25B), e dar uma olhada no resumo da publicação do PEPOP do mesmo jeito (Figura 25C).

A parte seguinte da descrição da molécula diz respeito às interações com o(s) canal(-is) iônico(s) nas quais ela é possivelmente envolvida. As informações relacionadas com essas interações são fornecidas (Figura 26A) juntas a uma apreciação de confiança e um vínculo para a publicação que deu origem aos dados (Figura 26B). As informações funcionais são também indicadas como descritas no Swiss-Prot, enquanto a implementação do BioΨ (Figura 26A). Emfim, os dados relativos aos efeitos naturais da molécula, incluindo a LD50, se for conhecida, nos organismos sensíveis, e os sintomas consequentes observados são apresentados no final da tabela (Figura 26C).

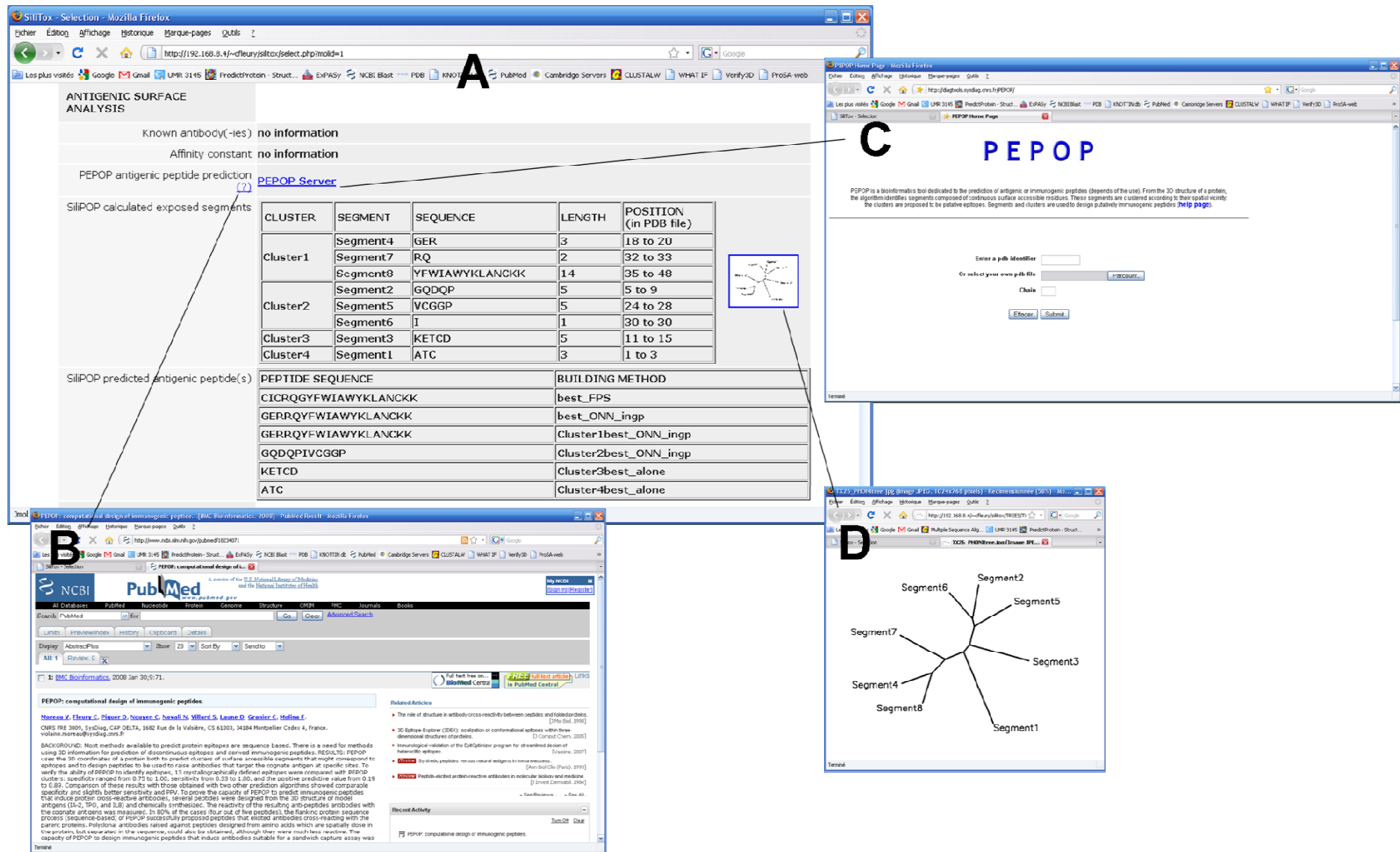


Figura 25. Dados do SiliTox resultando da análise da biomolécula PnTx2-6 por SiliPop. Em (A), “FPS” significa “Flanking Protein Sequence”, “ONN” significa “Optimized Nearest Neighbors”, “best” significa que o segmento de referência é o melhor dentro de todos os outros, “ClusterXbest” significa que o segmento de referência é o melhor dentro do seu próprio cluster, e finalmente “alone”, “ingp” e “inallgp” representam a região de extensão do peptídeo, aliás respectivamente “nenhuma extensão possível”, “extensão dentro do cluster”, e “extensão dentro do conjunto inteiro de segmentos”. (B) Entrada do PubMed para a publicação do PEPOP. (C) Página principal do servidor do PEPOP. (D) Zoom sobre a árvore de clustering dos segmentos da PnTx2-6 produzida por SiliPop.

A

LIGAND BINDING	
Known target channel	Voltage-gated sodium channel type 2 alpha subunit
Short name	rNav1.2
Binding site	site 3 putative source
Channel interacting residues	E1613 putative source
Toxin bioactive surface	R20 R32 K42 K47 K48 Y35 F36 W37 W40 Y41 putative source
Affinity constant	K0.5 = 23nM source
FUNCTIONAL ANNOTATION	
General function	Blocks voltage-gated sodium channels. Causes scratching, lacrimation, hypersalivation, sweating and agitation followed by spastic paralysis of the anterior and posterior extremities and death at dose levels of 0.79 mg/mouse. It significantly activates voltage-dependent sodium channels. Insecticidal to the larval and adult forms of the house fly.
BF	x
BA list	x
BEA list	x

B

Structure and activity analysis of two spider toxins that alter sodium channel inactivation kinetics

Matavel A, Fleury C, Oliveira L, Molina E, De Lima ME, Cruz J, Coimbra M, Richardson M, Ramos C, Beirão P.

In this present work, *LoPhoruzina nigroviridis* toxins PnTx2-5 and PnTx2-6 were shown to markedly delay the fast inactivation kinetics of neuronal-type sodium channels. Furthermore, our data show that they have significant differences in their interaction with the channel: PnTx2-6 has 6 times higher affinity than PnTx2-5, its effects are not reversible within 10-15 min washing, PnTx2-6 partially (59%) competes with the scorpion alpha-toxin Aa1 II, but not with the common fast-inactivator ω -conotoxin, thus suggesting a mode of action similar to site 3 toxins. However, PnTx2-6 is not removed by strong depolarizing pulses, as do the known site 3 toxins. We have also established the correct PnTx2-5 amino acid sequence and confirmed the sequence of PnTx2-6, in both cases establishing that the cysteines are in their oxidized form. A structural model of each toxin is proposed. They show structures with poor alpha-helix content. The model is supported by experimental and theoretical tests. A likely binding region on PnTx2-5 and PnTx2-6 is proposed based on their different affinities and sequence differences.

PMID: 1921838 (PubMed - as supplied by publisher)

C

NATURAL EFFECT	
Toxicity	death at dose levels of 0.79 mg/mouse
Sensitive species	mouse
Symptom(s)	scratching lacrimation hypersalivation sweating agitation paralysis death
APPLICATION	
Biotechnological	no information
Pharmaceutical	no information

Figura 26. Informações funcionais do SiliTox sobre a biomolécula PnTx2-6. **(A)** Dados sobre a interação com o receptor natural e anotação funcional da biomolécula segundo o banco Swiss-Prot (a descrição da função segundo o formalismo Bio Ψ será adicionada em breve). **(B)** Entrada do PubMed relacionada à fonte da informação disponível no SiliTox sobre a interação da PnTx2-6 com o canal Nav1.2 de rato. **(C)** Efeitos naturais, sintomas e aplicações conhecidas da biomolécula selecionada.

DISCUSSÃO

Durante o estudo, nos concentramos nas toxinas peptídicas animais pois elas representam um grande problema de saúde pública em vários lugares do planeta – especialmente no Brasil – e por causa do potencial fantástico delas a serem usadas em diversas áreas biológicas. É preciso entender o modo de ação dessas toxinas para poder prevenir ou tratar os danos causados nos acidentes com esses animais, e para poder tirar vantagem das suas atividades biológicas naturais para outros fins. O entendimento da função de uma proteína conta com a compreensão dos seus mecanismos de interação, o que depende do conhecimento da sua estrutura.

A bioinformática é a ciência da gerência, da mineração e da extração de conhecimento a partir de dados biológicos, particularmente das sequências e das estruturas. Há mais de 20 anos que as abordagens de bioinformática vêm para auxiliar os experimentalistas, seja para organizar os dados existentes ou para extrair novo conhecimento a partir desses dados.

Para fornecer um suporte aos métodos experimentais do estudo, nos desenvolvemos quatro ferramentas de bioinformática dedicados às investigações das relações estrutura-função-antigenicidade. Dentro desses quatro, três foram especificamente concebidas para atender o caso das toxinas peptídicas animais. Essas ferramentas foram chamadas SiliTox, SiliMod e SiliPop segundo a biologia “*in silico*”, enquanto os sufixos “-Tox”, “-Mod” e “-Pop” vêm respectivamente de “*Toxin*”, “*Model*” e “*ePitOPe*”.

As técnicas de modelagem comparativa têm comprovado que são eficientes para calcular modelos 3D de boa qualidade (Moult et al. 2007). Segundo a precisão dos modelos assim gerados, eles podem ser usados para identificar resíduos críticos envolvidos na catálise, numa ligação ou na estabilidade da estrutura, para examinar as interações da proteína com outras moléculas e para guiar o planejamento de um

experimento (Ginalski 2006). A cobertura da conservação estrutural entre a proteína alvo e a proteína *template*, e a exatidão do alinhamento constituem os dois fatores principais que determinam a qualidade do modelo (Marti-Renom et al. 2000; Kryshchuk et al. 2005). No caso da modelagem automática como no algoritmo do SiliMod, não há nenhuma fase de refinamento manual do alinhamento. Além disso, a identidade de sequência entre o alvo e o *template* em alguns casos para as toxinas peptídicas aracnídicas investigadas neste trabalho, pode cair abaixo de 30%. Todavia, mais que a identidade de sequências entre alvo e *template*, a distribuição dos resíduos conservados das sequências nos alinhamentos múltiplos incluindo o alvo, o *template* e as sequências intermediárias e homólogas, constitui um parâmetro relevante para avaliar a qualidade de um modelo (Cozzetto and Tramontano 2005; Kryshchuk et al. 2005; Tress et al. 2005). Os alinhamentos múltiplos de todas as toxinas peptídicas aracnídicas ricas em cisteínas apresentadas neste trabalho mostram uma alta conservação das posições ocupadas pelos resíduos de cisteína e das suas conectividades, dentro dos grupos estruturais, e dadas as restrições impostas pelas pontes disulfeto nas subfamílias, somente poucos enovelamentos são possíveis. Além do mais, dentro de 30-50% de identidade de sequências entre alvo e *template*, os melhores modelos da CASP apresentam com a estrutura experimental uma RMSD dos átomos C α tipicamente incluída entre 2 e 4Å, com erros principalmente localizados nas regiões de alças (Tabela 4). Apesar do que as toxinas peptídicas aracnídicas terem um conteúdo rico em alças nas suas estruturas, nenhum modelo produzido pelo SiliMod no experimento de modelagem automático apresentou uma estrutura nativa com uma RMSD superior a 4Å, e 73% dos modelos mostraram uma RMSD até menor que 2.5Å. Esses resultados demonstram que a modelagem comparativa baseada na distribuição e na conectividade dos resíduos de cisteína constitui uma estratégia pertinente para obter modelos 3D valiosos de toxinas peptídicas aracnídicas ricas em cisteínas.

Modelagem precisa das diferenças entre estruturas diferentes é uma das aplicações mais relevantes biologicamente da modelagem comparativa pois essas mudanças

estruturais associadas com propriedades de superfície correspondem geralmente a novas funções e/ou especificidades (Ginalski 2006). Foi mostrado que os modelos comparativos carregam informações adicionais em relação à estrutura *template* (Chakravarty and Sanchez 2004; Chakravarty et al. 2005). Segundo a propriedade derivada da estrutura que se quer investigar (área acessível na superfície, vizinhança de um resíduo, potencial eletrostático), modelos comparativos baseados em *templates* com uma identidade de sequências de 25-40% são tão úteis quanto estruturas RMN (Chakravarty et al. 2005). Portanto, os modelos do SiliMod são ferramentas valiosas para investigar as interações das toxinas aracnídicas com outras proteínas. PEPOP demonstrou sua capacidade de prever peptídeos que mimetizam uma região da superfície de uma proteína que pode ser reconhecida por um anticorpo. Vários dos peptídeos preditos foram sintetizados e foram capazes induzir uma resposta imune em animais. Após coletar e purificar os anticorpos recrutados pelos peptídeos, foi mostrado que os anticorpos eram também capazes reconhecer a proteína nativa por reação cruzada (Moreau et al. 2008). Esses resultados são muito encorajantes e tal estratégia de síntese de mimotopos convem muito bem ao caso das toxinas animais. De fato, o processo de imunização clássica que consiste em inocular o animal com a proteína nativa para recuperar anticorpos específicos é complicado quando a proteína alvo é tóxica para o organismo. Portanto o método PEPOP está sendo aplicado nos nossos laboratórios para induzir anticorpos anti-toxina específicos.

No entanto, não é a única aplicação da ferramenta. PEPOP e SiliPop realizam predições das regiões potencialmente antigênicas na superfície de uma proteína, *i.e.* regiões mais suscetíveis de interagir ou serem reconhecidas por uma outra proteína. Mais geralmente, a interação Ac/Ag é um caso especial de interação proteína-proteína, já que ela apresenta características compartilhadas por todos os tipos de interações proteína-proteína. Sabe-se que as toxinas peptídicas animais se ligam nos canais iônicos. Conseqüentemente, os peptídeos preditos por SiliPop nessas toxinas são suscetíveis de mimetizar regiões potencialmente interagindo, que seja com um anticorpo ou com um outro tipo de ligante protéico (receptor por exemplo).

Para investigar mais as toxinas da *Phoneutria* que são de um interesse especial no nosso laboratório, decidimos construir os modelos moleculares das toxinas PnTx2-5 e PnTx2-6 para ter uma idéia da interação com os canais sódicos. Já que essas toxinas apresentam uma atividade similar àquela das α -toxinas de escorpiões e compartilham sítios de ligação que se sobreponham parcialmente (Matavel et al., 2009), começamos comparando as estruturas primárias das toxinas PnTx2-5 e PnTx2-6 com àquelas das α -toxinas de escorpiões. Não achamos nenhuma similaridade significativa nas sequências, até nas regiões supostamente participantes da interação com o canal. Todavia, o possível sítio de ligação da PnTx2-6 e propriedades eletrofisiológicas similares das PnTx2-5 e PnTx2-6 com α -toxinas de escorpiões nos incitaram a verificar a hipótese que, apesar da ausência de similaridade de sequência, elas talvez tenham superfícies bioativas similares. Foi mostrado que as α -toxinas de escorpiões têm um *core* hidrofóbico conservado (predominantemente aromático) cercado de resíduos básicos (Fontecilla-Camps et al. 1988; Kharrat et al. 1989). A Figura 27 mostra a comparação das superfícies preditas das PnTx2-5 e PnTx2-6 com a potente α -toxin AaH II e com a muito estudada α -like toxina BmKM1 do escorpião *Buthus martensii* Karsch. Propriedades similares de superfície podem ser identificadas, com uma área hidrofóbica conservada cercada de cargas positivas. Essa observação é coerente com a hipótese que a superfície ativa das toxinas modificadoras de *gating* é um arranjo de resíduos hidrofóbicos e básicos (Kharrat et al. 1989; Takahashi et al. 2000). Além do mais, dentro dos resíduos que não são idênticos entre as duas toxinas, dois deles são localizados no *core* hidrofóbico e são expostos na superfície. Uma tirosina (na posição 35) e um triptofano (na posição 37), dois aminoácidos conhecidos por serem importantes numa interação proteína-proteína (Chakrabarti and Janin 2002), são ausentes do *core* hidrofóbico da PnTx2-5. Os dados de fluorescência da PnTx2-6 mostram que os resíduos de triptofano são expostos ao solvente, como predito nos modelos. Os resíduos aromáticos Phe36, Trp40 e Tyr41 são também expostos nas duas toxinas. Elas contêm os mesmos resíduos de aminoácidos carregados, nas mesmas localizações. Em particular, os resíduos carregados positivamente Arg20, Arg32, Lys42, Lys47 e Lys48 cercam o *patch* hidrofóbico nas duas estruturas (Figuras 27C e 27D). Este arranjo fornece o

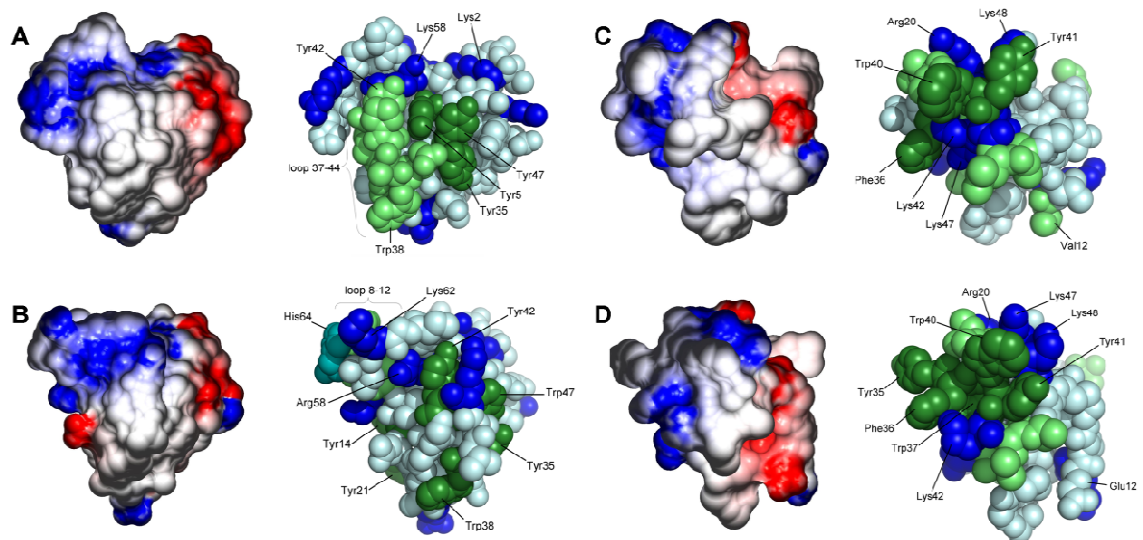


Figura 27. Comparação das superfícies bioativas das neurotoxinas modificadoras de canal de sódio que se ligam ao sítio 3 dos canais de sódio de mamíferos : **(A)** AaH II, **(B)** BmKM1, **(C)** PnTx2-5, e **(D)** PnTx2-6. O potencial eletrostático é mapeado na superfície da estrutura de cada molécula (lado esquerdo). O potencial é mostrado entre -5.0 e $+5.0$ kT/e , com k representando a constante de Boltzmann, T a temperatura, e e a carga eletrônica. Vermelho indica regiões de potencial eletrostático negativo; branco indica regiões neutras e azul indica regiões positivas. Estes mapas foram desenhados usando o programa GRASP2. A representação em esferas gerada pelo programa PyMOL é mostrada para cada molécula (lado direito). Os resíduos de aminoácidos presumidos ou conhecidos como envolvidos na interação com o canal estão coloridos e etiquetados. Em azul são representados os resíduos positivamente carregados. Os resíduos hidrofóbicos são enfatizados em verde claro e os resíduos aromáticos são mostrados em verde escuro.

requisito básico esperado para a interação com o sítio 3 do canal de sódio e se conforma à “área de reconhecimento” (“*recognition patch*”) encontrado nas superfícies interagindo das proteínas (Chakrabarti and Janin 2002). Além disso, a predição por PEPOP sobre PnTx2-5 e PnTx2-6 aponta a parte C-terminal das duas toxinas, incluindo os resíduos aromáticos Phe36, Trp40 e Tyr41, como sendo a melhor região potencialmente antigénica. Esse resultado é coerente com as observações experimentais e reforça a hipótese que PnTx2-5 e PnTx2-6 se ligam no sítio 3 do canal de sódio deste lado das suas superfícies respectivas. Sugerimos que a disposição das cargas junto ao patch aromático que elas cercam (resíduos Phe36, Trp40 e Tyr41) compõem a superfície bioativa das PnTx2-5 e PnTx2-6. A presença adicional de Tyr35 e Trp37 na PnTx2-6 pode explicar a sua maior afinidade pelo canal.

A análise em larga-escala dos peptídeos produzida por SiliPop para todas as estruturas de toxinas aracnídicas presentes no banco SiliTox deve fornecer informações sobre as características das superfícies que têm de ser compartilhadas por toxinas de espécies, estruturas e tamanhos diferentes, dado que elas se ligam ao mesmo canal iônico. Inversamente, alvos potenciais comuns poderiam ser inferidos a partir de toxinas que compartilham superfícies bioativas similares. Baseando-se nessas observações, anticorpos poderiam ser selecionados contra uma região comum a uma família funcional inteira de toxinas, independentemente da espécie da qual elas provêm. O sucesso desse tipo de experimento pode levar à produção de anticorpos poliespecíficos, a serem usados como tratamento eficaz contra vários tipos de envenenamentos animais.

CONCLUSÃO & PERSPECTIVAS

Quatro ferramentas bioinformáticas foram desenvolvidos dentro do contexto deste trabalho. PEPOP e SiliTox estão disponíveis nos endereços <http://diagtools.sysdiag.cnrs.fr/PEPOP/> e <http://192.168.8.4/~cfleury/silitox/> respectivamente. SiliMod e SiliPop estão para serem disponibilizados pela Internet.

Até agora, SiliTox contém todas as toxinas peptídicas ricas em cisteínas conhecidas dos venenos de aranhas e escorpiões. O banco de dados é operacional e todas as informações estão organizadas para serem facilmente recuperadas através da interface simples do SiliTox. A estrutura do banco SiliTox permite evolução, e futuros melhoramentos do banco são esperados. Por exemplo, planejamos incluir informações anotadas sobre toxinas de caramujo marinos e curtas toxinas de serpentes ao lado das curtas toxinas aracnídicas, para fazer que as investigações sobre as relações estrutura-função-antigenicidade das toxinas peptídicas sejam extendidas ao conjunto de todas as toxinas peptídicas animais ricas em cisteínas conhecidas.

Além disso, temos a intenção de adicionar uma nova e importante função ao banco de dados, que é a anotação funcional das funções das toxinas segundo o formalismo Bio Ψ (Maziere et al. 2004). A linguagem Bio Ψ está sendo desenvolvida atualmente no laboratório francês SysDiag (UMR3145 CNRS / Bio-Rad), em Montpellier. Será logo possível traduzir a função biológica descrita na linguagem natural num conjunto de elementos básicos de ação ("*Basic Element of Action*") escritos numa linguagem formal que permite a interpretação por um computador. Quando a fase final do desenvolvimento da linguagem Bio Ψ estiver completa, a linguagem será implementada no SiliTox, a interface do banco de dados permitirá ao usuário seleccionar toxinas segundo o modo de ação delas em diversos níveis: das funcionalidades gerais ao mecanismo bioquímico primário de ação.

As estruturas 3D de 350 toxinas peptídicas aracnídicas ricas em cisteínas foram preditas por SiliMod. Usando Modeller, o program calculou as estruturas 3D por modelagem comparativa segundo o critério de similaridade de motivos e conectividades dos resíduos de cisteína. Junto às 111 estruturas 3D que já estavam indexadas no PDB, um total de 461 estruturas 3D de toxinas são agora disponíveis no SiliTox. Para uma melhor avaliação da qualidade dos modelos, pretendemos integrar um sistema de verificação multi-critério das estruturas das toxinas. Vinculando o banco SiliTox aos serviços propostos pelo servidor do CBS, seremos capaz fornecer uma avaliação automática (incluindo *plots* de Ramachandran, score de Verify3D e de ProSA, etc.) para cada toxina presente no SiliTox. Um código de cores permitirá ao usuário conhecer rapidamente a confiabilidade do modelo. A avaliação da qualidade dos dados é um assunto que está sendo estudado no laboratório SysDiag.

PEPOP demonstrou seu potencial a guiar os experimentalistas que querem localizar um epitopo ou conceber peptídeos imunogênicos para gerar anticorpos capazes de se ligar em sítios específicos de proteínas (Moreau et al. 2008). De um modo mais geral, o programa pode ser usado para prever qualquer superfície de interação numa estrutura 3D protéica. SiliPop é uma derivação de PEPOP, adaptada aos grandes conjuntos de estruturas de proteínas a serem analisadas de uma vez e com os mesmos parâmetros. Em comparação ao algoritmo do PEPOP, o algoritmo do SiliPop foi ligeiramente modificado para melhor lidar com pequenas proteínas. Todas as toxinas aracnídicas peptídicas e ricas em cisteínas indexadas no SiliTox foram sistematicamente submetidas à análise por SiliPop e os peptídeos potencialmente antigênicos preditos que cobrem a superfície acessível inteira dessas toxinas são disponíveis no SiliTox. Para permitir ao usuário desenhar os peptídeos antigênicos segundo seus próprios critérios, um vínculo para a interface *web* de desenho de peptídeo por PEPOP foi previsto e está presente no SiliTox.

No que diz respeito às toxinas PnTx2-5 e PnTx2-6, os experimentos envolvendo a síntese química dos peptídeos preditos por PEPOP estão sendo realizados. Os ensaios sobre a atividade biológica e o potencial imunogênico desses peptídeos devem permitir validar logo nossas hipóteses.

Para concluir, SiliTox é o primeiro banco de informações sobre as relações estrutura-função-antigenicidade relacionadas às toxinas animais. O banco de dados SiliTox é dedicado à auxiliar e guiar a concepção de experimentos para investigar a atividade, as propriedades antigênicas e as aplicações terapêuticas e biotecnológicas potenciais das toxinas peptídicas animais. Dessa maneira, SiliTox constitui um recurso útil e único para os toxinologistas das áreas da imunologia, da bioquímica, da farmacologia, da eletrofisiologia e da biotecnologia.

REFERÊNCIAS BIBLIOGRÁFICAS

- Adams, M. E. (2004). "Agatoxins: ion channel specific toxins from the American funnel web spider, *Agelenopsis aperta*." Toxicon **43**(5): 509-525.
- Altschul, S. F., E. M. Gertz, R. Agarwala, A. A. Schaffer and Y. K. Yu (2008). "PSI-BLAST pseudocounts and the minimum description length principle." Nucleic Acids Res.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-3402.
- Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(96): 223-230.
- Araujo, D. A., M. N. Cordeiro, C. R. Diniz and P. S. Beirao (1993). "Effects of a toxic fraction, PhTx2, from the spider *Phoneutria nigriventer* on the sodium current." Naunyn Schmiedebergs Arch Pharmacol **347**(2): 205-208.
- Argos, P. (1988). "An investigation of protein subunit and domain interfaces." Protein Eng **2**(2): 101-113.
- Atassi, M. Z. and J. A. Smith (1978). "A proposal for the nomenclature of antigenic sites in peptides and proteins." Immunochemistry **15**(8): 609-610.
- Bairoch, A. (1993). "The PROSITE dictionary of sites and patterns in proteins, its current status." Nucleic Acids Res **21**(13): 3097-3103.
- Bairoch, A., B. Boeckmann, S. Ferro and E. Gasteiger (2004). "Swiss-Prot: juggling between evolution and stability." Brief Bioinform **5**(1): 39-55.
- Baker, D. and A. Sali (2001). "Protein structure prediction and structural genomics." Science **294**(5540): 93-96.
- Barhanin, J., J. R. Giglio, P. Leopold, A. Schmid, S. V. Sampaio and M. Lazdunski (1982). "Tityus serrulatus venom contains two classes of toxins. Tityus gamma toxin is a new tool with a very high affinity for studying the Na⁺ channel." J Biol Chem **257**(21): 12553-12558.

Batori, V., E. P. Friis, H. Nielsen and E. L. Roggen (2006). "An in silico method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context." J Mol Recognit **19**(1): 21-29.

Becerril, B., S. Marangoni and L. D. Possani (1997). "Toxins and genes isolated from scorpions of the genus Tityus." Toxicon **35**(6): 821-835.

Bechis, G., F. Sampieri, P. M. Yuan, T. Brando, M. F. Martin, C. R. Diniz and H. Rochat (1984). "Amino acid sequence of toxin VII, a beta-toxin from the venom of the scorpion Tityus serrulatus." Biochem Biophys Res Commun **122**(3): 1146-1153.

Benjamin, D. C., J. A. Berzofsky, I. J. East, F. R. Gurd, C. Hannum, S. J. Leach, E. Margoliash, J. G. Michael, A. Miller, E. M. Prager and et al. (1984). "The antigenic structure of proteins: a reappraisal." Annu Rev Immunol **2**: 67-101.

Bennett-Lovsey, R. M., A. D. Herbert, M. J. Sternberg and L. A. Kelley (2008). "Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre." Proteins **70**(3): 611-625.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and E. W. Sayers (2009). "GenBank." Nucleic Acids Res **37**(Database issue): D26-31.

Benzinger, G. R., J. W. Kyle, K. M. Blumenthal and D. A. Hanck (1998). "A specific interaction between the cardiac sodium channel and site-3 toxin anthopleurin B." J Biol Chem **273**(1): 80-84.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-242.

Berzofsky, J. A. (1984). "Mechanisms of antigen-specific, genetically restricted, T cell-B cell interaction." Surv Immunol Res **3**(2-3): 103-106.

Billen, B., F. Bosmans and J. Tytgat (2008). "Animal peptides targeting voltage-activated sodium channels." Curr Pharm Des **14**(24): 2492-2502.

Blumenthal, K. M. and A. L. Seibert (2003). "Voltage-gated sodium channel toxins: poisons, probes, and future promise." Cell Biochem Biophys **38**(2): 215-238.

Boehm, M. K., J. M. Woof, M. A. Kerr and S. J. Perkins (1999). "The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling." J Mol Biol **286**(5): 1421-1447.

Bogan, A. A. and K. S. Thorn (1998). "Anatomy of hot spots in protein interfaces." J Mol Biol **280**(1): 1-9.

Bonnycastle, L. L., J. S. Mehroke, M. Rashed, X. Gong and J. K. Scott (1996). "Probing the basis of antibody reactivity with a panel of constrained peptide libraries displayed by filamentous phage." J Mol Biol **258**(5): 747-762.

Bontems, F., B. Gilquin, C. Roumestand, A. Menez and F. Toma (1992). "Analysis of side-chain organization on a refined model of charybdotoxin: structural and functional implications." Biochemistry **31**(34): 7756-7764.

Bontems, F., C. Roumestand, P. Boyot, B. Gilquin, Y. Doljansky, A. Menez and F. Toma (1991). "Three-dimensional structure of natural charybdotoxin in aqueous solution by 1H-NMR. Charybdotoxin possesses a structural motif found in other scorpion toxins." Eur J Biochem **196**(1): 19-28.

Bontems, F., C. Roumestand, B. Gilquin, A. Menez and F. Toma (1991). "Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins." Science **254**(5037): 1521-1523.

Bork, P. and T. J. Gibson (1996). "Applying motif and profile searches." Methods Enzymol **266**: 162-184.

Bosmans, F., B. Brone, Y. M. Sun, R. H. Zhu, Y. M. Xiong, D. C. Wang, E. Van Kerkhove and J. Tytgat (2005). "Pharmacological comparison of two different insect models using the scorpion alpha-like toxin BmK M1 from *Buthus martensii* Karsch." Protein Pept Lett **12**(4): 363-367.

Bosmans, F., M. F. Martin-Eauclaire and J. Tytgat (2005). "The depressant scorpion neurotoxin LqqIT2 selectively modulates the insect voltage-gated sodium channel." Toxicon **45**(4): 501-507.

Bosmans, F., M. F. Martin-Eauclaire and J. Tytgat (2007). "Differential effects of five 'classical' scorpion beta-toxins on rNav1.2a and DmNav1 provide clues on species-selectivity." Toxicol Appl Pharmacol **218**(1): 45-51.

Bosmans, F. and J. Tytgat (2007). "Voltage-gated sodium channel modulation by scorpion alpha-toxins." Toxicon **49**(2): 142-158.

Bourinet, E., S. C. Stotz, R. L. Spaetgens, G. Dayanithi, J. Lemos, J. Nargeot and G. W. Zamponi (2001). "Interaction of SNX482 with domains III and IV inhibits activation gating of alpha(1E) (Ca(V)2.3) calcium channels." Biophys J **81**(1): 79-88.

Boutet, E., D. Lieberherr, M. Tognolli, M. Schneider and A. Bairoch (2007). "UniProtKB/Swiss-Prot." Methods Mol Biol **406**: 89-112.

Bradley, P., K. M. Misura and D. Baker (2005). "Toward high-resolution de novo structure prediction for small proteins." Science **309**(5742): 1868-1871.

Brone, B., J. Tytgat, D. C. Wang and E. Van Kerkhove (2003). "Characterization of Na(+) currents in isolated dorsal unpaired median neurons of *Locusta migratoria* and effect of the alpha-like scorpion toxin BmK M1." J Insect Physiol **49**(2): 171-182.

Casari, G., C. Sander and A. Valencia (1995). "A method to predict functional residues in proteins." Nat Struct Biol **2**(2): 171-178.

Catterall, W. A. (1980). "Neurotoxins that act on voltage-sensitive sodium channels in excitable membranes." Annu Rev Pharmacol Toxicol **20**: 15-43.

Catterall, W. A., S. Cestele, V. Yarov-Yarovoy, F. H. Yu, K. Konoki and T. Scheuer (2007). "Voltage-gated ion channels and gating modifier toxins." Toxicon **49**(2): 124-141.

Ceard, B., M. E. De Lima, P. E. Bougis and M. F. Martin-Eauclaire (1992). "Purification of the main beta-toxin from *Tityus serrulatus* scorpion venom using high-performance liquid chromatography." Toxicon **30**(1): 105-110.

Cestele, S., R. B. Ben Khalifa, M. Pelhate, H. Rochat and D. Gordon (1995). "Alpha-scorpion toxins binding on rat brain and insect sodium channels reveal divergent allosteric modulations by brevetoxin and veratridine." J Biol Chem **270**(25): 15153-15161.

Cestele, S. and W. A. Catterall (2000). "Molecular mechanisms of neurotoxin action on voltage-gated sodium channels." Biochimie **82**(9-10): 883-892.

Chakrabarti, P. and J. Janin (2002). "Dissecting protein-protein recognition sites." Proteins **47**(3): 334-343.

Chakravarty, S. and R. Sanchez (2004). "Systematic analysis of added-value in simple comparative models of protein structure." Structure **12**(8): 1461-1470.

Chakravarty, S., L. Wang and R. Sanchez (2005). "Accuracy of structure-derived properties in simple comparative models of protein structures." Nucleic Acids Res **33**(1): 244-259.

Chen, H., S. Lu, E. Leipold, D. Gordon, A. Hansel and S. H. Heinemann (2002). "Differential sensitivity of sodium channels from the central and peripheral nervous system to the scorpion toxins Lqh-2 and Lqh-3." Eur J Neurosci **16**(4): 767-770.

Chen, Y. H., J. T. Yang and H. M. Martinez (1972). "Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion." Biochemistry **11**(22): 4120-4131.

Chong, Y., J. L. Hayes, B. Sollod, S. Wen, D. T. Wilson, P. G. Hains, W. C. Hodgson, K. W. Broady, G. F. King and G. M. Nicholson (2007). "The omega-atracotoxins: selective blockers of insect M-LVA and HVA calcium channels." Biochem Pharmacol **74**(4): 623-638.

Chothia, C. (1974). "Hydrophobic bonding and accessible surface area in proteins." Nature **248**(446): 338-339.

Chothia, C. and J. Janin (1975). "Principles of protein-protein recognition." Nature **256**(5520): 705-708.

Chothia, C. and A. M. Lesk (1986). "The relation between the divergence of sequence and structure in proteins." Embo J **5**(4): 823-826.

Chuang, R. S., H. Jaffe, L. Cribbs, E. Perez-Reyes and K. J. Swartz (1998). "Inhibition of T-type voltage-gated calcium channels by a new scorpion toxin." Nat Neurosci **1**(8): 668-674.

Cleveland, D. W., S. G. Fischer, M. W. Kirschner and U. K. Laemmli (1977). "Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis." J Biol Chem **252**(3): 1102-1106.

Codd, E. F. (1970). "A relational model of data for large shared data banks." MD Comput **15**(3): 162-166.

Coetzee, W. A., Y. Amarillo, J. Chiu, A. Chow, D. Lau, T. McCormack, H. Moreno, M. S. Nadal, A. Ozaita, D. Pountney, M. Saganich, E. Vega-Saenz de Miera and B. Rudy (1999). "Molecular diversity of K⁺ channels." Ann N Y Acad Sci **868**: 233-285.

Corzo, G., P. Escoubas, M. Stankiewicz, M. Pelhate, C. P. Kristensen and T. Nakajima (2000). "Isolation, synthesis and pharmacological characterization of delta-palutoxins IT, novel insecticidal toxins from the spider *Paracoelotes luctuosus* (Amaurobiidae)." Eur J Biochem **267**(18): 5783-5795.

Corzo, G., P. Escoubas, E. Villegas, I. Karbat, D. Gordon, M. Gurevitz, T. Nakajima and N. Gilles (2005). "A spider toxin that induces a typical effect of scorpion alpha-toxins but competes with beta-toxins on binding to insect sodium channels." Biochemistry **44**(5): 1542-1549.

Corzo, G., N. Gilles, H. Satake, E. Villegas, L. Dai, T. Nakajima and J. Haupt (2003). "Distinct primary structures of the major peptide toxins from the venom of the spider *Macrothele gigas* that bind to sites 3 and 4 in the sodium channel." FEBS Lett **547**(1-3): 43-50.

Corzo, G., J. K. Sabo, F. Bosmans, B. Billen, E. Villegas, J. Tytgat and R. S. Norton (2007). "Solution structure and alanine scan of a spider toxin that affects the activation of mammalian voltage-gated sodium channels." J Biol Chem **282**(7): 4643-4652.

Cozzetto, D. and A. Tramontano (2005). "Relationship between multiple sequence alignments and quality of protein comparative models." Proteins **58**(1): 151-157.

Craik, D. J., N. L. Daly and C. Wayne (2001). "The cystine knot motif in toxins and implications for drug design." Toxicon **39**(1): 43-60.

Cuff, J. A., M. E. Clamp, A. S. Siddiqui, M. Finlay and G. J. Barton (1998). "JPred: a consensus secondary structure prediction server." Bioinformatics **14**(10): 892-893.

Dandekar, T., B. Snel, M. Huynen and P. Bork (1998). "Conservation of gene order: a fingerprint of proteins that physically interact." Trends Biochem Sci **23**(9): 324-328.

Darbon, H., E. Jover, F. Couraud and H. Rochat (1983). "Alpha-scorpion neurotoxin derivatives suitable as potential markers of sodium channels. Preparation and characterization." Int J Pept Protein Res **22**(2): 179-186.

Das, R., B. Qian, S. Raman, R. Vernon, J. Thompson, P. Bradley, S. Khare, M. D. Tyka, D. Bhat, D. Chivian, D. E. Kim, W. H. Sheffler, L. Malmstrom, A. M. Wollacott, C. Wang, I. Andre and D. Baker (2007). "Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home." Proteins **69 Suppl 8**: 118-128.

de Figueiredo, S. G., M. E. de Lima, M. Nascimento Cordeiro, C. R. Diniz, D. Patten, R. F. Halliwell, J. Gilroy and M. Richardson (2001). "Purification and amino acid sequence of a highly insecticidal toxin from the venom of the brazilian spider *Phoneutria nigriventer* which inhibits NMDA-evoked currents in rat hippocampal neurones." Toxicon **39**(2-3): 309-317.

de la Vega, R. C. and L. D. Possani (2007). "Novel paradigms on scorpion toxins that affects the activating mechanism of sodium channels." Toxicon **49**(2): 171-180.

DeLano, W. L. (2002). "The PyMOL Molecular Graphics System." on World Wide Web <http://www.pymol.org>.

Dill, K. A. (1990). "The meaning of hydrophobicity." Science **250**(4978): 297-298.

Diochot, S., M. Salinas, A. Baron, P. Escoubas and M. Lazdunski (2007). "Peptides inhibitors of acid-sensing ion channels." Toxicon **49**(2): 271-284.

Eddy, S. R. (1998). "Profile hidden Markov models." Bioinformatics **14**(9): 755-763.

Eisenberg, D., R. Luthy and J. U. Bowie (1997). "VERIFY3D: assessment of protein models with three-dimensional profiles." Methods Enzymol **277**: 396-404.

Eisenberg, D., E. M. Marcotte, I. Xenarios and T. O. Yeates (2000). "Protein function in the post-genomic era." Nature **405**(6788): 823-826.

Eitan, M., E. Fowler, R. Herrmann, A. Duval, M. Pelhate and E. Zlotkin (1990). "A scorpion venom neurotoxin paralytic to insects that affects sodium current inactivation: purification, primary structure, and mode of action." Biochemistry **29**(25): 5941-5947.

el Ayeb, M., E. M. Bahraoui, C. Granier and H. Rochat (1986). "Use of antibodies specific to defined regions of scorpion alpha-toxin to study its interaction with its receptor site on the sodium channel." Biochemistry **25**(21): 6671-6678.

Enshell-Seijffers, D., D. Denisov, B. Groisman, L. Smelyanski, R. Meyuhas, G. Gross, G. Denisova and J. M. Gershoni (2003). "The mapping and reconstitution of a conformational discontinuous B-cell epitope of HIV-1." J Mol Biol **334**(1): 87-101.

Escoubas, P. and L. Rash (2004). "Tarantulas: eight-legged pharmacists and combinatorial chemists." Toxicon **43**(5): 555-574.

Eswar, N., D. Eramian, B. Webb, M. Y. Shen and A. Sali (2008). "Protein structure modeling with MODELLER." Methods Mol Biol **426**: 145-159.

Fagerstam, L. G., A. Frostell, R. Karlsson, M. Kullman, A. Larsson, M. Malmqvist and H. Butt (1990). "Detection of antigen-antibody interactions by surface plasmon resonance. Application to epitope mapping." J Mol Recognit **3**(5-6): 208-214.

Fariselli, P., F. Pazos, A. Valencia and R. Casadio (2002). "Prediction of protein--protein interaction sites in heterocomplexes with neural networks." Eur J Biochem **269**(5): 1356-1361.

Ferrat, G., F. Bosmans, J. Tytgat, C. Pimentel, B. Chagot, N. Gilles, T. Nakajima, H. Darbon and G. Corzo (2005). "Solution structure of two insect-specific spider toxins and their pharmacological interaction with the insect voltage-gated Na⁺ channel." Proteins **59**(2): 368-379.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature **340**(6230): 245-246.

Fiser, A. and A. Sali (2003). "Modeller: generation and refinement of homology-based protein structure models." Methods Enzymol **374**: 461-491.

Fletcher, J. I., B. E. Chapman, J. P. Mackay, M. E. Howden and G. F. King (1997). "The structure of versutoxin (delta-atracotoxin-Hv1) provides insights into the binding of site 3 neurotoxins to the voltage-gated sodium channel." Structure **5**(11): 1525-1535.

Fontecilla-Camps, J. C., R. J. Almassy, F. L. Suddath and C. E. Bugg (1982). "The three-dimensional structure of scorpion neurotoxins." Toxicon **20**(1): 1-7.

Fontecilla-Camps, J. C., C. Habersetzer-Rochat and H. Rochat (1988). "Orthorhombic crystals and three-dimensional structure of the potent toxin II from the scorpion *Androctonus australis* Hector." Proc Natl Acad Sci U S A **85**(20): 7443-7447.

Frank, R. (2002). "The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports--principles and applications." J Immunol Methods **267**(1): 13-26.

Froy, O., N. Zilberberg, D. Gordon, M. Turkov, N. Gilles, M. Stankiewicz, M. Pelhate, E. Loret, D. A. Oren, B. Shaanan and M. Gurevitz (1999). "The putative bioactive surface of insect-selective scorpion excitatory neurotoxins." J Biol Chem **274**(9): 5769-5776.

Furtado, P. B., P. W. Whitty, A. Robertson, J. T. Eaton, A. Almogren, M. A. Kerr, J. M. Woof and S. J. Perkins (2004). "Solution structure determination of monomeric human IgA2 by X-ray and neutron scattering, analytical ultracentrifugation and constrained modelling: a comparison with monomeric human IgA1." J Mol Biol **338**(5): 921-941.

Gavin, A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelman, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga (2002). "Functional organization of the yeast proteome by systematic analysis of protein complexes." Nature **415**(6868): 141-147.

Gelly, J. C., J. Gracy, Q. Kaas, D. Le-Nguyen, A. Heitz and L. Chiche (2004). "The KNOTTIN website and database: a new information system dedicated to the knottin scaffold." Nucleic Acids Res **32**(Database issue): D156-159.

Geysen, H. M., S. J. Rodda and T. J. Mason (1986). "A priori delineation of a peptide which mimics a discontinuous antigenic determinant." Mol Immunol **23**(7): 709-715.

Gilles, N., C. Blanchet, I. Shichor, M. Zaninetti, I. Lotan, D. Bertrand and D. Gordon (1999). "A scorpion alpha-like toxin that is active on insects and mammals reveals an unexpected specificity and distribution of sodium channel subtypes in rat brain neurons." J Neurosci **19**(20): 8730-8739.

Gilquin, B., J. Racape, A. Wrisch, V. Visan, A. Lecoq, S. Grissmer, A. Menez and S. Gasparini (2002). "Structure of the BgK-Kv1.1 complex based on distance restraints identified by double mutant cycles. Molecular basis for convergent evolution of Kv1 channel blockers." J Biol Chem **277**(40): 37406-37413.

Ginalski, K. (2006). "Comparative modeling for protein structure prediction." Curr Opin Struct Biol **16**(2): 172-177.

Ginalski, K., N. V. Grishin, A. Godzik and L. Rychlewski (2005). "Practical lessons from protein structure prediction." Nucleic Acids Res **33**(6): 1874-1891.

Glaser, F., D. M. Steinberg, I. A. Vakser and N. Ben-Tal (2001). "Residue frequencies and pairing preferences at protein-protein interfaces." Proteins **43**(2): 89-102.

Gobel, U., C. Sander, R. Schneider and A. Valencia (1994). "Correlated mutations and residue contacts in proteins." Proteins **18**(4): 309-317.

Goldin, A. L. (1999). "Diversity of mammalian voltage-gated sodium channels." Ann N Y Acad Sci **868**: 38-50.

Gordon, D., I. Karbat, N. Ilan, L. Cohen, R. Kahn, N. Gilles, K. Dong, W. Stuhmer, J. Tytgat and M. Gurevitz (2007). "The differential preference of scorpion alpha-toxins for insect or mammalian sodium channels: implications for improved insect control." Toxicon **49**(4): 452-472.

Gordon, D., M. F. Martin-Eauclaire, S. Cestele, C. Kopeyan, E. Carlier, R. B. Khalifa, M. Pelhate and H. Rochat (1996). "Scorpion toxins affecting sodium current inactivation bind to distinct homologous receptor sites on rat brain and insect sodium channels." J Biol Chem **271**(14): 8034-8045.

Gotoh, O. (1992). "Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences." J Biol Chem **267**(1): 83-90.

Goudet, C., I. Huys, E. Clynen, L. Schoofs, D. C. Wang, E. Waelkens and J. Tytgat (2001). "Electrophysiological characterization of BmK M1, an alpha-like toxin from *Buthus martensi* Karsch venom." FEBS Lett **495**(1-2): 61-65.

Gracy, J., D. Le-Nguyen, J. C. Gelly, Q. Kaas, A. Heitz and L. Chiche (2008). "KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007." Nucleic Acids Res **36**(Database issue): D314-319.

Guan, R. J., Y. Xiang, X. L. He, C. G. Wang, M. Wang, Y. Zhang, E. J. Sundberg and D. C. Wang (2004). "Structural mechanism governing cis and trans isomeric states and an intramolecular switch for cis/trans isomerization of a non-proline peptide bond observed in crystal structures of scorpion toxins." J Mol Biol **341**(5): 1189-1204.

Guntert, P. (2004). "Automated NMR structure calculation with CYANA." Methods Mol Biol **278**: 353-378.

Gurevitz, M., D. Gordon, S. Ben-Natan, M. Turkov and O. Froy (2001). "Diversification of neurotoxins by C-tail 'wiggling': a scorpion recipe for survival." Faseb J **15**(7): 1201-1205.

Gurevitz, M., I. Karbat, L. Cohen, N. Ilan, R. Kahn, M. Turkov, M. Stankiewicz, W. Stuhmer, K. Dong and D. Gordon (2007). "The insecticidal potential of scorpion beta-toxins." Toxicon **49**(4): 473-489.

Gutman, G. A., K. G. Chandy, S. Grissmer, M. Lazdunski, D. McKinnon, L. A. Pardo, G. A. Robertson, B. Rudy, M. C. Sanguinetti, W. Stuhmer and X. Wang (2005). "International Union of Pharmacology. LIII. Nomenclature and molecular relationships of voltage-gated potassium channels." Pharmacol Rev **57**(4): 473-508.

Halperin, I., B. Ma, H. Wolfson and R. Nussinov (2002). "Principles of docking: An overview of search algorithms and a guide to scoring functions." Proteins **47**(4): 409-443.

Halperin, I., H. Wolfson and R. Nussinov (2003). "SiteLight: binding-site prediction using phage display libraries." Protein Sci **12**(7): 1344-1359.

Harris, L. J., S. B. Larson, K. W. Hasel and A. McPherson (1997). "Refined structure of an intact IgG2a monoclonal antibody." Biochemistry **36**(7): 1581-1597.

Harris, L. J., E. Skaletsky and A. McPherson (1998). "Crystallographic structure of an intact IgG1 monoclonal antibody." J Mol Biol **275**(5): 861-872.

Hassani, O., P. Mansuelle, S. Cestele, M. Bourdeaux, H. Rochat and F. Sampieri (1999). "Role of lysine and tryptophan residues in the biological activity of toxin VII (Ts gamma) from the scorpion *Tityus serrulatus*." Eur J Biochem **260**(1): 76-86.

He, Q. Y., Q. Z. He, X. C. Deng, L. Yao, E. Meng, Z. H. Liu and S. P. Liang (2008). "ATDB: a uni-database platform for animal toxins." Nucleic Acids Res **36**(Database issue): D293-297.

He, X. L., H. M. Li, Z. H. Zeng, X. Q. Liu, M. Wang and D. C. Wang (1999). "Crystal structures of two alpha-like scorpion toxins: non-proline cis peptide bonds and implications for new binding site selectivity on the sodium channel." J Mol Biol **292**(1): 125-135.

Ho, Y., A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Rasmussen, J. R. Andersen, L. E. Johansen, L. H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sorensen, J. Matthiesen, R. C. Hendrickson, F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. Hogue, D. Figgeys and M. Tyers (2002). "Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry." Nature **415**(6868): 180-183.

Hobohm, U. and C. Sander (1995). "A sequence property approach to searching protein databases." J Mol Biol **251**(3): 390-399.

Holm, L., S. Kaariainen, P. Rosenstrom and A. Schenkel (2008). "Searching protein structure databases with DaliLite v.3." Bioinformatics **24**(23): 2780-2781.

Hopp, T. P. and K. R. Woods (1981). "Prediction of protein antigenic determinants from amino acid sequences." Proc Natl Acad Sci U S A **78**(6): 3824-3828.

Housset, D., C. Habersetzer-Rochat, J. P. Astier and J. C. Fontecilla-Camps (1994). "Crystal structure of toxin II from the scorpion *Androctonus australis Hector* refined at 1.3 Å resolution." J Mol Biol **238**(1): 88-103.

Jacobs, T. (2005). "No pain, no gain?" Nat Biotechnol **23**(8): 934.

Janin, J., S. Miller and C. Chothia (1988). "Surface, subunit interfaces and interior of oligomeric proteins." J Mol Biol **204**(1): 155-164.

Jasti, J., H. Furukawa, E. B. Gonzales and E. Gouaux (2007). "Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH." Nature **449**(7160): 316-323.

Jemmerson, R. and Y. Paterson (1986). "Mapping epitopes on a protein antigen by the proteolysis of antigen-antibody complexes." Science **232**(4753): 1001-1004.

Jin, L., B. M. Fendly and J. A. Wells (1992). "High resolution functional analysis of antibody-antigen interactions." J Mol Biol **226**(3): 851-865.

Jones, S. and J. M. Thornton (1995). "Protein-protein interactions: a review of protein dimer structures." Prog Biophys Mol Biol **63**(1): 31-65.

Jones, S. and J. M. Thornton (1996). "Principles of protein-protein interactions." Proc Natl Acad Sci U S A **93**(1): 13-20.

Jones, S. and J. M. Thornton (1997). "Prediction of protein-protein interaction sites using patch analysis." J Mol Biol **272**(1): 133-143.

Jungo, F. and A. Bairoch (2005). "Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein knowledgebase." Toxicon **45**(3): 293-301.

Kaas, Q., J. C. Westermann, R. Halai, C. K. Wang and D. J. Craik (2008). "ConoServer, a database for conopeptide sequences and structures." Bioinformatics **24**(3): 445-446.

Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers **22**(12): 2577-2637.

Karbat, I., L. Cohen, N. Gilles, D. Gordon and M. Gurevitz (2004). "Conversion of a scorpion toxin agonist into an antagonist highlights an acidic residue involved in voltage sensor trapping during activation of neuronal Na⁺ channels." Faseb J **18**(6): 683-689.

Karbat, I., M. Turkov, L. Cohen, R. Kahn, D. Gordon, M. Gurevitz and F. Frolow (2007). "X-ray structure and mutagenesis of the scorpion depressant toxin LqhIT2

reveals key determinants crucial for activity and anti-insect selectivity." J Mol Biol **366**(2): 586-601.

Karplus, K., C. Barrett, M. Cline, M. Diekhans, L. Grate and R. Hughey (1999). "Predicting protein structure using only sequence information." Proteins Suppl **3**: 121-125.

Kharrat, R., H. Darbon, C. Granier and H. Rochat (1990). "Structure-activity relationships of scorpion alpha-neurotoxins: contribution of arginine residues." Toxicon **28**(5): 509-523.

Kharrat, R., H. Darbon, H. Rochat and C. Granier (1989). "Structure/activity relationships of scorpion alpha-toxins. Multiple residues contribute to the interaction with receptors." Eur J Biochem **181**(2): 381-390.

Kneller, D. G., F. E. Cohen and R. Langridge (1990). "Improvements in protein secondary structure prediction by an enhanced neural network." J Mol Biol **214**(1): 171-182.

Kozlov, S. and E. Grishin (2005). "Classification of spider neurotoxins using structural motifs by primary structure features. Single residue distribution analysis and pattern analysis techniques." Toxicon **46**(6): 672-686.

Kramer, A., T. Keitel, K. Winkler, W. Stocklein, W. Hohne and J. Schneider-Mergener (1997). "Molecular basis for the binding promiscuity of an anti-p24 (HIV-1) monoclonal antibody." Cell **91**(6): 799-809.

Kryshtafovych, A., C. Venclovas, K. Fidelis and J. Moult (2005). "Progress over the first decade of CASP experiments." Proteins **61 Suppl 7**: 225-236.

Larsen, T. A., A. J. Olson and D. S. Goodsell (1998). "Morphology of protein-protein interfaces." Structure **6**(4): 421-427.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Cryst. **26**: 283-291.

Le Nguyen, D., A. Heitz, L. Chiche, B. Castro, R. A. Boigegrain, A. Favel and M. A. Coletti-Previero (1990). "Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure." Biochimie **72**(6-7): 431-435.

Lee, B. and F. M. Richards (1971). "The interpretation of protein structures: estimation of static accessibility." J Mol Biol **55**(3): 379-400.

Legros, C., B. Ceard, H. Vacher, P. Marchot, P. E. Bougis and M. F. Martin-Eauclaire (2005). "Expression of the standard scorpion alpha-toxin AaH II and AaH II mutants leading to the identification of some key bioactive elements." Biochim Biophys Acta **1723**(1-3): 91-99.

Leipold, E., S. Lu, D. Gordon, A. Hansel and S. H. Heinemann (2004). "Combinatorial interaction of scorpion toxins Lqh-2, Lqh-3, and LqhalphalIT with sodium channel receptor sites-3." Mol Pharmacol **65**(3): 685-691.

Lesage, F. and M. Lazdunski (2000). "Molecular and functional properties of two-pore-domain potassium channels." Am J Physiol Renal Physiol **279**(5): F793-801.

Lewis, R. J. and M. L. Garcia (2003). "Therapeutic potential of venom peptides." Nat Rev Drug Discov **2**(10): 790-802.

Li-Smerin, Y. and K. J. Swartz (1998). "Gating modifier toxins reveal a conserved structural motif in voltage-gated Ca²⁺ and K⁺ channels." Proc Natl Acad Sci U S A **95**(15): 8585-8589.

Li, H. M., D. C. Wang, Z. H. Zeng, L. Jin and R. Q. Hu (1996). "Crystal structure of an acidic neurotoxin from scorpion *Buthus martensii* Karsch at 1.85 Å resolution." J Mol Biol **261**(3): 415-431.

Li, Y. J. and Y. H. Ji (2000). "Binding characteristics of BmK I, an alpha-like scorpion neurotoxic polypeptide, on cockroach nerve cord synaptosomes." J Pept Res **56**(4): 195-200.

Liang, S. (2004). "An overview of peptide toxins from the venom of the Chinese bird spider *Selenocosmia huwena* Wang [= *Ornithoctonus huwena* (Wang)]." Toxicon **43**(5): 575-585.

Liang, S., J. Zhang, S. Zhang and H. Guo (2004). "Prediction of the interaction site on the surface of an isolated protein structure by analysis of side chain energy scores." Proteins **57**(3): 548-557.

Lijnzaad, P. and P. Argos (1997). "Hydrophobic patches on protein subunit interfaces: characteristics and prediction." Proteins **28**(3): 333-343.

Lippens, G., J. Najib, S. J. Wodak and A. Tartar (1995). "NMR sequential assignments and solution structure of chlorotoxin, a small scorpion toxin that blocks chloride channels." Biochemistry **34**(1): 13-21.

Liu, G., Y. Shen, H. S. Atreya, D. Parish, Y. Shao, D. K. Sukumaran, R. Xiao, A. Yee, A. Lemak, A. Bhattacharya, T. A. Acton, C. H. Arrowsmith, G. T. Montelione and T. Szyperski (2005). "NMR data collection and analysis protocol for high-throughput protein structure determination." Proc Natl Acad Sci U S A **102**(30): 10487-10492.

Liu, L. H., F. Bosmans, C. Maertens, R. H. Zhu, D. C. Wang and J. Tytgat (2005). "Molecular basis of the mammalian potency of the scorpion alpha-like toxin, BmK M1." Faseb J **19**(6): 594-596.

Lo Conte, L., C. Chothia and J. Janin (1999). "The atomic structure of protein-protein recognition sites." J Mol Biol **285**(5): 2177-2198.

Long, S. B., E. B. Campbell and R. Mackinnon (2005). "Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel." Science **309**(5736): 897-903.

Long, S. B., E. B. Campbell and R. Mackinnon (2005). "Voltage sensor of Kv1.2: structural basis of electromechanical coupling." Science **309**(5736): 903-908.

Loret, E. P., R. M. del Valle, P. Mansuelle, F. Sampieri and H. Rochat (1994). "Positively charged amino acid residues located similarly in sea anemone and scorpion toxins." J Biol Chem **269**(24): 16785-16788.

Loret, E. P., P. Mansuelle, H. Rochat and C. Granier (1990). "Neurotoxins active on insects: amino acid sequences, chemical modifications, and secondary structure estimation by circular dichroism of toxins from the scorpion *Androctonus australis Hector*." Biochemistry **29**(6): 1492-1501.

Loughney, K., R. Kreber and B. Ganetzky (1989). "Molecular analysis of the para locus, a sodium channel gene in *Drosophila*." Cell **58**(6): 1143-1154.

Luban, J. and S. P. Goff (1995). "The yeast two-hybrid system for studying protein-protein interactions." Curr Opin Biotechnol **6**(1): 59-64.

Maertens, C., E. Cuypers, M. Amininasab, A. Jalali, H. Vatanpour and J. Tytgat (2006). "Potent modulation of the voltage-gated sodium channel Nav1.7 by OD1, a toxin from the scorpion *Odonthobuthus doriae*." Mol Pharmacol **70**(1): 405-414.

Manoleras, N. and R. S. Norton (1994). "Three-dimensional structure in solution of neurotoxin III from the sea anemone *Anemonia sulcata*." Biochemistry **33**(37): 11051-11061.

Marcotte, E. M., M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates and D. Eisenberg (1999). "Detecting protein function and protein-protein interactions from genome sequences." Science **285**(5428): 751-753.

Markus, M. A., K. T. Dayie, P. Matsudaira and G. Wagner (1994). "Effect of deuteration on the amide proton relaxation rates in proteins. Heteronuclear NMR experiments on villin 14T." J Magn Reson B **105**(2): 192-195.

Marti-Renom, M. A., U. Pieper, M. S. Madhusudhan, A. Rossi, N. Eswar, F. P. Davis, F. Al-Shahrour, J. Dopazo and A. Sali (2007). "DBAli tools: mining the protein structure space." Nucleic Acids Res **35**(Web Server issue): W393-397.

Marti-Renom, M. A., A. C. Stuart, A. Fiser, R. Sanchez, F. Melo and A. Sali (2000). "Comparative protein structure modeling of genes and genomes." Annu Rev Biophys Biomol Struct **29**: 291-325.

Martin, M. F., L. G. Garcia y Perez, M. el Ayeb, C. Kopeyan, G. Bechis, E. Jover and H. Rochat (1987). "Purification and chemical and biological characterizations of seven toxins from the Mexican scorpion, *Centruroides suffusus suffusus*." J Biol Chem **262**(10): 4452-4459.

Martin, M. F. and H. Rochat (1986). "Large scale purification of toxins from the venom of the scorpion *Androctonus australis Hector*." Toxicon **24**(11-12): 1131-1139.

Matavel, A., J. S. Cruz, C. L. Penaforte, D. A. Araujo, E. Kalapothakis, V. F. Prado, C. R. Diniz, M. N. Cordeiro and P. S. Beirao (2002). "Electrophysiological characterization and molecular identification of the *Phoneutria nigriventer* peptide toxin PnTx2-6." FEBS Lett **523**(1-3): 219-223.

Matavel, A., C. Fleury, L. Oliveira, F. Molina, M. E. De Lima, J. Cruz, M. Cordeiro, M. Richardson, C. Ramos and P. Beirao (2009). "Structure and activity analysis of two spider toxins that alter sodium channel inactivation kinetics." Biochemistry.

Maziere, P., C. Granier and F. Molina (2004). "A description scheme of biological processes based on elementary bricks of action." J Mol Biol **339**(1): 77-88.

McGuffin, L. J., K. Bryson and D. T. Jones (2000). "The PSIPRED protein structure prediction server." Bioinformatics **16**(4): 404-405.

McGuffin, L. J. and D. T. Jones (2003). "Improvement of the GenTHREADER method for genomic fold recognition." Bioinformatics **19**(7): 874-881.

Meloen, R. H., W. C. Puijk and J. W. Slootstra (2000). "Mimotopes: realization of an unlikely concept." J Mol Recognit **13**(6): 352-359.

Menez, A. (1998). "Functional architectures of animal toxins: a clue to drug design?" Toxicon **36**(11): 1557-1572.

Mer, G., E. Kellenberger and J. F. Lefevre (1998). "Alpha-helix mimicry of a beta-turn." J Mol Biol **281**(2): 235-240.

Middleton, R. E., V. A. Warren, R. L. Kraus, J. C. Hwang, C. J. Liu, G. Dai, R. M. Brochu, M. G. Kohler, Y. D. Gao, V. M. Garsky, M. J. Bogusky, J. T. Mehl, C. J. Cohen and M. M. Smith (2002). "Two tarantula peptides inhibit activation of multiple sodium channels." Biochemistry **41**(50): 14734-14747.

Mizuguchi, K., C. M. Deane, T. L. Blundell and J. P. Overington (1998). "HOMSTRAD: a database of protein structure alignments for homologous families." Protein Sci **7**(11): 2469-2471.

Moreau, V., C. Fleury, D. Piquer, C. Nguyen, N. Novali, S. Villard, D. Laune, C. Granier and F. Molina (2008). "PEPOP: computational design of immunogenic peptides." BMC Bioinformatics **9**: 71.

Moreau, V., C. Granier, S. Villard, D. Laune and F. Molina (2006). "Discontinuous epitope prediction based on mimotope analysis." Bioinformatics **22**(9): 1088-1095.

Moult, J., K. Fidelis, A. Kryshchuk, B. Rost, T. Hubbard and A. Tramontano (2007). "Critical assessment of methods of protein structure prediction-Round VII." Proteins **69 Suppl 8**: 3-9.

Mourier, G., D. Servent, S. Zinn-Justin and A. Menez (2000). "Chemical engineering of a three-fingered toxin with anti- $\alpha 7$ neuronal acetylcholine receptor activity." Protein Eng **13**(3): 217-225.

Mumey, B. M., B. W. Bailey, B. Kirkpatrick, A. J. Jesaitis, T. Angel and E. A. Dratz (2003). "A new method for mapping discontinuous antibody epitopes to reveal structural features of proteins." J Comput Biol **10**(3-4): 555-567.

Murzin, A. G., S. E. Brenner, T. Hubbard and C. Chothia (1995). "SCOP: a structural classification of proteins database for the investigation of sequences and structures." J Mol Biol **247**(4): 536-540.

Narasimhan, L., J. Singh, C. Humblet, K. Guruprasad and T. Blundell (1994). "Snail and spider toxins share a similar tertiary structure and 'cystine motif'." Nat Struct Biol **1**(12): 850-852.

Neuvirth, H., R. Raz and G. Schreiber (2004). "ProMate: a structure based prediction program to identify the location of protein-protein binding sites." J Mol Biol **338**(1): 181-199.

Norel, R., F. Sheinerman, D. Petrey and B. Honig (2001). "Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis." Protein Sci **10**(11): 2147-2161.

Norton, R. S. and P. K. Pallaghy (1998). "The cystine knot structure of ion channel toxins and related polypeptides." Toxicon **36**(11): 1573-1583.

Novotny, J., M. Handschumacher, E. Haber, R. E. Bruccoleri, W. B. Carlson, D. W. Fanning, J. A. Smith and G. D. Rose (1986). "Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains)." Proc Natl Acad Sci U S A **83**(2): 226-230.

Odorico, M. and J. L. Pellequer (2003). "BEPITOPE: predicting the location of continuous epitopes and patterns in proteins." J Mol Recognit **16**(1): 20-22.

Ohlson, T., B. Wallner and A. Elofsson (2004). "Profile-profile methods provide improved fold-recognition: a study of different profile-profile alignment methods." Proteins **57**(1): 188-197.

Olamendi-Portugal, T., B. I. Garcia, I. Lopez-Gonzalez, J. Van Der Walt, K. Dyason, C. Ulens, J. Tytgat, R. Felix, A. Darszon and L. D. Possani (2002). "Two new scorpion toxins that target voltage-gated Ca²⁺ and Na⁺ channels." Biochem Biophys Res Commun **299**(4): 562-568.

Olmea, O. and A. Valencia (1997). "Improving contact predictions by the combination of correlated mutations and other sources of sequence information." Fold Des **2**(3): S25-32.

Omecinsky, D. O., K. E. Holub, M. E. Adams and M. D. Reily (1996). "Three-dimensional structure analysis of mu-agatoxins: further evidence for common motifs among neurotoxins with diverse ion channel specificities." Biochemistry **35**(9): 2836-2844.

Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells and J. M. Thornton (1997). "CATH--a hierarchic classification of protein domain structures." Structure **5**(8): 1093-1108.

Ortiz, A. R., C. E. Strauss and O. Olmea (2002). "MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison." Protein Sci **11**(11): 2606-2621.

Ostrow, K. L., A. Mammoser, T. Suchyna, F. Sachs, R. Oswald, S. Kubo, N. Chino and P. A. Gottlieb (2003). "cDNA sequence and in vitro folding of GsMTx4, a specific peptide inhibitor of mechanosensitive channels." Toxicon **42**(3): 263-274.

Overbeek, R., M. Fonstein, M. D'Souza, G. D. Pusch and N. Maltsev (1999). "Use of contiguity on the chromosome to predict functional coupling." In Silico Biol **1**(2): 93-108.

Padlan, E. A. (1990). "On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands." Proteins **7**(2): 112-124.

Pallaghy, P. K., K. J. Nielsen, D. J. Craik and R. S. Norton (1994). "A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides." Protein Sci **3**(10): 1833-1839.

Parker, J. M., D. Guo and R. S. Hodges (1986). "New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites." Biochemistry **25**(19): 5425-5432.

Pervushin, K., R. Riek, G. Wider and K. Wuthrich (1997). "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." Proc Natl Acad Sci U S A **94**(23): 12366-12371.

Petrey, D. and B. Honig (2003). "GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences." Methods Enzymol **374**: 492-509.

Pimenta, A. M. and M. E. De Lima (2005). "Small peptides, big world: biotechnological potential in neglected bioactive peptides from arthropod venoms." J Pept Sci **11**(11): 670-676.

Pizzi, E., R. Cortese and A. Tramontano (1995). "Mapping epitopes on protein surfaces." Biopolymers **36**(5): 675-680.

Plewczynski, D. and K. Ginalski (2009). "The interactome: predicting the protein-protein interactions in cells." Cell Mol Biol Lett **14**(1): 1-22.

Possani, L. D., B. Becerril, M. Delepierre and J. Tytgat (1999). "Scorpion toxins specific for Na⁺-channels." Eur J Biochem **264**(2): 287-300.

Possani, L. D., E. Merino, M. Corona, F. Bolivar and B. Becerril (2000). "Peptides and genes coding for scorpion toxins that affect ion-channels." Biochimie **82**(9-10): 861-868.

Qian, B., S. Raman, R. Das, P. Bradley, A. J. McCoy, R. J. Read and D. Baker (2007). "High-resolution structure prediction and the crystallographic phase problem." Nature **450**(7167): 259-264.

Retief, J. D. (2000). "Phylogenetic analysis using PHYLIP." Methods Mol Biol **132**: 243-258.

Revillard, J. P., L. Adorini, M. Goldman, D. Kabelitz and H. Waldmann (1998). "Apoptosis: potential for disease therapies." Immunol Today **19**(7): 291-293.

Ricciardi, A., M. H. le Du, M. Khayati, F. Dajas, J. C. Boulain, A. Menez and F. Ducancel (2000). "Do structural deviations between toxins adopting the same fold reflect functional differences?" J Biol Chem **275**(24): 18302-18310.

Rieping, W., M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin and M. Nilges (2007). "ARIA2: automated NOE assignment and data integration in NMR structure calculation." Bioinformatics **23**(3): 381-382.

Rochat, H., C. Rochat, F. Sampieri, F. Miranda and S. Lissitzky (1972). "The amino-acid sequence of neurotoxin II of *Androctonus australis* Hector." Eur J Biochem **28**(3): 381-388.

Rochfeld, A. and J. Morejon (1989). La méthode Merise - Tome 3 Gamme opératoire. Paris, Editions d'organisation.

Rodriguez de la Vega, R. C. and L. D. Possani (2004). "Current views on scorpion toxins specific for K⁺-channels." Toxicon **43**(8): 865-875.

Rodriguez de la Vega, R. C. and L. D. Possani (2005). "Overview of scorpion toxins specific for Na⁺ channels and related peptides: biodiversity, structure-function relationships and evolution." Toxicon **46**(8): 831-844.

Rogers, J. C., Y. Qu, T. N. Tanada, T. Scheuer and W. A. Catterall (1996). "Molecular determinants of high affinity binding of alpha-scorpion toxin and sea anemone toxin in the S3-S4 extracellular loop in domain IV of the Na⁺ channel alpha subunit." J Biol Chem **271**(27): 15950-15962.

Rohl, C. A., C. E. Strauss, K. M. Misura and D. Baker (2004). "Protein structure prediction using Rosetta." Methods Enzymol **383**: 66-93.

Rose, G. D. and R. Wolfenden (1993). "Hydrogen bonding, hydrophobicity, packing, and protein folding." Annu Rev Biophys Biomol Struct **22**: 381-415.

Rosengren, K. J., D. Wilson, N. L. Daly, P. F. Alewood and D. J. Craik (2002). "Solution structures of the cis- and trans-Pro30 isomers of a novel 38-residue toxin from the venom of *Hadronyche infensa* sp. that contains a cystine-knot motif within its four disulfide bonds." Biochemistry **41**(10): 3294-3301.

Rost, B., C. Sander and R. Schneider (1994). "PHD--an automatic mail server for protein secondary structure prediction." Comput Appl Biosci **10**(1): 53-60.

Rost, B., C. Sander and R. Schneider (1994). "Redefining the goals of protein secondary structure prediction." J Mol Biol **235**(1): 13-26.

Rost, B., G. Yachdav and J. Liu (2004). "The PredictProtein server." Nucleic Acids Res **32**(Web Server issue): W321-326.

Rudolf, M. P., M. Vogel, F. Kricek, C. Ruf, A. W. Zurcher, R. Reuschel, M. Auer, S. Miescher and B. M. Stadler (1998). "Epitope-specific antibody response to IgE by mimotope immunization." J Immunol **160**(7): 3315-3321.

Rychlewski, L., L. Jaroszewski, W. Li and A. Godzik (2000). "Comparison of sequence profiles. Strategies for structural predictions using sequence information." Protein Sci **9**(2): 232-241.

Sadreyev, R. I. and N. V. Grishin (2004). "Estimates of statistical significance for comparison of individual positions in multiple sequence alignments." BMC Bioinformatics **5**: 106.

Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Mol Biol Evol **4**(4): 406-425.

Saphire, E. O., P. W. Parren, R. Pantophlet, M. B. Zwick, G. M. Morris, P. M. Rudd, R. A. Dwek, R. L. Stanfield, D. R. Burton and I. A. Wilson (2001). "Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design." Science **293**(5532): 1155-1159.

Saunal, H. and M. H. Van Regenmortel (1995). "Mapping of viral conformational epitopes using biosensor measurements." J Immunol Methods **183**(1): 33-41.

Sautiere, P., S. Cestele, C. Kopeyan, A. Martinage, H. Drobecq, Y. Doljansky and D. Gordon (1998). "New toxins acting on sodium channels from the scorpion *Leiurus*

quinquestriatus hebraeus suggest a clue to mammalian vs insect selectivity." Toxicon **36**(8): 1141-1154.

Schaffer, A. A., L. Aravind, T. L. Madden, S. Shavirin, J. L. Spouge, Y. I. Wolf, E. V. Koonin and S. F. Altschul (2001). "Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements." Nucleic Acids Res **29**(14): 2994-3005.

Sheshberadaran, H. and L. G. Payne (1988). "Protein antigen-monoclonal antibody contact sites investigated by limited proteolysis of monoclonal antibody-bound antigen: protein "footprinting"." Proc Natl Acad Sci U S A **85**(1): 1-5.

Shichor, I., E. Zlotkin, N. Ilan, D. Chikashvili, W. Stuhmer, D. Gordon and I. Lotan (2002). "Domain 2 of Drosophila para voltage-gated sodium channel confers insect properties to a rat brain channel." J Neurosci **22**(11): 4364-4371.

Siddiqui, A. S., U. Dengler and G. J. Barton (2001). "3Dee: a database of protein structural domains." Bioinformatics **17**(2): 200-201.

Siemens, J., S. Zhou, R. Piskorowski, T. Nikai, E. A. Lumpkin, A. I. Basbaum, D. King and D. Julius (2006). "Spider toxins activate the capsaicin receptor to produce inflammatory pain." Nature **444**(7116): 208-212.

Smith, G. P. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." Science **228**(4705): 1315-1317.

Smith, G. R. and M. J. Sternberg (2002). "Prediction of protein-protein interactions by docking methods." Curr Opin Struct Biol **12**(1): 28-35.

Sollod, B. L., D. Wilson, O. Zhaxybayeva, J. P. Gogarten, R. Drinkwater and G. F. King (2005). "Were arachnids the first to use combinatorial peptide libraries?" Peptides **26**(1): 131-139.

Srinivasan, K. N., V. Sivaraja, I. Huys, T. Sasaki, B. Cheng, T. K. Kumar, K. Sato, J. Tytgat, C. Yu, B. C. San, S. Ranganathan, H. J. Bowie, R. M. Kini and P. Gopalakrishnakone (2002). "kappa-Hefutoxin1, a novel toxin from the scorpion *Heterometrus fulvipes* with unique structure and function. Importance of the functional diad in potassium channel selectivity." J Biol Chem **277**(33): 30040-30047.

Stephen, C. W., P. Helminen and D. P. Lane (1995). "Characterisation of epitopes on human p53 using phage-displayed peptide libraries: insights into antibody-peptide interactions." J Mol Biol **248**(1): 58-78.

Sun, Y. M., F. Bosmans, R. H. Zhu, C. Goudet, Y. M. Xiong, J. Tytgat and D. C. Wang (2003). "Importance of the conserved aromatic residues in the scorpion alpha-like toxin BmK M1: the hydrophobic surface region revisited." J Biol Chem **278**(26): 24125-24131.

Swartz, K. J. and R. MacKinnon (1995). "An inhibitor of the Kv2.1 potassium channel isolated from the venom of a Chilean tarantula." Neuron **15**(4): 941-949.

Takahashi, H., J. I. Kim, H. J. Min, K. Sato, K. J. Swartz and I. Shimada (2000). "Solution structure of hanatoxin1, a gating modifier of voltage-dependent K(+) channels: common surface features of gating modifier toxins." J Mol Biol **297**(3): 771-780.

Tan, P. T., A. Veeramani, K. N. Srinivasan, S. Ranganathan and V. Brusica (2006). "SCORPION2: a database for structure-function analysis of scorpion toxins." Toxicon **47**(3): 356-363.

Tardieu, H., A. Rochfeld and R. Colletti (1983). La méthode Merise - Tome 1 Principes et outils. Paris, Editions d'organisation.

Tardieu, H., A. Rochfeld, R. Colletti, G. Panet and G. Vahée (1985). La méthode Merise - Tome 2 Démarches et pratiques. Paris, Editions d'organisation.

Tedford, H. W., N. Gilles, A. Menez, C. J. Doering, G. W. Zamponi and G. F. King (2004). "Scanning mutagenesis of omega-atracotoxin-Hv1a reveals a spatially restricted epitope that confers selective activity against insect calcium channels." J Biol Chem **279**(42): 44133-44140.

Thompson, J. D., T. J. Gibson and D. G. Higgins (2002). "Multiple sequence alignment using ClustalW and ClustalX." Curr Protoc Bioinformatics **Chapter 2**: Unit 2 3.

Thornton, J. M., M. S. Edwards, W. R. Taylor and D. J. Barlow (1986). "Location of 'continuous' antigenic determinants in the protruding regions of proteins." Embo J **5**(2): 409-413.

Tong, A. H., B. Drees, G. Nardelli, G. D. Bader, B. Brannetti, L. Castagnoli, M. Evangelista, S. Ferracuti, B. Nelson, S. Paoluzi, M. Quondam, A. Zucconi, C. W. Hogue, S. Fields, C. Boone and G. Cesareni (2002). "A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules." Science **295**(5553): 321-324.

Trakselis, M. A., S. C. Alley and F. T. Ishmael (2005). "Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics." Bioconj Chem **16**(4): 741-750.

Tress, M., I. Ezkurdia, O. Grana, G. Lopez and A. Valencia (2005). "Assessment of predictions submitted for the CASP6 comparative modeling category." Proteins **61 Suppl 7**: 27-45.

Tsai, C. J., S. L. Lin, H. J. Wolfson and R. Nussinov (1996). "Protein-protein interfaces: architectures and interactions in protein-protein interfaces and in protein cores. Their similarities and differences." Crit Rev Biochem Mol Biol **31**(2): 127-152.

Tsai, C. J., S. L. Lin, H. J. Wolfson and R. Nussinov (1997). "Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect." Protein Sci **6**(1): 53-64.

Valadon, P., G. Nussbaum, L. F. Boyd, D. H. Margulies and M. D. Scharff (1996). "Peptide libraries define the fine specificity of anti-polysaccharide antibodies to *Cryptococcus neoformans*." J Mol Biol **261**(1): 11-22.

Van Regenmortel, M. H. (2001). "Antigenicity and immunogenicity of synthetic peptides." Biologicals **29**(3-4): 209-213.

Van Regenmortel, M. H. V. (1996). "Mapping Epitope Structure and Activity: From One-Dimensional Prediction to Four-Dimensional Description of Antigenic Specificity." Methods **9**(3): 465-472.

Vargas, O., M. F. Martin and H. Rochat (1987). "Characterization of six toxins from the venom of the Moroccan scorpion *Buthus occitanus mardochei*." Eur J Biochem **162**(3): 589-599.

Vita, C., E. Drakopoulou, J. Vizzavona, S. Rochette, L. Martin, A. Menez, C. Roumestand, Y. S. Yang, L. Ylisastigui, A. Benjouad and J. C. Gluckman (1999). "Rational engineering of a miniprotein that reproduces the core of the CD4 site

interacting with HIV-1 envelope glycoprotein." Proc Natl Acad Sci U S A **96**(23): 13091-13096.

Vita, C., C. Roumestand, F. Toma and A. Menez (1995). "Scorpion toxins as natural scaffolds for protein engineering." Proc Natl Acad Sci U S A **92**(14): 6404-6408.

Vriend, G. (1990). "WHAT IF: a molecular modeling and drug design program." J Mol Graph **8**(1): 52-56, 29.

Wang, C. G., N. Gilles, A. Hamon, F. Le Gall, M. Stankiewicz, M. Pelhate, Y. M. Xiong, D. C. Wang and C. W. Chi (2003). "Exploration of the functional site of a scorpion alpha-like toxin by site-directed mutagenesis." Biochemistry **42**(16): 4699-4708.

Wang, G. and R. L. Dunbrack, Jr. (2004). "Scoring profile-to-profile sequence alignments." Protein Sci **13**(6): 1612-1626.

Wang, X., M. Connor, R. Smith, M. W. Maciejewski, M. E. Howden, G. M. Nicholson, M. J. Christie and G. F. King (2000). "Discovery and characterization of a family of insecticidal neurotoxins with a rare vicinal disulfide bridge." Nat Struct Biol **7**(6): 505-513.

Wang, X. H., M. Connor, D. Wilson, H. I. Wilson, G. M. Nicholson, R. Smith, D. Shaw, J. P. Mackay, P. F. Alewood, M. J. Christie and G. F. King (2001). "Discovery and structure of a potent and highly specific blocker of insect calcium channels." J Biol Chem **276**(43): 40306-40312.

Wang, Y., S. Bryant, R. Tatusov and T. Tatusova (2000). "Links from genome proteins to known 3-D structures." Genome Res **10**(10): 1643-1647.

Warrell, D. A. (1993). "Venomous bites and stings in the tropical world." Med J Aust **159**(11-12): 773-779.

Welling, G. W., W. J. Weijer, R. van der Zee and S. Welling-Wester (1985). "Prediction of sequential antigenic regions in proteins." FEBS Lett **188**(2): 215-218.

Whetstone, P. A. and B. D. Hammock (2007). "Delivery methods for peptide and protein toxins in insect control." Toxicon **49**(4): 576-596.

White, J. (2000). "Bites and stings from venomous animals: a global overview." Ther Drug Monit **22**(1): 65-68.

White, J., D. Warrell, M. Eddleston, B. J. Currie, I. M. Whyte and G. K. Isbister (2003). "Clinical toxinology--where are we now?" J Toxicol Clin Toxicol **41**(3): 263-276.

Wicher, D. and H. Penzlin (1998). "omega-Toxins affect Na⁺ currents in neurosecretory insect neurons." Receptors Channels **5**(6): 355-366.

Wiederstein, M. and M. J. Sippl (2007). "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins." Nucleic Acids Res **35**(Web Server issue): W407-410.

Wun, H. L., D. T. Leung, K. C. Wong, Y. L. Chui and P. L. Lim (2001). "Molecular mimicry: anti-DNA antibodies may arise inadvertently as a response to antibodies generated to microorganisms." Int Immunol **13**(9): 1099-1107.

Xiao, Y. and S. Liang (2003). "Inhibition of neuronal tetrodotoxin-sensitive Na⁺ channels by two spider toxins: hainantoxin-III and hainantoxin-IV." Eur J Pharmacol **477**(1): 1-7.

Xiao, Y., J. Tang, W. Hu, J. Xie, C. Maertens, J. Tytgat and S. Liang (2005). "Jingzhaotoxin-I, a novel spider neurotoxin preferentially inhibiting cardiac sodium channel inactivation." J Biol Chem **280**(13): 12069-12076.

Xiao, Y., J. Tang, Y. Yang, M. Wang, W. Hu, J. Xie, X. Zeng and S. Liang (2004). "Jingzhaotoxin-III, a novel spider toxin inhibiting activation of voltage-gated sodium channel in rat cardiac myocytes." J Biol Chem **279**(25): 26220-26226.

Young, L., R. L. Jernigan and D. G. Covell (1994). "A role for surface hydrophobicity in protein-protein recognition." Protein Sci **3**(5): 717-729.

Yu, F. H. and W. A. Catterall (2004). "The VGL-chanome: a protein superfamily specialized for electrical signaling and ionic homeostasis." Sci STKE **2004**(253): re15.

Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein and M. Snyder (2001). "Global analysis of protein activities using proteome chips." Science **293**(5537): 2101-2105.

Zilberberg, N., O. Froy, E. Loret, S. Cestele, D. Arad, D. Gordon and M. Gurevitz (1997). "Identification of structural elements of a scorpion alpha-neurotoxin important for receptor site recognition." J Biol Chem **272**(23): 14810-14816.

Zlotkin, E., M. Gurevitz, E. Fowler and M. E. Adams (1993). "Depressant insect selective neurotoxins from scorpion venom: chemistry, action, and gene cloning." Arch Insect Biochem Physiol **22**(1-2): 55-73.

APÊNDICES

Apêndice 1: Dicionário de dados do SiliTox

Apêndice 2: *Scripts* adaptados do Modeller

Apêndice 3: Grupos estruturais para SiliMod – alinhamentos

Apêndice 4: Grupos estruturais para SiliMod – árvores *Neighbor-Joining*

Apêndice 5: Artigo publicado 1: “*PEPOP: Computational design of immunogenic peptides*”

Apêndice 6: Artigo publicado 2: “*Structure and activity analysis of two spider toxins that alter sodium channel inactivation kinetics*”

APÊNDICE 1: Dicionário de dados do SiliTox

Author

Field	Type	Null	Default	Comment
<u>autID</u>	int(10)	No		author unique ID
name	varchar(50)	No		author second name
firstname	varchar(30)	No		author first name

BA

Field	Type	Null	Default	Comment
<u>baID</u>	int(10)	No		biological activity unique ID

BAofBF

Field	Type	Null	Default	Comment
<u>babfIDbf</u>	int(10)	No		ID of the biological function that includes the biological activity - refers to BF table
<u>babfIDba</u>	int(10)	No		ID of the biological activity played inside the biological function - refers to BA table

BEA

Field	Type	Null	Default	Comment
<u>beaID</u>	int(10)	No		basic element of action unique ID

BEAofBA

Field	Type	Null	Default	Comment
<u>beabaIDba</u>	int(10)	No		ID of the biological activity to which the basic element of action contributes - refers to BA table
<u>beabaIDbea</u>	int(10)	No		ID of the basic element of action involved in the biological activity - refers to BEA table

BF

Field	Type	Null	Default	Comment
<u>bfID</u>	int(10)	No		biological function unique ID

BioEntity

Field	Type	Null	Default	Comment
<u>entityID</u>	int(10)	No		bioentity unique ID
molnb	int(10)	Yes	NULL	number of biomolecules composing the bioentity

BioMolecule

Field	Type	Null	Default	Comment
<u>molID</u>	int(10)	No		biomolecule unique ID
type	enum('protein', 'nucleicacid', 'cDNA', 'chemicalcompound', 'sugar', 'ion')	No		type of biomolecule
subtype	varchar(100)	Yes	NULL	subtype of biomolecule
iontransport	varchar(30)	Yes	NULL	type of ion transported by

				the channel
channelsubtype	varchar(50)	Yes	NULL	channel subtype
family	int(10)	Yes	NULL	ID of the family to which belongs the biomolecule - refers to Family table
sequence	text	Yes	NULL	amino acid or nucleotide sequence
fasta	varchar(20)	Yes	NULL	biomolecule code in Swiss-Prot fasta format
length	int(10)	Yes	NULL	biomolecule sequence length
species	int(10)	Yes	NULL	ID of the species to which belongs the biomolecule - refers to Species table
genericname	varchar(100)	Yes	NULL	biomolecule generic name
shortname	varchar(100)	Yes	NULL	biomolecule short name
othername	varchar(100)	Yes	NULL	biomolecule other name
othername2	varchar(100)	Yes	NULL	biomolecule second other name
othername3	varchar(100)	Yes	NULL	biomolecule third other name
othername4	varchar(100)	Yes	NULL	biomolecule fourth other name
swissp	varchar(20)	Yes	NULL	biomolecule Swiss-Prot access number
swissp2	varchar(20)	Yes	NULL	biomolecule Swiss-Prot second access number
swissp3	varchar(20)	Yes	NULL	biomolecule Swiss-Prot third access number

pfam	vchar(20)	Yes	NULL	biomolecule Pfam access number
pdb	vchar(25)	Yes	NULL	biomolecule PDB code
pdb2	vchar(25)	Yes	NULL	biomolecule second PDB code
pdb3	vchar(25)	Yes	NULL	biomolecule third PDB code
3Dmethod	vchar(100)	Yes	NULL	biomolecule 3D determination method
otheracnumb	vchar(30)	Yes	NULL	other access number referring to the biomolecule
knottin	vchar(200)	Yes	NULL	biomolecule classification as a knottin
2Dstruc	vchar(200)	Yes	NULL	
3Dstruc	vchar(20)	Yes	NULL	biomolecule 3D structure file name
chain	vchar(10)	Yes	NULL	biomolecule chain code in 3D structure file
cyspattern	int(10)	Yes	NULL	ID of the of biomolecule cysteine pattern - refers to CysPattern table
fold	int(10)	Yes	NULL	ID of the of biomolecule fold - refers to Fold table
similarity	text	Yes	NULL	similarity field of the biomolecule in Swiss-Prot
toxicity	text	Yes	NULL	toxicity field of the biomolecule in Swiss-Prot
comment	text	Yes	NULL	function field of the biomolecule in Swiss-Prot

clustertree	varchar(20)	Yes	NULL	file name of biomolecule surface accessible segment clustering by SiliPop
-------------	-------------	-----	------	---

ClinicalApplication

Field	Type	Null	Default	Comment
<u>appID</u>	int(10)	No		application unique ID
apptype	enum('biotechno', 'pharmaco', 'other')	No		type of clinical application
description	text	Yes	NULL	description of clinical application

Cluster

Field	Type	Null	Default	Comment
<u>clusterID</u>	Int(10)	No		cluster unique ID
cluname	varchar(20)	Yes	NULL	cluster name
molcluster	Int(10)	Yes	NULL	ID of the biomolecule to which belongs the cluster - refers to BioMolecule table
cluclassLEN	Int(10)	Yes	NULL	cluster ranking as for length criteria
cluclassACC	Int(10)	Yes	NULL	cluster ranking as for accessibility criteria
cluclassHYD	Int(10)	Yes	NULL	cluster ranking as for hydrophobicity criteria
cluclassTURN	Int(10)	Yes	NULL	cluster ranking as for beta-turn criteria
cluclassWRYP	Int(10)	Yes	NULL	cluster ranking as for amino acid W, R, Y, P criteria
cluscoreLEN	Int(10)	Yes	NULL	cluster score as for length criteria

cluscoreACC	Int(10)	Yes	NULL	cluster score as for accessibility criteria
cluscoreHYD	Int(10)	Yes	NULL	cluster score as for hydrophobicity criteria
cluscoreTURN	Int(10)	Yes	NULL	cluster score as for beta-turn criteria
cluscoreWRYP	Int(10)	Yes	NULL	cluster score as for amino acid W, R, Y, P criteria

CysPattern

Field	Type	Null	Default	Comment
<u>cyspattID</u>	int(10)	No		cysteine pattern unique ID
cyspattern	varchar(200)	Yes	NULL	cysteine pattern in PROSITE format
cysnumber	int(10)	Yes	NULL	number of cysteine residues
kozlov	varchar(10)	Yes	NULL	cysteine pattern classification according to Kozlov and Grishin
ICKclass	varchar(20)	Yes	NULL	cysteine pattern classification for knottins
Nclass	varchar(20)	Yes	NULL	cysteine pattern classification for not knottins
cyscomment	text	Yes	NULL	comment about the cysteine pattern

Distance

Field	Type	Null	Default	Comment
<u>distIDseg1</u>	int(10)	No	0	ID of the first segment - refers to Segment table
<u>distIDseg2</u>	int(10)	No	0	ID of the second segment - refers to Segment table

seg1name	varchar(20)	Yes	NULL	first segment name
seg2name	varchar(20)	Yes	NULL	second segment name
moldist	int(10)	Yes	NULL	ID of the biomolecule to which belong the segments - refers to BioMolecule table
N_N	Float	Yes	NULL	distance from N-term extremity of first segment to N-term extremity of second segment
N_M	Float	Yes	NULL	distance from N-term extremity of first segment to the middle of second segment
N_C	Float	Yes	NULL	distance from N-term extremity of first segment to C-term extremity of second segment
M_N	Float	Yes	NULL	distance from the middle of first segment to N-term extremity of second segment
M_M	Float	Yes	NULL	distance from the middle of first segment to the middle of second segment
M_C	Float	Yes	NULL	distance from the middle of first segment to C-term extremity of second segment
C_N	Float	Yes	NULL	distance from C-term extremity of first segment to N-term extremity of second segment
C_M	Float	Yes	NULL	distance from C-term extremity of first segment to the middle of second segment
C_C	Float	Yes	NULL	distance from C-term extremity of first segment to C-term extremity of second segment
distminvalue	Float	Yes	NULL	minimum distance value between the two segments
distminnature	varchar(10)	Yes	NULL	type of distance corresponding to

				minimum value
--	--	--	--	---------------

DisuMolecule

Field	Type	Null	Default	Comment
<u>disumolIDmol</u>	int(10)	No	0	ID of the biomolecule to which belongs the disulfide bridge - refers to BioMolecule table
<u>disumolIDdis</u>	int(10)	No	0	ID of the disulfide bridge - refers to Disulfide table
disuquality	varchar(50)	Yes	NULL	quality of the disulfide bridge describing if it has been experimentally determined or predicted

Disulfide

Field	Type	Null	Default	Comment
<u>disID</u>	int(10)	No		disulfide bridge unique ID
CCsequence	varchar(50)	Yes	NULL	sequence position of disulfide bridge forming cysteine pair of residues
CCrelative	varchar(50)	Yes	NULL	relative position of disulfide bridge forming cysteine pair of residues
CCconvICK	varchar(50)	Yes	NULL	cysteine residue pair numbering for disulfide bridges in the ICK motif, according to ICK numbering convention
CCconvNOick	varchar(50)	Yes	NULL	cysteine residue pair numbering for disulfide bridges out of ICK motif

EntityPlayBF

Field	Type	Null	Default	Comment
<u>ebfIDent</u>	Int(10)	No		ID of the bioentity involved in the biological function - refers to BioEntity table
<u>ebfIDbf</u>	Int(10)	No		ID of the biological function played by the bioentity - refers to BF table

Family

Field	Type	Null	Default	Comment
<u>famID</u>	int(10)	No		family unique ID
famname	varchar(100)	No		family name

Fold

Field	Type	Null	Default	Comment
<u>foldID</u>	int(10)	No		fold unique ID
scop	varchar(30)	Yes	NULL	fold class according to SCOP
cath	varchar(30)	Yes	NULL	fold class according to CATH
dali	varchar(30)	Yes	NULL	fold class according to DALI
comment	text	Yes	NULL	comment about the fold

FuncDomInMol

Field	Type	Null	Default	Comment
<u>fdmolIDmol</u>	int(10)	No		ID of the biomolecule owning the functional domain - refers to BioMolecule table

<u>fdmollIDdom</u>	int(10)	No		ID of the functional domain owned by the biomolecule - refers to FuncDomain
--------------------	---------	----	--	---

FuncDomPlayBA

Field	Type	Null	Default	Comment
<u>fdbaIDdom</u>	int(10)	No		ID of the functional domain involved in the biological activity - refers to FuncDomain table
<u>fdbaIDba</u>	int(10)	No		ID of the biological activity in which the functional domain is involved - refers to BA table

FuncDomain

Field	Type	Null	Default	Comment
<u>funcdomID</u>	int(10)	No		functional domain unique ID

FuncMotInDom

Field	Type	Null	Default	Comment
<u>fmdomIDdom</u>	int(10)	No		ID of the functional domain which contain the functional motif - refers to FuncDomain table
<u>fmdomIDmot</u>	int(10)	No		ID of the functional motif included in the functional domain - refers to FuncMotif table

FuncMotPlayBEA

Field	Type	Null	Default	Comment
-------	------	------	---------	---------

<u>fmbeaIDmot</u>	int(10)	No		ID of the functional motif responsible for the basic element of action - refers to FuncMotif table
<u>fmbeaIDbea</u>	int(10)	No		ID of the basic element of action realized by the functional motif - refers to BEA table

FuncMotif

Field	Type	Null	Default	Comment
<u>funcmotID</u>	Int(10)	No		functional motif unique ID

Interact

Field	Type	Null	Default	Comment
<u>chatoxIDcha</u>	int(10)	No		ID of the ion channel - refers to BioMolecule table
<u>chatoxIDtox</u>	int(10)	No		ID of the toxin - refers to BioMolecule table
bindingsite	varchar(100)	Yes	NULL	binding site on the channel
interactquality	varchar(100)	Yes	NULL	quality of the interaction describing if it has been experimentally determined or predicted
receptorAA	varchar(100)	Yes	NULL	amino acid residues involved in the interaction on the receptor side (ion channel)
bioactiveAA	varchar(100)	Yes	NULL	amino acid residues involved in the interaction on the toxin side
affinityconst	varchar(50)	Yes	NULL	affinity constant
paratopeAA	varchar(100)	Yes	NULL	amino acid residues of the antibody involved in the interaction with the

				biomolecule
epitopeAA	varchar(100)	Yes	NULL	amino acid residues of the biomolecule recognized by the antibody

IsTemplate

Field	Type	Null	Default	Comment
tempIDmol	int(10)	No		ID of the template biomolecule - refers to BioMolecule table
targetIDmol	int(10)	No		ID of the target biomolecule - refers to BioMolecule table
modelIDmod	int(10)	No		ID of the 3D model - refers to Model table

Model

Field	Type	Null	Default	Comment
<u>modelID</u>	int(10)	No		model unique ID
modIDmol	int(10)	No		ID of the biomolecule to which corresponds the model
modelfile	varchar(30)	No		model file name
alifile	varchar(30)	No		alignment file name
MOF	double	No		Modeller Objective Function value

MolComposeEntity

Field	Type	Null	Default	Comment
<u>compIDent</u>	int(10)	No		ID of the bioentity including the biomolecule - refers to BioEntity table

<u>compIDmol</u>	int(10)	No		ID of the biomolecule composing the bioentity - refers to BioMolecule table
------------------	---------	----	--	---

NatEffByBF

Field	Type	Null	Default	Comment
<u>nebfIDeff</u>	int(10)	No		ID of the natural effect observed together with the biological function - refers to NaturalEffect table
<u>nebfIDbf</u>	int(10)	No		ID of the biological function related to the observed natural effect - refers to BF table

NatEffByMolecule

Field	Type	Null	Default	Comment
<u>nemolIDmol</u>	int(10)	No		ID of the biomolecule causing the observed natural effect - refers to BioMolecule table
<u>nemolIDeff</u>	int(10)	No		ID of the natural effect observed for the biomolecule - refers to NaturalEffect table

NaturalEffect

Field	Type	Null	Default	Comment
<u>effID</u>	int(10)	No		natural effect unique ID
action	text	Yes	NULL	description of natural effect
animal	text	Yes	NULL	animal on which the natural effect are observed
LD50	varchar(20)	Yes	NULL	lethal dose

OriMolecule

Field	Type	Null	Default	Comment
<u>fromIDori</u>	int(10)	No		ID of the origin of the biomolecule - refers to Origin table
<u>fromIDmol</u>	int(10)	No		ID of the biomolecule coming from the origin - refers to BioMolecule table

Origin

Field	Type	Null	Default	Comment
<u>oriID</u>	int(10)	No		biomolecule origin unique ID
tissuename	varchar(50)	Yes	NULL	name of the tissue or gland in which the biomolecule is naturally produced
subcellloc	varchar(50)	Yes	NULL	subcellular location of the biomolecule
tissuespe	text	Yes	NULL	biomolecule tissue specificity

PaperRefEntity

Field	Type	Null	Default	Comment
<u>papIDref</u>	int(10)	No		ID of the bibliographic reference citing the bioentity - refers to RefBiblio table
<u>papIDent</u>	int(10)	No		ID of the bioentity cited in the bibliographic reference - refers to BioEntity table

PaperRefMolecule

Field	Type	Null	Default	Comment
<u>papIDref</u>	Int(10)	No		ID of the bibliographic reference citing the biomolecule - refers to RefBiblio table

<u>papIDmol</u>	Int(10)	No		ID of the biomolecule cited in the bibliographic reference - refers to BioMolecule table
-----------------	---------	----	--	--

Peptide

Field	Type	Null	Default	Comment
<u>pepID</u>	int(10)	No		potentially antigenic peptide unique ID
molpep	int(10)	Yes	NULL	ID of the biomolecule from which the peptide is predicted - refers to BioMolecule table
peppdb	varchar(20)	Yes	NULL	name of the biomolecule structure
pepname	varchar(20)	Yes	NULL	peptide name
Pepsequence	text	Yes	NULL	peptide amino acid sequence
peplength	int(10)	Yes	NULL	peptide sequence length
pepmethod	text	Yes	NULL	method of extension used to build the peptide
Pepcomment	text	Yes	NULL	comment about the peptide

Publish

Field	Type	Null	Default	Comment
<u>pubIDref</u>	int(10)	No		ID of the bibliographic reference published by the author - refers to RefBiblio table
<u>pubIDaut</u>	int(10)	No		ID of the author involved in the publication of the bibliographic reference - refers to Author table
pos	int(10)	Yes	NULL	authorship position in the published reference

RefBiblio

Field	Type	Null	Default	Comment
<u>refbibID</u>	int(10)	No		bibliographic reference unique ID
medline	varchar(20)	Yes	NULL	Medline access number of the bibliographic reference
pubmed	varchar(20)	Yes	NULL	PubMed access number of the bibliographic reference
doi	varchar(40)	Yes	NULL	DOI access number of the bibliographic reference
title	text	Yes	NULL	title of the bibliographic reference
date	varchar(20)	Yes	NULL	publication date
journal	varchar(100)	Yes	NULL	journal of publication
volume	varchar(30)	Yes	NULL	journal volume containing the reference

SegInPep

Field	Type	Null	Default	Comment
<u>seginpepIDpep</u>	int(10)	No		ID of the peptide containing the segment - refers to Peptide table
<u>seginpepIDseg</u>	int(10)	No		ID of the segment composing the peptide - refers to Segment table
Molseginpep	int(10)	No		ID of the biomolecule owning both segment and peptide - refers to BioMolecule table

Segment

Field	Type	Null	Default	Comment
<u>segmentID</u>	int(10)	No		solvent accessible segment unique ID
segname	varchar(20)	Yes	NULL	segment name
molseg	int(10)	Yes	NULL	ID of the biomolecule from which were selected the segment - refers to BioMolecule table
cluseg	int(10)	Yes	NULL	ID of the cluster to which the segment belongs - refers to Cluster table
Segsequence	text	Yes	NULL	segment amino acid sequence
seglength	int(10)	Yes	NULL	segment sequence length
segfirstaa	int(10)	Yes	NULL	position of the first amino acid of the segment
seglastaa	int(10)	Yes	NULL	position of the last amino acid of the segment
segclassLEN	int(10)	Yes	NULL	segment ranking as for length criteria
segclassACC	int(10)	Yes	NULL	segment ranking as for accessibility criteria
segclassHYD	int(10)	Yes	NULL	segment ranking as for hydrophobicity criteria
segclassTURN	int(10)	Yes	NULL	segment ranking as for beta-turn criteria
segclassWRYP	int(10)	Yes	NULL	segment ranking as for amino acid W, R, Y, P criteria
segscoreLEN	int(10)	Yes	NULL	segment score as for length criteria
segscoreACC	int(10)	Yes	NULL	segment score as for accessibility criteria
segscoreHYD	int(10)	Yes	NULL	segment score as for hydrophobicity criteria

segscoreTURN	int(10)	Yes	NULL	segment score as for beta-turn criteria
segscoreWRYP	int(10)	Yes	NULL	segment score as for amino acid W, R, Y, P criteria

Species

Field	Type	Null	Default	Comment
<u>speID</u>	int(10)	No		biomolecule species unique ID
ncbitaxid	int(10)	Yes	NULL	NCBI access number of the species
spelatname	varchar(100)	Yes	NULL	species latin name
specomname	varchar(100)	Yes	NULL	species common name
speothername	varchar(100)	Yes	NULL	species other name
taxo	text	Yes	NULL	complete taxonomy of the species
class	varchar(50)	Yes	NULL	species class

StrucMotInFold

Field	Type	Null	Default	Comment
<u>smfolIDfold</u>	int(10)	No		ID of the fold which contains the structural motif - refers to Fold table
<u>smfolIDmot</u>	int(10)	No		ID of the structural motif composing the fold - refers to StrucMotif table

StrucMotif

Field	Type	Null	Default	Comment
<u>strmotID</u>	int(10)	No		structural motif unique ID

SympByMolecule

Field	Type	Null	Default	Comment
<u>symolIDmol</u>	int(10)	No		ID of the biomolecule responsible for the symptom - refers to BioMolecule table
<u>symolIDsymp</u>	int(10)	No		ID of the symptom caused by the biomolecule - refers to Symptom table

SympByNatEff

Field	Type	Null	Default	Comment
<u>syneIDsymp</u>	Int(10)	No		ID of the symptom present in the natural effect - refers to Symptom table
<u>syneIDeff</u>	Int(10)	No		ID of the natural effect containing the symptom - refers to NaturalEffect table

Symptom

Field	Type	Null	Default	Comment
<u>sympID</u>	int(10)	No		symptom unique ID
sympname	varchar(30)	Yes	NULL	symptom name

TempGroup

Field	Type	Null	Default	Comment
<u>tempgpID</u>	int(10)	No		template group unique ID
tempgpname	varchar(50)	No		template group name
level	smallint(10)	No		template group level of classification

mainfunc	text	Yes	NULL	main function of the template group members
----------	------	-----	------	---

ToxinTempGroup

Field	Type	Null	Default	Comment
<u>toxingpIDtox</u>	int(10)	No		ID of the toxin classified into the template group - refers to BioMolecule table
<u>toxingpIDgp</u>	int(10)	No		ID of the template group in which the toxin is classified - refers to TempGroup table
structype	enum('model', 'template', 'target', 'valid')	No		type of toxin structure whether it is a template for modeling or a target to be modeled

APÊNDICE 2: *Scripts* adaptados do Modeller

salgn.py

```
from modeller import *
log.verbose()
env = environ()

env.io.atom_files_directory = ['/home/cfleury/public_html/silimod/tmp/PDB/']

aln = alignment(env)
for (code, chain) in (('1QDP', 'A'), ('1VTX', 'A')):

    mdl = model(env, file=code, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(mdl, atom_files=code, align_codes=code+chain)

for (weights, write_fit, whole) in (((1., 0., 0., 0., 1., 0.), False, True),
                                    ((1., 0.5, 1., 1., 1., 0.), False, True),
                                    ((1., 1., 1., 1., 1., 0.), True, False)):
    aln.salign(rms_cutoff=3.5, normalize_pp_scores=False,
               rr_file='$(LIB)/asl.sim.mat', overhang=30,
               gap_penalties_ld=(-450, -50),
               gap_penalties_3d=(0, 3), gap_gap_score=0, gap_residue_score=0,

               dendrogram_file='/home/cfleury/public_html/silimod/tmp/TXR1_MACRV_salign.tree',

               alignment_type='tree',
               feature_weights=weights,
               improve_alignment=True, fit=True, write_fit=write_fit,
               write_whole_pdb=whole, output='ALIGNMENT QUALITY')

aln.write(file='/home/cfleury/public_html/silimod/tmp/TXR1_MACRV_salign.pap',
          alignment_format='PAP')
aln.write(file='/home/cfleury/public_html/silimod/tmp/TXR1_MACRV_salign.ali',
          alignment_format='PIR')

aln.salign(rms_cutoff=1.0, normalize_pp_scores=False,
           rr_file='$(LIB)/asl.sim.mat', overhang=30,
           gap_penalties_ld=(-450, -50), gap_penalties_3d=(0, 3),
           gap_gap_score=0, gap_residue_score=0, dendrogram_file='lis3A.tree',
           alignment_type='progressive', feature_weights=[0]*6,
           improve_alignment=False, fit=False, write_fit=True,
           write_whole_pdb=False, output='QUALITY')
```

multalign2d.py

```
from modeller import *
log.verbose()
env = environ()

env.libs.topology.read(file='$(LIB)/top_heav.lib')

aln = alignment(env)
aln.append(file='/home/cfleury/public_html/silimod/tmp/TXR1_MACRV_salign.ali',
          align_codes='all')
aln_block = len(aln)

aln.append(file='/home/cfleury/public_html/silimod/tmp/TXR1_MACRV.ali',
          align_codes='TXR1_MACRV')

aln.salign(output='', max_gap_length=20,
           gap_function=True,
           alignment_type='PAIRWISE', align_block=aln_block,
           feature_weights=(1., 0., 0., 0., 0., 0.), overhang=0,
           gap_penalties_ld=(-450, 0),
           gap_penalties_2d=(0.35, 1.2, 0.9, 1.2, 0.6, 8.6, 1.2, 0., 0.),
           similarity_flag=True)
```



```
aln.write(file='/home/cfleury/public_html/silimod/tmp/TXR1_MACRV_multalign2d.ali',
alignment_format='PIR')
aln.write(file='/home/cfleury/public_html/silimod/tmp/TXR1_MACRV_multalign2d.pap',
alignment_format='PAP')
```

singalign2d.py

```
from modeller import *

log.verbose()
env = environ()
aln = alignment(env)
mdl = model(env, file='/home/cfleury/public_html/silimod/tmp/PDB/1QDP',
model_segment=('FIRST:A', 'LAST:A'))
aln.append_model(mdl, align_codes='1QDPA',
atom_files='/home/cfleury/public_html/silimod/tmp/PDB/1QDP.pdb')
aln.append(file='/home/cfleury/public_html/silimod/tmp/TXDT1_HADVE.ali',
align_codes='TXDT1_HADVE')
aln.align2d()
aln.write(file='/home/cfleury/public_html/silimod/tmp/TXDT1_HADVE_singalign2d.ali',
alignment_format='PIR')
aln.write(file='/home/cfleury/public_html/silimod/tmp/TXDT1_HADVE_singalign2d.pap',
alignment_format='PAP')
```

model.py

```
from modeller import *
from modeller.automodel import *

class MyModel(automodel):
    def special_patches(self, aln):
        self.patch(residue_type='DISU', residues=(self.residues['1'],
self.residues['16']))
        self.patch(residue_type='DISU', residues=(self.residues['8'],
self.residues['21']))
        self.patch(residue_type='DISU', residues=(self.residues['15'],
self.residues['36']))
        self.patch(residue_type='DISU', residues=(self.residues['17'],
self.residues['43']))

log.verbose()
env = environ()

env.io.atom_files_directory = ['/home/cfleury/public_html/silimod/tmp/PDB/']

a = MyModel(env,
alnfile =
'/home/cfleury/public_html/silimod/tmp/TXR1_MACRV_multalign2d.ali',
knowns = ('1QDPA', '1VTXA'),
sequence = 'TXR1_MACRV')

a.starting_model= 1
a.ending_model= 50

a.make()
```


Kozclass1102

JZT11_CHIJI 2A2V spi na k	-----	ECRKMFGGCS	-----	VSDS	CA	--	HLGCKPTL	-----	KY	-----	AWDGT	F	-----	34																			
TXG2_PLEGU spi k	-----	ECRKMFGGCS	-----	VSDS	CA	--	HLGCKPTL	-----	KY	-----	AWDGT		-----	33																			
TX1_STRCF spi k	-----	DCTRMFGACR	-----	RSDS	CP	--	HLGCKPTS	-----	KY	-----	AWDGT	I	-----	34																			
TX1_SCOGR 1LA4 spi k	-----	TCRYLFGGCK	-----	TTAD	CK	--	HLACRSDG	-----	KY	-----	AWDGT	F	-----	34																			
TX1_HETMC spi k	-----	ECRYLFGGCS	-----	STSD	CK	--	HLSCRSDW	-----	KY	-----	AWDGT	F	S	-----	35																		
TXHN1_GRARO 1D1H spi k ca	-----	ECRYLFGGCK	-----	TTSD	CK	--	HLGCKFRD	-----	KY	-----	AWD	F	T	F	S	-----	35																
TXHN2_GRARO spi k	-----	ECRYLFGGCK	-----	TTAD	CK	--	HLGCKFRD	-----	KY	-----	AWD	F	T	F	S	-----	35																
TX1_PSACA spi k	-----	SECRWFMGGCD	-----	STLD	CK	--	HLCKMGL	-----	YY	-----	AWDGT	F	-----	35																			
TX2_PSACA spi k	-----	GACRWFLLGGCK	-----	STSD	CE	--	HLCKMGL	-----	DY	-----	AWDGT	F	-----	35																			
TX3_CERCR spi na		GVDKEGCRKLLGGCT	-----	IDDD	CP	--	HLGCKKY	-----	WH	-----	GWDGT	F	-----	39																			
TX482_HYSGI spi ca		GVDKAGCRYMFGGCS	-----	VNDD	CP	--	RLGCKSLF	-----	SY	-----	AWD	L	T	F	S	D	-----	41															
TX3_PSACA spi	-----	ECRWYLGCK	-----	EDSE	CE	--	HLQCHSYW	-----	EW	-----	LWDGS	F	-----	34																			
TXH5_ORNHU spi	-----	ECRWYLGCK	-----	QDGD	CK	--	HLQCHSNY	-----	EW	-----	VWDGT		-----	33																			
TXPR1_THRPR spi na k ca	-----	ECRYWLGCK	-----	AGQT	CK	--	HLVCSRRH	-----	GW	-----	VWDGT	F	S	-----	35																		
WGRTX_GRARO 1KOZ spi k ca	-----	DCVRFWVKCS	-----	QTS	CP	--	HLACKSKW	-----	PR	--	NI	VWDGS	V	-----	36																		
TX2_HETMC spi k	-----	ECRYFWGECN	-----	DEMV	CE	--	HLVCKEKW	-----	PITYKI		VWDRT	F	-----	38																			
TXG1D_PLEGU spi k	---	DGECGGFWWKC	GS	--	GKPA	C	--	PKYVCS	PKW	--	GL	---	QNF	P	M	P	-----	36															
TXG1E_PLEGU spi k	---	EGECGGFWWKC	GS	--	GKPA	C	--	PKYVCS	PKW	--	GL	---	QNF	P	M	P	-----	36															
JZTX3_CHIJI 2I1T spi na k	---	DGECGGFWWKC	GR	--	GKPP	C	--	KGYACS	KTW	--	GW	---	AVE	E	A	P	-----	36															
JZTX1_CHIJI spi na	-----	ACGQFWWKC	GE	--	GKPP	C	--	ANFACK	IGL	---	YL	---	C	--	I	W	S	P	-----	33													
MTX4_GRARO 1TYK spi	-----	GCLF	FWWKC	NP	--	NDDK	CR	--	PKLK	CS	KL	F	-----	KL	---	N	F	S	F	-----	34												
TXHA1_SELHA 1NIX spi na	-----	ECKG	FGK	SC	VP	--	GKNE	CS	--	GYAC	NS	RD	-----	KW	---	K	V	L	L	-----	33												
TXHA3_SELHA 2JTB spi na	-----	GCKG	FGD	S	CTP	--	GKNE	CP	--	NYAC	SS	KH	-----	KW	---	K	V	Y	L	-----	33												
TXH1_ORNHU 1QK6 spi ca	-----	ACKG	VFD	A	CTP	--	GKNE	CP	--	NRVC	S	DKH	-----	KW	---	K	W	K	L	-----	33												
TXHA4_SELHA 1NIY spi na	-----	ECLG	FGK	G	CNP	--	SNDQ	CK	SS	N	L	V	C	R	KH	-----	RW	---	K	Y	E	I	-----	35									
TXHA5_SELHA spi na	-----	ECLG	FGK	G	CNP	--	SNDQ	CK	S	A	N	L	V	C	R	KH	-----	RW	---	K	Y	E	I	-----	35								
TXH4_ORNHU 1MB6 spi na	-----	ECL	E	I	F	K	A	C	N	P	--	SNDQ	CK	S	S	K	L	V	C	R	K	T	-----	RW	---	K	Y	Q	I	-----	35		
TX3_PARSR spi na	-----	DCLG	F	L	W	K	C	N	P	--	SNDK	CR	--	PNL	V	C	S	R	K	D	-----	KW	---	K	Y	Q	I	-----	34				
TX1_CERCR spi na	-----	DCLG	W	F	K	S	C	D	P	--	KNDK	CK	--	NYT	C	S	R	R	D	-----	RW	---	K	Y	D	L	-----	33					
TX2_CERCR spi na	-----	DCLG	W	F	K	S	C	D	P	--	KNDK	CK	--	NYT	C	S	R	R	D	-----	RW	---	K	Y	Y	L	-----	33					
VSTX3_GRARO spi k	-----	DCLG	W	F	K	G	C	D	P	--	DNDK	CE	--	GYK	C	N	R	R	D	-----	KW	---	K	Y	K	L	W	-----	34				
TX2_THEBL spi k	----	DDCL	G	M	F	S	S	C	D	P	--	KNDK	CP	--	NRV	C	R	S	R	D	-----	QW	---	K	Y	K	L	W	-----	35			
TX3_THEBL spi k	----	DDCL	G	M	F	S	S	C	D	P	--	NNDK	CP	--	NRV	C	R	V	R	D	-----	QW	---	K	Y	K	L	W	-----	35			
TX1_THEBL spi k	----	AA	C	L	G	M	F	S	C	D	P	--	NNDK	CP	--	N	R	E	C	N	R	K	H	-----	KW	---	K	Y	K	L	W	-----	35
JZTX7_CHIJI 1ZJQ spi na	-----	GCG	L	M	A	G	C	D	G	-----	KSTF	CS	--	GYNC	S	P	T	W	-----	KW	---	V	Y	A	R	P	-----	34					
TXH3_ORNHU spi	-----	DC	A	G	Y	M	R	E	C	-----	KEK	L	C	S	--	GYV	C	S	S	R	W	-----	KW	---	V	L	P	A	P	-----	32		
VSTX1_GRARO 1S6X spi k	-----	ECG	K	F	M	W	K	C	K	-----	NSND	CK	--	DLV	C	S	S	R	W	-----	KW	---	V	L	A	S	P	F	-----	34			
TXP5_BRASM spi	-----	SC	V	D	F	Q	T	K	C	K	-----	KDSD	CG	--	KLE	C	S	S	R	W	-----	KW	---	V	Y	P	S	P	F	-----	34		

MTX2_GRARO 1LUP spi	-----YCQKWMWTCDE---E-RKCE--GLVCR-----LWCKRIINM-----	31
TXP2_PARSR spi k	-----YCQKWMWTCDE---E-RKCE--GLVCR-----LWCKRIINM-----	31
TXP1_PARSR 1V7F spi k	-----YCQKWMWTCDS---A-RKCE--GLVCR-----LWCKKII-----	29
AF2_GRARO spi	-----YCQKWMWTCDE---E-RKCE--GLVCR-----LWCKKKIEW-----	31
VSTX2_GRARO spi k	-----YCQKWMWTCDE---E-RKCE--GLVCR-----LWCKKKIEEG-----	32
AF1_GRARO spi	-----YCQKWLWTCDS---E-RKCE--DMVCR-----LWCKKRL-----	29
TX1_GRARO spi na	-----YCQKWMWTCDS---K-RKCE--DMVCR-----LWCKKRL-----	29
TXPR2_THRPR spi na ca	-----YCQKWMWTCDS---E-RKCE--GMVCR-----LWCKKKLW-----	30
JZT12_CHIJI spi k	-----YCQKWMWTCDS---E-RKCE--GYVCE-----LWCKYNL-----	29
JZTX5_CHIJI spi na k	-----YCQKWMWTCDS---K-RACE--GLRCK-----LWCKKII-----	29
TXHP2_HETVE 1EMX spi k ca	----DDCGKLFSGCDT---N-ADCE--GYVCR-----LWCKLDW-----	30
TXU2_HETVE spi ca	----DDCGGLFSGCDS---N-ADCE--GYVCR-----LWCKYKL-----	30
TXJ1_HETVE spi ca	----DDCGTLFSGCDT---S-KDCE--GYVCH-----LWCKYK-----	29
TXHP3_HETVE spi k ca	----ECGTLFSGCST---H-ADCE--GFICK-----LWCKRYERTW-----	31
TX5B_HETVE spi ca	----ADCGWLFHSCES---N-ADCE--NWACATTG---RF-RYLCKYQI-----	36
TXHP1_HETVE spi k ca	----DCGTIWHYCGT---DQSECE--GWKCS-----RQLCKYVIDW-----	33
TX5A_HETVE spi ca	---DDDGGWIMDDCTS---D-SDCP--NWCCKTG---FV-KNICKYEM-----	37
TXFK1_PSACA 1X5V spi	----ACGILHDNCVYVPAQNP--CRG-L-QCRYGK-----GLVQV-----	33
TXP7_APTSC 1WQB spi	---WLGCARVKEACG--PWEWP--CSG-L-KDGESE-----GHPQ-----	32
TOT1A_HADVE 1AXH spi ca	---SPTCIPSGQPC---PYNEN--CSQ---SCTFKE-NENGNQVKRQD-----	37
TOT1C_HADVE spi ca	---SSTCIPSGQPC---PYNEN--CSQ---SCTFKE-NENGNQVKRQD-----	37
TOT1B_HADVE spi ca	---SSTCIPSGQPC---PYNEN--CSQ---SCTYKE-NENGNQVKRQD-----	37
TOT1E_HADVE spi ca	---SPTCIPSGQPC---PYNEN--CSQ---SCTYKE-NENGNQVKRQD-----	37
TOT1B_HADFO spi ca	---SPTCIRSGQPC---PYNEN--CSQ---SCTFKT-NENGNQVKRQD-----	37
TOT1D_HADVE spi ca	---SPTCIPSGQPC---PYNEN--CSK---SCTYKE-NENGNQVKRQD-----	37
TOT1B_HADIN spi ca	---SPTCIPTGQPC---PYNEN--CSQ---SCTYKA-NENGNQVKRQD-----	37
TOT1A_HADIN spi ca	---STCTPTDQPC---PYHES--CSG---SCTYKA-NENGNQVKRQD-----	36
TOT1C_HADIN spi ca	---SSTCIRTQPC---PYNES--CSG---SCTYKA-NENGNQVKRQD-----	37
TOT1F_HADVE spi ca	---SAVCIPSGQPC---PYSKY--CSG---SCTYKT-NENGNQVKRQD-----	37
TOT1A_ATRRO spi ca	---SSVCIPSGQPC---PYNEH--CSG---SCTYKE-NENGNQVKRQD-----	37
TXH10_ORNHU 1Y29 spi ca	----KCLPPGKPCYGATQKIP--CG---VCSHNK-----CT-----	28
TXFK2_PSACA spi	----RCLPAGKTCVRGPMRVP--CG---SCSQNK-----CT-----	28
TXH9_ORNHU spi	----IICAPEGGPC---VAGIG--CAG-L-RCSGAKLGLAG---SQ-----	34
TXMG7_MACGS spi	----IICAPEGGPC---VVGIG--CKG-Y-SCAPGLLGLVG---HQ-----	34
TOT2A_HADIN spi ca	--GVLDCVVNTLGCS--SDKD--CGM-TPSCTL-----G---IAPS---VGGIVGGLL	42
TOT2B_HADIN spi ca	--GVVDCVLNLTGCS--SDKD--CGM-TPSCTL-----G---IAPS---VGGIVGGLL	42
TOT2A_HADVE 1G9P spi ca	---LLACLFNGRCS--SNRD--CEL-TPVCKR-----G---SVSSGPGLVGGILGGIL	45
TXMG5_MACGS 2GX1 spi na	----GCKLTFWKCK--NKKE--CGW-N-ACAL-----G---IIMPR-----	29
TXR3_MACRV spi	----GCKLTFWKCK--NKKE--CGW-N-ACAL-----G---IIMPR-----	29
TXM11_MACGS spi na	----SCKLTFWRCK--KDKE--CGW-N-ICT-----G---LQIPP-----	28

TXMG8_MACGS spi	-----DCKGLFRQCK--KSSE--CKG-S-SCESDL---TG---LQIFNLP-----	34
TXAG_AGEOP 2E2S spi ca	-----GGCLPHNRFCN--ALSGPRCSG-L-KCKELS---IW--DSRRL-----	35
TXAG_AGEOR spi ca	-----GGCLPHNRFCN--ALSGPRCSG-L-KCKELS---IW--DSRRL-----	35
SCX1_OPICA sco	-----GDCLPHLKRCK--ENND--CS--K-KCKRRG---TNP-EKRGR-----	33
SCX2_OPICA sco	-----GDCLPHLKRCK--ENND--CS--K-KCKRRG---ANP-EKRGR-----	33
SCXC_SCOMA 1C6W sco	-----GDCLPHLKLCK--ENKD--CS--K-KCKRRG---TNI-EKRGR-----	33
IPTXA_PANIM 1IE6 sco ca	-----GDCLPHLKRCK--ADND--CG--K-KCKRRG---TNA-EKRGR-----	33
TX27_PHORI spi	-----IACAPRGLLCF--RDKE--CKG-L-TCKGRF---VNTWPTFQLV-----	36
TX32_PHOKE spi	-----IACAPRGQLCF--SDKL--CIG-L-RCKSRV---ANMWPTFQLV-----	36
TX27_PHONI spi	-----IACAPRFLCN--SDKE--CKG-L-RCKSRI---ANMWPTFQSQ-----	36
TX13_PHONI spi	-----VFCRSNGQQCT--SDGQ--CY--G-KCMTAF---MG---KIQMR-----	32
TX24_PHONI spi	-----VFCRFNGQQCT--SDGQ--CY--G-KCRTAF---MG---KIQM-----	31
T244_PHONI spi	-----VFCRFNGQQCT--SDGQ--CY--G-KCRTAF---LR---MIQMG-----	33
TX17_PHORI spi	-----AFCRFNGQQCT--SDGQ--CN--G-RCINAF---QG---RIQIG-----	32
TX21_PHOKE spi	-----AFCKYNGEQCT--SDGQ--CN--G-RCRTAF---MG---KIQMG-----	32
F256_OLIOR spi	-----APCTYPGQQCK--SDDE--CH--G-TCKTAF---IG---RIQMR-----	32
TXMG1_MACGS spi	-----CMGYDIHCTD--RLP--CFG--LECVKTSGY-WWYKKTYSR-RKS-----	38
TXMG2_MACGS spi	-----CMGYDIECNE--NLP--CKHRKLECVETSGY-WWYKRYRPIKG-----	41
TXLT4_LASPA spi	-----CGVDAPCDK--DRPD--CSY--AECLRPSGYGWWHGTYYSYRKRER-----	42
TXVL2_CORVA spi	-----ACSRAGENCY--KSGRCD--GLYCKAYV-----VTCYKP-----	31
SCXC1_MESMA sco	-----GCNRLNKKCN--SDGD--CRY-GERCISTG---VNY---YCRPDFGP-----	37
TXP1_PSACA 1LMM spi na	-----EDCIPKWKGCVN--RHGD--CE--GLECWKRR--RSF--EVVVPKTPKT-----	40
TX22_PHONI spi	-----MPCPKILKQCK--SDED--CR--GWKCFGFS---IKDKMISR-----	35
TXM10_MACGS spi	-----ACLAEYQKCEG--STVPCP--GLSCSAGR---FRKTKLTK-----	35

Kozclass1202

TOT2A_ATRIL spi ca	RLLDVLSRVCSPDANCCGLTPICKMG----LQVPKVGGLLGGLLGGIL	45
TOT2B_ATRIL spi ca	RLLDVLSRVCSDDANCCGLTPTCKMG----LQVPKVGGLLGGLLGGIL	45
TX29_PHONI spi na	SF--CIPFKPKSDENCCGKFK--CKTTGIVKLRW-----	32
TXL1_ORNHU 1QK7 spi	---GCLGDK--CDYNGCCSGYV--CSR--TWKVVLAGPW-----	32

Kozclass1403

TXDT1_ATTRO 1QDP spi na	CAKKNWCGKNED--CCPMKCIYAWYNQQ---GSQTTITGLFK-K--	42
-------------------------------	--	----

```

TXDT1_HADSP | spi | na      CAKKRNWCGKNEDECCPMKCIYAWYNQQ----GSQSTITGLFK-KC-- 42
TXDT1_HADVE | 1VTX | spi | na CAKKRNWCGKTED-CCCPMKCVYAWYNEQ----GSQSTISALWK-KC-- 42
TDM1A_MISBR | spi | na      CAKKREWCAKTED-CCCPMKCIYAWYNEQ----SSQTTFSGMFK-KC-- 42
TDT1B_HADVE | spi | na      CSRSDGWCGKTED-CCCPMKCIKAWYKQN----GNCQNTISAIWK-NC-- 42
TXM14_MACGS | spi | na      CARKRAWCEKTEN-CCCPMKCIYAWYNGQ----SSDHTISTIWT-SLPK 44
TXMG4_MACGS | spi | na      CGSKRAWCKEKKD-CCCGYNCVYAWYNQQ----SSLERKWKYLFTEG-- 43
TXR1_MACRV | spi          CGTNRAWCRNAKDHCCCGYSCVKPIWASKPEDDGYWKKFG-----G-- 43
* . . * : : * : . : : : : : : : :

```

Kozclass1502

```

TJT1A_HADVE | spi | k      -TICTGADRPCAACCPCCPGTSCQGPESNGVYTRNF-- 36
TJT1B_HADVE | spi | k      -TICTGADRPCAACCPCCPGTSCQGPENGVSYTRND-- 36
TJT1A_HADFO | spi | k      SPTCTGADRPCAACCPCCPGTSCQGPENGVSYTRND-- 37
TJT1C_HADVE | 1DL0 | spi | k -AICTGADRPCAACCPCCPGTSCK-AESNGVSYTRKDEP 37
. ***** : . * . * . * :

```

Kozclass2110

```

TXC2_HOLCU | spi          -----ADCVGDGQRCADWAGPY---CCS-----GYYCSRSM-----PYRFRSDS----- 38
TXMG3_AGEAP | spi          -----ADCVGDGQRCADWAGPY---CCS-----GYYCSRSM-----PYRFRSDS----- 38
TXC3_HOLCU | spi          -----ADCVGDGQRCADWFGPY---CCS-----GYYCSRSM-----PYRFRSDS----- 38
TXDP2_PARLU | 1V91 | spi | na -----ACVGDGQRCASWSGPY---CD-----GYYCSRSM-----PYRFRNNS----- 37
TXDP1_PARLU | 1V90 | spi | na -----GCLGEGEKCADWSGPS---CD-----GFYCSRSM-----PYRFRNNS----- 37
TXDP4_PARLU | spi | na      -----ACATKNQRCASWAGPY---CD-----GFYCSRSY-----PGMRPNS----- 37
TXMG2_AGEAP | spi          -----ECATKNKRCADWAGPW---CD-----GLYCSRSY-----PGMRPSS----- 37
TXMG1_AGEAP | 1EIT | spi       -----ECVPENGHCRDWY-DE---CE-----GFYCSRQP-----PKIQRNNS----- 36
TXC1_HOLCU | spi          -----SCVGEYGRCSAY-ED---CD-----GYYCNSQP-----PYLQRNNS----- 36
TXMG4_AGEAP | spi          -----ACVGENQQCADWAGPH---CD-----GYYCTRYF-----PKIQRNNS----- 37
TXMG5_AGEAP | spi          -----ACVGENKQCADWAGPH---CD-----GYYCTRYF-----PKIQRNNS----- 37
TXMG6_AGEAP | spi          -----DCVGESQQCADWAGPH---CD-----GYYCTRYF-----PKIQRNNS----- 37
TXMO_HADIN | 1KQH | spi | na -----KCLAEAADCSPPWSGDS---CK-----PYLCSIFP-----YPSRPPKRW----- 38
TXDP3_PARLU | spi          -----ADCLNEGDCADWSGPS---CG-----EMWCSPGF-----GKRCKK----- 36
SFI3_SEGFL | spi          -----KECMVDGTVCYIHNHND---CG-----SCLLNGPIARPWEM-MVGNKGGPKA----- 46
SFI4_SEGFL | spi          -----KECMVDGTVCYIHNHND---CG-----SCLLNGPIARPWKM-MVGNKGGPKA----- 46
SFI6_SEGFL | spi          -----KECMTDETVCYIHNHND---CG-----SCLLNGPIARPWEM-MVGNKGGPKA----- 46
SFI1_SEGFL | spi          -----KECMTDGTVCYIHNHND---CG-----SCLLNGPIARPWEM-MVGNKGGPKA----- 46

```

```

SFI5_SEGFL|spi -----KECMVDGTVCYIHNHND---CG-----SCLAPNGPLARPWEM-LVGNKKGPKA----- 46
SFI8_SEGFL|spi -----KECMADGTVCYIHNHND---CG-----SCLAPNGPLARPWEM-LVGNKKGPKA----- 46
SFI7_SEGFL|spi -----KECMADGTVCYIHNHND---CG-----SCLAPNGPLARPWEV-LVGNKKGPKA----- 46
SFI2_SEGFL|spi -----KECMADGTVCYIHNHND---CG-----SCLAPNGPYARPWEM-LVGNKKGPKA----- 46
TXC1_CUPSA|spi|ca -----SCIPKHEEC-TNDKHN---CRK---GLFKLKQSTFDDESGQP---TERAAGRPMGHQALETGLNIF--- 60
TXC9_CUPSA|spi|ca -----KDDKNCIPKHHEC-TNDKKN---CKK---GLTKMKCKFTVADAKGAT---SERADSSLLQKFGFTGLHIKGLF--- 68
TX35_PHONI|spi|ca -----GCIGRNEEC-KFDRHG---CC-----WPWSCSWNKEGQPE---DVEEESLKIGK----- 45
TX1A_GEOA2|spi|ca -----EKSCITWRNSC-MHNDKG---CC-----FPWSCVWSQTVSRNSSR-KEKKQQLRW----- 47
TX19_PHOKE|spi -----GKCADAWKSCDN---LP---CV-V---NGYSRTCMS-----ANRNEEETKKLREHFGR----- 46
TX33A_PHONI|spi|ca -----G-CADAYKSCNHP--RT---CDGY---NGYKRACIS-----GSNKCK----- 36
TX3A_PHONI|spi|ca -----ACADVYKECWYPE-KP---CK-----DRACQTL-----GMTKCKA----- 34
TX32_PHONI|spi|ca -----ACAGLYKKCKGGA-SP---CE-----DRPCKDLA-----MGNLCK----- 34
TX37_PHORI|spi|ca -----ACAGLYKKCKGV-NT---CE-----NRPCKDLA-----MGNLCK--KFVEFFGG----- 43
TX31_PHONI|spi|ca -----AECAAVYERCGKY-KR---CE-----ERPCKNIV-----MDNITCK--FISE----- 40
TX325_SEGFL|spi|ca -----GSCIESGKSCTHSRSMKNGLCPP-----KSRCKRQIQHRHDYLGKRYSRCS----- 49
TXI92_DIGCA|spi|na AKDGDVEGPAGCKKYDVECDSGE-----CCQQLWYKWRPLDCRLKSG-----FFSSKVRDV----- 56

```

Kozclass2310

```

TOG4A_AGEAP|1IVA|spi|ca -----KKKCIADYGRCKWGGTPCC-----RGRGCIISIMGTN---EKKPRLIMEGLGLA---- 48
TOG4B_AGEAP|1AGG|spi|ca -----EDNCIAEDYGKCTWGGTKCC-----RGRPCRSMIGTN---EETPRLIMEGLSFA---- 48
TXI11_DIGCA|spi|na AKDGDVKGPAGCMKYKSGDCR--GKTCDQQYLWYKWRNLACRFVVEVFKKDWNDIS----- 58

```

NOT KNOTTINS :

6C_N3.x.x.5.4

```

TX2_EURCA|spi IFECVFSDIEKEGK-PCKPKGEKKSGGWCKIKLTKI 39
TXP1_BRASM|spi IFECVFSDIEKEGK-PCKPKGEKKSGGWCKIKLTKI 39
TXP1_BARRH|spi ILECFVSDIKKEGK-PCKPKGEKKTGGWRCKIKLTKI 39
TXP2_BARRH|spi IFECVFSDIKKEGK-PCKPKGEKKTGGWRCKIKLTKI 39
TX1_EURCA|spi IFECVFSDIEKEGK-PCKPKGEKKTGGWRCKIKLTKI 39
TXH21_ORNHU|1I25|spi LFECSFSCIEKEGDKPKKK---KCKGGWKCKFNMVKV 37
TXH22_ORNHU|spi LFECSFSCIEKEGDKPKKK---KCKGGWKCKFNMVKV 37
TXH7_ORNHU|spi -FECSFSCIEKEGDKPKKK---KCKGGWKCKFNMVKV 36

```


6C_Nx.3.x.4.1

KA101_CENNO	1PJV	sco	k	-----AVC-VYRTCDKDKRR-GYRSG-KCINNACKYYPY	32
KA102_CENNO	sco	k		-----VAC-VYRTCDKDKTSR-KYRSG-KCINNACKYYPY	32
KAX51_LEIQH	1SCY	sco	k	-----AFC-NLRMCQLSRSLS-GLL-G-KCIGDKCEVVKH	31
KAX53_MESMA	sco	k		-----AVC-NLKRQCQLSRSLS-GLL-G-KCIGDKCEVVKH	31
KAX52_ANDMA	1PNH	sco	k	-----TVC-NLRRQCQLSRSLS-GLL-G-KCIGVKCEVVKH	31
KAX54_MESTA	sco	k		-----AFC-NLRRCELSRSLS-GLL-G-KCIGEECKVVPY	31
KAX55_MESTA	sco	k		-----AFC-NLRRCELSRSLS-GLL-G-KCIGEECKVVPH	31
KA191_MESMA	1Q2K	sco	k	-----AAC-YSSDCRVKQVAM-GFSSG-KCINSKCKCYK-	31
KAX28_CENEL	sco	k		---TVINVKCTSPKQCLPKPKDLYGPHAGAKCMNGKCKCYNN	39
KAX29_CENEL	sco	k		---TIINVKCTSPKQCLPKPKDLYGPHAGAKCMNGKCKCYNN	39
KAX21_CENNO	1SXM	sco	k	---TIINVKCTSPKQCSKPKKELYGSSAGAKCMNGKCKCYNN	39
KAX2B_CENEL	sco	k		---TIINVKCTSPKQCLLPCKEITYGIHAGAKCMNGKCKCYKI	39
KAX2C_CENEL	sco	k		---TIINVKCTSPKQCLPPCKEITYGRHAGAKCMNGKCHCSKI	39
KAX22_CENMA	1MTX	sco	k	---TIINVKCTSPKQCLPPCKAQFGQSAGAKCMNGKCKCYPH	39
KAX25_CENLM	1HLY	sco	k	---TVIDVKCTSPKQCLPPCKAQFIRAGAKCMNGKCKCYPH	39
KAX23_CENLL	sco	k		---ITINVKCTSPQQCLRPCKDRFGQHAGGKCINGKCKCYP-	38
KAX2A_CENEL	sco	k		---IFINVKCSLPQQCLRPCKDRFGQHAGGKCINGKCKCYP-	38
KAX41_TITSE	1HP2	sco	k	---VFINAKCRGSPKCLPKCKEAIKKAAG-KCMNGKCKCYP-	37
KAX45_TITCO	sco	k		---VFINVKCRGSPKCLPKCKEAIKKSAG-KCMNGKCKCYP-	37
KAX43_TITDI	sco	k		---VFINVKCTGSKQCLPAKKAAVGKAAG-KCMNGKCKCYP-	37
KAX44_TITCA	sco	k		---VFINVKCRGSKKCLPAKKAAVGKAAG-KCMNGKCKCYP-	37
KAX24_CENNO	sco	k		---TIINEKCFATSQCWTPCKKAIGSLQS-KCMNGKCKCYNG	38
KAX32_LEIQH	1AGT	sco	k	--GVPINVSTGSPQCIKPKKDA-GMRFG-KCMNRKCHCTPK	38
KAX33_LEIQH	sco	k		--GVPINVPTGSPQCIKPKKDA-GMRFG-KCMNRKCHCTPK	38
KAX31_ANDMA	2KTX	sco	k	--GVEINVKCSGSPQCLPKKDA-GMRFG-KCMNRKCHCTPK	38
KAX34_LEIQH	sco	k		--GVPINVKCTGSPQCLPKKDA-GMRFG-KCINGKCHCTPK	38
KAX38_BUTSI	sco	k		--GVPINVKCRGSPQCIQPERDA-GMRFG-KCMNGKCHCTPQ	38
KAX37_ORTSC	1SCO	sco	k	--GVIINVKCKISRQCLEPKKKA-GMRFG-KCMNGKCHCTPK	38
KAX36_MESMA	1BKT	sco	k	---VGINVCKKHSQGCLPKPKDA-GMRFG-KCINGKCDCTPK	37
KAX39_BUTOC	sco	k		-GNVGIPIVSKKHSQGCIKPKKDA-GMRFG-KCMNRKCDCT--	37
KA162_MESMA	1M2S	sco	k	---GLIDVCKCFASSECWIAKVKVTGSGQG-KCQNNQCRCY--	36
KA163_MESMA	sco	k		---FGLIDVKCFASSECWIAKVKVTGSGVQG-KCQNNQCRCY--	37
KAX17_LEIQH	sco	k		---GLIDVRCYDSRQCWIAKVKVTGSTQG-KCQNKQCRCY--	36
KA161_MESTA	sco	k		---DLIDVKCISSQECWIAKVKVTGRFEG-KCQNRQCRCY--	36
KAX11_LEIQH	2CRD	sco	k	---QFTNVSCTTSKECWSVQRLHNTSRG-KCMNKKCRCY-	37
KAX1C_LEIQH	sco	k		---QFTDVSTTSKECWSVQRLHNTSIG-KCMNKKCRCY-	37
KAX1D_LEIQH	sco	k		---QFTNVSCTTSKECWSVEKLYNTSRG-KCMNKKCRCY-	37
KAX12_LEIQH	1LIR	sco	k	---QFTQESCTASNQWSIEKRLHNTNRG-KCMNKKCRCY-	37

KAX16_MESMA	2BMT	sco	k	---QFTNVSCSASSQCWPVCKKLFQTYRG-KCMNSKCRQYS- 37
KAX14_CENLM		sco	k	---VFIDVSVSVSKCEWAPCKAAVGTDRG-KCMGKKCRQYX- 37
KAX1B_CENNO		sco	k	---TFIDVDCTVSKCEWAPCKAAFVDRG-KCMGKKCKQYV- 37
KAX13_MESTA		sco	k	---QFTDVDCSVSKCEWSVCKDLFGVDRG-KCMGKKCRQYQ- 37
KAX15_MESMA	1BIG	sco	k	---QFTDVKCTGSKQCWPVCKQMFQKPNG-KCMNGKCRQYS- 37
KGX21_BUTEU	1J5J	sco	k	---RPTDIKCSSEYQCFPVCKSRFGKTNG-RCVNGFCDQF-- 36
KGX22_MESMA		sco	k	---RPTDIKCSASYQCFPVCKSRFGKTNG-RCVNGLCDQF-- 36
KAX71_PANIM	2PTA	sco	k	----DKTISCTNPKQCYPHCKKETGYPNA-KCMNRKCKQF-- 35
KAX72_PANIM	1C49	sco	k	-----TISCTNEKQCYPHCKKETGYPNA-KCMNRKCKQFGR 35
KAX1A_PARTR		sco	k	----EVDMRCKSSKECLVKQKQATGRPNG-KCMNRKCKQYPR 37
KAX42_TITSE	1TSK	sco	k	---VVIGQRRCYRSPDCYSACKKLVGKATG-KCTNGRCDQ--- 35
KA151_ANDAU		sco	k	--QNETNKKCQGGG-CASVCRRVIGVAAG-KCINGRCVQYP- 37
KA155_ANDAU		sco	k	--QVETNKKCQGGG-CASVCRRVIGVAAG-KCINGRCVQYP- 37
KA153_ANDMA		sco	k	--QIETNKKCQGGG-CASVCRKVIIGVAAG-KCINGRCVQYP- 37
KA152_MESMA		sco	k	--QVETNVKQGGG-CASVCRKAIIGVAAG-KCINGRCVQYP- 37
KA157_ANDAM		sco	k	NGQVQTNKKCKGGG-CASVCRKEIGVAAG-KCINGRCVQ--- 37
KA158_MESMA		sco	k	--QVQTNVKCQGGG-CASVCRREIGVAAG-KCINGRCVQYRN 38
KA156_TITDI	2AXK	sco	k	--QIDTNVKCSGSSKCVKICIDRYNTRGA-KCINGRCTQYP- 38
KA181_TITCA	2JP6	sco	k	---TGPQTTTC-QAAMCEAGCKGL-GKSMK-SCQGDTCCKKA- 35
KA182_TITDI		sco	k	---TGPQTTTC-QASTCEAGCKQI-GKSMK-SCQGDTCCKKA- 34
KA183_TITDI		sco		--GEGPKSDC-KPDLCEAAKDL-GKPMDFCKDGTCKCKD- 36

8C_N2.4.5

KA121_TITSE	1C56	sco	k	WCSTCLDLACGASRECYDPCFKAFGRAHGKCMNNKCRQYT 40
KA122_TITTR		sco	k	WCSTCLDLACGASRECYDPCFKAFGRAHGKCMNNKCRQYT 40
KA123_TITCO		sco	k	WCSTCLDLACGASRECYDPCFKAFGRAHGKCMNNKCRQYT 40

8C_N3.5.3

SCXA_BUTOM		sco	na	--VRDAYIAQNYNCVYDCARDAYCNELCTKNGAKS----GYCEWFGPHGDACWIDLPNNVPIK--VEG-KCHRK 66
SCXB_BUTOM		sco	na	--VRDAYIAQNYNCVYDCARDAYCNELCTKNGAKS----GHCEWFGPHGDACWIDLPNNVPIK--VEG-KCHRK 66
SCXD_BUTOM		sco	na	--VRDAYIAQNYNCVYHCGRDAYCNELCTKNGAKSRTRGGYCHWFGPHGDACWIDLPNNVPIK--VEG-KCHRK 70
SC12_MESMA	1OMY	sco	na	--VRDAYIAQNYNCVYHCGRDAYCNELCTKNGAKS----GSCPYLGEHKFACYCKDLDPNVPIR--VPG-KCH-- 64
SCX3_LEIQU	1LQQ	sco	na	--VRDAYIAKNYNCVYECFRDSYCNELCTKNGASS----GYCQWAGKYGNACWYALPDNVPIR--VPG-KCH-- 64
SCXA_LEIQH	2ASC	sco	na	--VRDAYIAKNYNCVYECFRDAYCNELCTKNGASS----GYCQWAGKYGNACWYALPDNVPIR--VPG-KCHRK 66

SCXC_BUTOM	sco	na	--VRDAYIAQNYNCVYTCFKDAHCNDLCTKNGASS----	GYCQWAGKYGNACWYALPDNVPIR--	IPG-KCHRK	66	
SIX1_BUTOC	sco	na	--VRDAYIAQNYNCVYFCMKDDYCNLDCTKNGASS----	GYCQWAGKYGNACWYALPDNVPIR--	IPG-KCHS-	65	
SCXB_ANDAU	sco	na	DSARDAYIAKNDNCVYECFQDSYCNLDCTKNGAKS----	GTCDWIGTYGDACLWYALPDNVPIK--	LSG-EC---	65	
SCX1_MESMA	1DJT	sco	na	--VRDAYIAKPHNCVYECARNEYCNLDCTKNGAKS----	GYCQWVGKYGNACWYALPDNVPIR--	VPG-KCH--	64
SC16_MESMA	sco	na	--VRDAYIAKPHNCVYECARNEYCNLDCTKNGAKS----	GYCQWVGKYGNACWYALPDNVPIR--	VPG-KCH--	64	
SCX2_MESMA	sco	na	--VRDAYIAKPHNCVYECARNEYCNLDCTKNGAKS----	GYCQWVGKYGNACWYALPDNVPIR--	IPG-NCH--	64	
SCM2_MESMA	1CHZ	sco	na	--VRDAYIAKPHNCVYECARNEYCNLDCTKNGAKS----	GYCQWSGKYGNACWYALPDNVPIR--	VPG-KCH--	64
SCX7_MESMA	1KV0	sco	na	--VRDGYIALPHNCAYGCLNNEFCNLDCTKNGAKI----	GYCNIVGKYGNACWYALPDNVPIR--	VPG-RCHPA	66
SC10_MESMA	sco	na	--VRDGYIALPHNCAYGCLNNEFCNLDCTKNGAKI----	GYCNIQGYGNACWYALPDNVPIR--	VPG-RCHPS	66	
SCX4_MESMA	1SN4	sco	na	--VRDAYIAKPENCVYHCAGNEGCNKLDCTDNGAES----	GYCQWGGRYGNACWYALPDNVPIR--	VPG-KCH--	64
SC13_MESMA	sco	na	--VRDAYIAKPENCVYHCAGNEGCNKLDCTDNGAES----	GYCQWGGRYGNACWYALPDNVPIR--	VPG-KCH--	64	
SCX9_MESMA	sco	na	--VRDAYIAKPENCVYHCAGNEGCNKLDCTDNGAES----	GYCQWGGRYGNACWYALPDNVPIR--	VPG-KCH--	64	
SCXB_MESMA	sco	na	--VRDAYIAKPENCVYHCAGNEGCNKLDCTDNGAES----	GYCQWGGRYGNACWYALPDNVPIR--	VPG-KCH--	64	
SCX3_MESMA	sco	na	--VRDAYIAKPENCVYHCAGNEGCNKLDCTDNGAES----	GYCQWGGRYGNACWYALPDNVPIR--	VWG-KCHG-	65	
SCXA_MESMA	sco	na	k	--VRDAYIAKPENCVYHCAGNEGCNKLDCTDNGAES----	GYCQWGGRYGNACWYALPDNVPIR--	VPG-KCQ--	64
SCX8_MESMA	1SNB	sco	na	--GRDAYIADSENCTYFCGSNPYCNLDCTENGAKS----	GYCQWAGRYGNACWYALPDNVPIR--	EPG-KCG--	64
MKTX2_MESMA	sco			--GRDAYIADSENCTYFCGSNPYCNLDCTENGAKS----	GYCQWAGRYGNACWYALPDNVPIR--	IPG-PCR--	64
MKTX1_MESMA	sco			--GRDAYIADSENCTYFCALNPYCNLDCTKNGAKS----	GYCQWAGRYGNACWYALPDNVPIR--	ISG-SCR--	64
MKTX3_MESMA	sco			--GRDAYIAKKENCTYFCALNPYCNLDCTKNGAKS----	GYCQWAGRYGNACWYALPDNVPIR--	IPG-PCI--	64
SCX1_BUTOC	sco	na	--GRDAYIAQPENCVYECANYSYCNLDCTKNGATS----	GYCQWLKGYGNACWYALPDNVPIR--	IPG-KCHF-	65	
SCX4_BUTOM	sco	na	--GRDAYIAQPENCVYECANYSYCNLDCTKNGAKS----	GYCQWLKGYGNACWYALPDNVPIR--	IPG-KCHF-	65	
SCX2_BUTOC	sco	na	--GRDAYIAQPENCVYECANYSYCNLDCTKNGAKS----	GYCQWLGRWGNACWYALPDNVPIR--	IEG-KCHF-	65	
SCXI_BUTOM	sco	na	--GRDAYIAQPENCVYECANYSYCNLDCTKNGAKS----	GYCQWLGRWGNACWYALPDNVPIR--	IEG-KCHFA	66	
SCX1_MESTA	2A7T	sco	na	--GEDGYIADGDNCTYICTFNNYCHALCTDKKGDSD----	GACDWWVPYGVVWYALPDNVPIR--	GSG-KCR--	64
SCX1_ORTSC	sco	na	--ERDGYIVQLHNCVYHCGLNPNYCNLDCTKNGATS----	GSYCQWMTKWGNACWYALPDNVPIR--	LDP-KCY--	66	
SCXE_BUTEU	sco	na	--ARDAYIADDRNCVYTCALNPYCDSECTKNGAD----	GSYCQWLGRFGNACWYALPDNVPIR--	IPGEECR--	66	
SCX2_LEIQH	sco	na	--IKDGYIVDDVNCTYFCGRNAYCNEECTKLGES----	GYCQWASPYGNACWYALPDNVPIR--	GPG-RCR--	64	
SCX8_ANDMA	sco	na	--LKDGYIVNDINCTYFCGRNAYCNEECTKLGES----	GYCQWASPYGNACWYALPDNVPIR--	GPG-RCND-	65	
SCXB_BUTOC	sco	na	--LKDGYIVDDRNCTYFCGTNAYCNEECTKLGES----	GYCQWVGKYGNACWYALPDNVPIR--	QAG-RCRS-	65	
SCX5_ANDMA	sco	na	--LKDGYIIDDLNCTFFCGRNAYCDDECTKKGES----	GYCQWASPYGNACWYALPDNVPIR--	EKG-RCN--	64	
SCX5_LEIQU	sco	na	--LKDGYIVDDKNCTFFCGRNAYCDECTKKGES----	GYCQWASPYGNACWYALPDNVPIR--	EKG-RCN--	64	
SC11_MESMA	sco	na	--VKDGYIADDRNCVYFCGRNAYCDDECTKNGAES----	GYCQWASKYGNACWYALPDNVPIR--	KPG-RCNGG	66	
SC15_MESMA	sco	na	--VRDGYIADDKNCAYFCGRNAYCDDECTKNGAES----	GYCQWAGVYGNACWYALPDNVPIR--	VPG-KCNGG	66	
SCAA_MESMA	sco	na	--VRDGYIADDKNCAYFCGRNAYCDDECTKNGAES----	GYCQWAGVYGNACWYALPDNVPIR--	VPG-KCNGG	66	
SCA3_MESMA	sco	na	--VRDGYIADDKNCAYFCGRNAYCDDECTKNGAES----	GYCQWAGVYGNACWYALPDNVPIR--	VPG-KCNGG	66	
SCA4_MESMA	sco	na	--VRDGYIADDKNCAYFCGRNAYCDDECTKNGAES----	GYCQWAGVYGNACWYALPDNVPIR--	VPG-KCNGG	66	
SCA1_MESMA	sco	na	--VRDGYIADDKNCAYFCGRNAYCDDECTKNGAES----	GYCQWAGVYGNACWYALPDNVPIR--	VPG-KCNGG	66	
SCBK_MESMA	sco	na	--VRDGYIADDKNCAYFCGRNAYCDDECTKNGAES----	GYCQWAGVYGNACWYALPDNVPIR--	VSG-ECQQ-	65	
SCXA_BUTEU	sco	na	--VRDGYIADDKNCAYFCGRNAYCDDECTKNGAES----	GKCWYAGQYGNACWYALPDNVPIR--	QKQVSG-KCN--	65	
SCX2_ANDAM	sco	na	--VRDGYIADAGNCGYTCVANDYCNTECTKNGAES----	GYCQWFGRYGNACWYALPDNVPIR--	VPG-KCN--	64	

SIX3G_LEIQH	sco	na	---DGYIRGG-DGCKVSCVINHVFCDFNEKAAGGSYG-YCWGWGLA---CWTEGLPAER-EWDYE-TDTCG---	61
SIX3H_LEIQH	sco	na	---DGYIRGG-DGCKVSCVINHVFCDFNEKAAGGSYG-YCWAAGLA---CWTEGLPAER-EWDYE-TDTCG---	61
SIX3E_LEIQH	sco	na	---DGYIRGG-DGCKVSCVINHVFCDFNEKAAGGSYG-YCWAAGLA---CWTEGLPADR-EWDYE-TDTCG---	61
SIX3F_LEIQH	sco	na	---DGYIRGG-DGCKVSCVINHVFCDFNEKAAGGSYG-YCWGWGLA---CWTEGLPADR-EWDYE-TDTCG---	61
SIX3B_LEIQH	sco	na	---DGYIRGG-DGCKVSCVINHVFCDFNEKAAGGSYG-YCWGWGLA---CWTEGLPADR-EWDYE-TNTCG---	61
SIX3C_LEIQH	sco	na	---DGYIRGG-DGCKVSCVINHVFCDFNEKAAGGSYG-YCWAAGLA---CWTEGLPAER-EWDYE-TNTCG---	61
SIX3D_LEIQH	sco	na	---DGYIRGG-DGCKVSCVINHVFCDFNEKAAGGSYG-YCWGWGLA---CWTEGLPAER-EWDYE-TNTCG---	61
SIX3A_LEIQH	sco	na	---DGYIRGG-DGCKVSCVINHVFCDFNEKAAGGSYG-YCWAAGLA---CWTEGLPADR-EWDYE-TNTCG---	61
SIX5_LEIQH	sco	na	---DGYIRGG-DGCKVSCVIDHVFCDFNEKAAGGSYG-YCWGWGLA---CWTEGLPADR-EWKYE-TNTCG---	61
SCAT_MESMA	sco	k	--DNGYLLNKYTGCKIWCVINNESCNSECKLRRGNYG-YCYFWKLA---CYTEGAPK-SELWAYE-TNKCNGKM	66
SCB1_LEIQH	sco	na	--DNGYLLNKATGCKVWCVINNASCNSECKLRRGNYG-YCYFWKLA---CYTEGAPK-SELWAYA-TNKCNGKL	66
SCAS_MESMA	sco	na	--DNGYLLDKYTGCKVWCVINNESCNSECKIRGGYYG-YCYFWKLA---CFTEQGARK-SELWNYN-TNKCNGKL	66
SCXB_CENNO	sco	na	-ARDGYPVDE-KGCKLSCILINDKWCNSACHSRGGKYG-YCYTGGLA---CYTEAVPDNVKVVWTYE-TNTC---	63
SIX6_BUTOC	sco	na	---DGYPKQK-NGCKYDCIINNKCNGIICKMHGGYYG-YCWGWGLA---CWTEGLPE-DKKWWE-TNKCGR---	62
SIXI_ORTSC	sco	na	---DGYPKQK-DGCKYSCINHKFCNSVCKSNGGDYG-YCWFWGLA---CWTEGLPD-NKMWKYE-TNTCG---	61
SIX2_BUTSI	sco	na	---DGYIKKS-KGCKVSCVINNVYCNSMCKSLGGSYG-YCWTYGLA---CWTEGLPN-AKRWKYE-TKTK---	61
SIX3_BUTSI	sco	na	---DGYILNS-KGCKVSCVVSIVYCNSMCKSSGGSYG-YCWTWGLA---CWTEGLPN-SKRWTSS-KNKC---	61
SIX4_BUTSI	sco	na	---DGYIKGN-KGCKVSCVINNVFCNSMCKSSGGSYG-YCWSWGLA---CWTEGLPA-AKKWLYAATNTCG---	62
SCX2_BUTEU	sco	na	--ADGYVKGK-SGCKISCFLDNLDLONADCKYYGGKLNWCIPDKSGY--CWTEPNKGWNS--IKSE-TNTC----	62
SCL3_LEIQH	1FH3	sco na	-VRDGYIAQP-ENCVYHCFPGSSGCDTLCKEKGGSYGH-CGFKVGHGLACWONALPDNVGIIVEG--EKCHS---	67
SCL7_LEIQH	sco	na	-VRDGYIAKP-ENCAHHCFFPGSSGCDTLCKENGGTGGH-CGFKVGHGTACWONALPDKVGIIVDG--VKCH---	66
SCXE_BUTOC	sco	na	-VRDGYIAQP-HNCAYHCLKIISGCDTLCKENGATSGH-CGHKSGHGSACWOKDLPDKVGIIVHG--EKCHR---	67
SCX3_BUTOM	sco	na	-GRDGYIAQP-ENCVYHCFPGSSGCDTLCKEKGATSGH-CGFLPGSGVACWONLNPKNVPIVVG--EKCH---	66
SCX3_ORTSC	sco	na	GVRDGYIAQP-HNCVYHCFPGSSGCDTLCKENGATQSSCFILRGTACWOKDLPDRVGIIVDG--EKCH---	67
SCX3_ANDAM	sco	na	-ARDGYIAQP-NNCVYHCIPSPGCDKLCRENGATSGK-CSFLAGSGLACWONVALPDNVPIKIIG--QKCT---	66
KURT_PARTR	1T1T	sco na ca	-KIDGYPVDY-WNCKRICWYNNKYCNDLCKGLKADSG-YCWGWTLN---CYTEQGLPDNARIKRS---RCRA---	63
KURT2_PARGR	sco	na ca	-KIDGYPVDY-WNCKRICWYNNKYCNDLCKGLKADSG-YCWGWTLN---CYTEQGLPDNARIKRS---RCRA---	63
KURT1_PARGR	sco	na ca	-KIDGYPVDN-WNCKRICWYNNKYCYDLCKGLKADSG-YCWGWTLN---CYTEGLPDNARIKRG---RCN---	62

: * . * * * * * * *

8C_N3.7.3

SCX1_TITBA	sco	na	KKDGYPVEADNCAFVCFGYDNAYCDKLGDK-KADSGYC-YWVHI--LCYTYGLPD---NEPTKTNGKC-----	62
SCX5_TITBA	sco	na	KKDGYPVEGDNCAFVCFGYDNAYCDKLGDK-KADSGYC-YWVHI--LCYTYGLPD---KEPTKTNGRCKP---	64
SCXV_TITSE	sco	na	KKDGYPVEGDNCAFACFGYDNAYCDKLGDK-KADDGYC-VWS-P--DCYTYGLPEHILKEPTKTSGRC-----	64
SC49A_TITCA	sco	na	-KDGYLVGNDGCKYCNLTPRGHYCANESRV-KGKDGYCYAWM----ACYTYSMPDW-VKTWSRSTNRGR---	64
SCX1_TITZU	sco	na	-KDGYLVGNDGCKYSCFTRPGTYCANESRV-KGKDGYCYAWM----ACYTYSMPNW-VKTWDTRATNRGR---	64
SC49B_TITCA	sco	na	KKEGYLVGNDGCKYGCITRPHQYCVHEELK-KGTDGYCAYWL----ACYTYNMPDW-VKTWSSATNKCK-----	64
SCX2_TITPA	sco	na	KKEGYLVGNDGCKYSCFTRPAQYCVHEELR-KGTDGYCAYWL----ACYTYNMPDH-VRTWSRATNRGRS---	65

SC48A_TITCA	sco na	NKDGYLMEDGDKMGLTRKASVCVDQKEV-GGKDGICYAWL----SCYTYNMPDS-VEIWDSKNNKCGK--- 65
SCX7_TITSE	1NPI sco na	-KEGYLMDHEGCKLSCFIRPSGYCGRECGIK-KGSSGYC-AWP----ACYTYGLPNW-VKVWDRATNKC----- 61
SCX7_TITBA	sco na	-KEGYLMDHEGCKLSCFIRPSGYCGSECKIK-KGSSGYC-AWP----ACYTYGLPNW-VKVWDRATNKC----- 61
SCX7_TITST	sco na	GKEGYLMDHEGCKLSCFIRPSGYCGRECTLK-KGSSGYC-AWP----ACYTYGLPNW-VKVWDRATNKC----- 62
SCX7_TITCO	sco na	-KEGYAMDHEGCKLSCFIRPSGYCGRECGYK-KGSSGYC-AWP----ACYTYGLPNW-VKVWERATNRC----- 61
SCX2_TITBA	sco na	-KEGYAMDHEGCKFSCFIRPAGFCFDGYCKTHLKASSGYC-AWP----ACYTYGVPSN-IKVWDYATNKC----- 62
SCXJ_TITBA	sco na	-KEGYAMDHEGCKFSCFIRPSGFCFDGYCKTHLKASSGYC-AWP----ACYTYGVPSN-IKVWDYATNKC----- 62
SCX2_TITSE	sco na	-KEGYAMDHEGCKFSCFIRPAGFCFDGYCKTHLKASSGYC-AWP----ACYTYGVPDH-IKVWDYATNKC----- 62
SCX2_TITST	sco na	-KEGYAMDHEGCKFSCFIRPAGFCFDGYCKTHLKASSGYC-AWP----ACYTYGVPDH-IKVWDYATNKC----- 62
NTXP_TITSE	sco	GREGYPADSKGCKITCFLTAAGYCNTECTLK-KGSSGYC-AWP----ACYTYGLPDS-VKIWTSETNKC----- 62
TTX6_TITSE	sco	GREGYPADSKGCKITCFLTAAGYCNTECTLK-KGSSGYC-AWP----ACYTYGLPES-VKIWTSETNKC----- 62
SCX38_TITCO	sco	GKEGYPADSKGCKVTCFFTGAGYCNTECKLQ-KASSGYC-AWP----ACYTYGLPDS-ASVWDSATNKC----- 62
SCX4_TITFA	sco na	KEGYPADSKGCKVTCFFTGAGYCNTECKLK-KASSGYC-AWP----ACYTYGLPDS-ASVWDSATNKC----- 62
SCXI_TITBA	sco na	GKEGYPVDSRGCKVTCFFTGAGYCDKECKLK-KASSGYC-AWP----ACYTYGLPDS-VPVYDNASNKCB----- 63
ARDI_TITDI	sco na	-KNGYIIEPKGCKYSFCWGSSTWCNRECKFK-KGSSGYC-AWP----ACWYGLPDN-VKIFDYNNKC----- 61
SCX9_BUTEU	sco na	ARDAYIAKPHNCVYECYNPKGSYCNDLTEN-GAESGYCQILGKYGNACWLIQLPDN---VPIRIPGKCH--- 66
SCXKI_MESMA	sco na	DSLSPWNEGDTCDYGCQRQTDFCNKICKLH-LASSGSCQQPAPFVKLCTQGLDIDNSFFFGALEKQCCKLRE 73

. * * : * . * * : : *

8C_N3.8.3

SCX5C_CENLL	sco na	KEGYLVNKSTGCKYGCFWLGNENCDMECKAKNQGGSYGYCYSFACWTEGLPDSTPTYPLPNKSCS--- 66
SCX5X_CENLL	sco na	KEGYLVNKSTGCKYGCFWLGNENCDMECKAKNQGGSYGYCYSFACWTEGLPDSTPTYPLPNKSCS--- 66
SCX5B_CENLL	sco na	KEGYLVNKSTGCKYGCFWLGNENCDKECKAKNQGGSYGYCYSFACWTEGLPDSTPTYPLPNKSCS--- 66
SCXR_CENLL	sco na	KEGYLVNKSTGCKYGCFWLGNENCDKECKAKNQGGSYGYCYSFACWTEGLPESTPTYPLPNKSCS--- 66
SCX4_CENEX	sco na	KEGYLVNKSTGCKYECFWLGNENCDKECKAKNQGGSYGYCYSFACWTEGLPESTSTYPLPNKSC--- 65
SCX1_CENSC	1VNB sco na	KEGYLVKSDGCKYDCFWLGNENCDTECKAKNQGGSYGYCYAFACWTEGLPESTPTYPLPNKSC--- 65
SCX2_CENSC	1JZA sco na	KEGYLVNKSTGCKYGCCLKGENEGCDKECKAKNQGGSYGYCYAFACWTEGLPESTPTYPLPNKSCS--- 66
SCX3_CENSC	2SN3 sco na	KEGYLVNKSTGCKYGCCLKGENEGCDKECKAKNQGGSYGYCYAFACWTEGLPESTPTYPLPNKSC--- 65
SCX3_CENSU	sco na	KEGYLVNKSTGCKYGCCLKGENEGCDKECKAKNQGGSYGYCYAFGCWTEGLPESTPTYPLPNKSCS--- 66
SCX5_CENNO	sco na	KEGYLVNKSTGCKYGCCLKGENEGCDKECKAKNQGGSYGYCYAFGCWTEGLPESTPTYPLPNKSCS--- 66
SCX1_CENNO	sco na	KEGYLVNKSTGCKYGCFWLGNENEGCDKECKAKNQGGSYGYCYAFGCWTEGLPESTPTYPLPNKTCSSK 68
SCX1_CENEX	sco na	KEGYLVSKSTGCKYECFWLGNENEGCDKECKAPNQGGYGYCHAFACWTEENLPESTPTYPIPGKSC--- 65
SCX2_CENEX	sco na	KEGYLVSKSTGCKYECFWLGNENEGCDKECKAPNQGGYGYCHAFACWTEENLPESTPTYPIPGKSC--- 65
SCX4_CENSC	sco na	KEGYMVNKSTGCSYSCPKTGESVYCDKECKAKNQGGSYGFQYSNCWTEGLPESTPTWPLDDKPCD--- 66
SCX8_CENSC	sco na	EKGYLVHEDTGCYKCTFSGENSYCDKECKS--QGGDSGICQSKACYQGLPEDTKTWPLIGKLC--- 63
SCX2_CENNO	1CN2 sco na	KEGYLVDKNTGCKYECCLKGDNDYCLRECKQQYKKGAGGYCYAFACWTEHLYEQAVVWPLPNKRC--- 66
SCX2_CENSU	sco na	KEGYLVSKSTGCKYECCLKGDNDYCLRECKQQYKKGSSGYCYAFACWTEHLYEQAVVWPLPNKTCN--- 66
SCX4_CENSU	sco na	KEGYLVNSYTGCKYECCLKGDNDYCLRECKQQYKKGSSGYCYAFGCWTEHLYEQAVVWPLPNKTCN--- 66

SCX6_CENSU	sco	na	KEGYLVNSYTGCKFECKFLGDNDYCKRECKQQYGKSSGGYCYAFGCWETHLYEQAVVWPLPNKTCN--	66
SCXM_CENSC	sco	na	KEGYLVNSYTGCKYECFLKLGNDYCLRECKRQQYGK-SGGYCYAFACWETHLYEQAVVWPLPNKTCN--	65
SCX4_CENNO	sco	na	KEGYLVNSYTGCKYECFKLGDNDYCLRECKQQYGKAGGGYCYAFGCWETHLYEQAVVWPLPNKTCN--	66
SC2B_CENLL	sco	na ca	KEGYLVNHSTGCKYECYKLGNDYCLRECKQQYGKAGGGYCYAFGCWETHLYEQAVVWPLPKTCN--	66
SCX1_CENII	sco	na	KEGYLVNHSTGCKYECYKLGNDYCLRECKQQYGKAGGGYCYAFGCWETHLYEQAVVWPLPKTCN--	66
SCX2_CENLL	sco	na	KEGYLVNHSTGCKYECFKLGDNDYCLRECKQQYGKAGGGYCYAFGCWETHLYEQAVVWPLPKTCN--	66
SCX1_CENLL	sco	na	KEGYLVNLSYTGCKYECYKLGNDYCLRECKQQYGKAGGGYCYAFGCWETHLYEQAVVWPLPKTCN--	66
SCX1_CENLI	sco	na	KEGYLVNHSTGCKYECFKLGDNDYCLRECKQQYGKAGGGYCYAFGCWETHLYEQAVVWPLPNKTCN--	66
SCX3_CENNO	sco	na	KEGYLVNLTGCKYECFKLGDNDYCLRECKARYGKAGGGYCYAFGCWETHLYEQAVVWPLPNKTCN--	66
SCE3_CENSC	sco	na	KEGYLVNYHTGCKYECFKLGDNDYCLRECKLRHKGSGGYCYAFGCWETHLYEQAVVWPLPKKCN--	66
SCX4_CENLL	sco	na	KEGYLVNYHDGCKYECYKLGNDYCLRECKLRYGKAGGGYCYAFGCWETHLYEQAVVWPLPKRCN--	66
SCX3_CENLL	sco	na	KEGYLVNYDYGCKYACFKLGDNDYCLRECKARYYKAGGGYCYAFACWETHLYEQAVVWPLPNKTCY--	66
SCX11_CENEX	sco	na	KEGYPVNIYTGCKYSCWLLGENEYCIAECK-EIGAGYG-YCHGFGWETHLYEQAVVWPLPNKTCY--	63
SCX7_CENLL	sco	na	KEGYLVNITYTGCKYICWKLGENKYCIDECK-EIGAGYG-YCYGFGWETHLYEQAVVWPLPNKTCY--	63
SCX6_CENLL	sco	na	KEGYLVNMKTGCKYGCYELGDNGYCDRCKAESGN-YG-YCYTVGCWETHLYEQAVVWPLPNKTCY--	64
SCX10_CENEX	sco	na	KDGYLVEV-TGCKKSCYKLGENDYCNRECKMKHRGGSYGYCYFFGCWETHLYEQAVVWPLPNKTCY--	64
SCX9_CENEX	sco	na	KDGYLVEV-TGCKKSCYKLGENDYCNRECKMKHRGGSYGYCYFFGCWETHLYEQAVVWPLPNKTCY--	64
SCX1_CENNO	sco	na	KDGYLVDA-KGCKKNKYKLGENDYCNRECKMKHRGGSYGYCYFFGCWETHLYEQAVVWPLPNKTCY--	65
SCXY_CENNO	sco	na	KDGYLVDV-KGCKKNKYKLGENDYCNRECKMKHRGGSYGYCYFFGCWETHLYEQAVVWPLPNKTCY--	65
SCXI_CENSC	2B3C sco	na	KDGYLVEK-TGCKKTCYKLGENDYCNRECKWKHIGGSYGYCYFFGCWETHLYEQAVVWPLPNKTCY--	64
SCXX_CENSC	sco	na	KEGYLVDV-KGCKKNWKLGDNDYCNRECKWKHIGGSYGYCYFFGCWETHLYEQAVVWPLPNKTCY--	64
SCX2_CENGR	sco	na	KDGYLVNKSTGCKYSCININDSHCNEECISSIRKGSYGYCYKFCYCIIMPSTQVYPIPGKTCSTE	68
SCX12_CENEX	sco	na	NDGYLFDKRRKRTLECIDKTGDKNCDRNCKKE--GGSFGKCSYSACWCKGLPGITPISRTPGKTC---	63
SCX9_CENSC	sco	na	EDGYLFDKRRKRTLECIDKTGDKNCDRNCKNE--GGSFGKCSYFACWCKGLPGITPISRTPGKTCKI-	65
SCX9_CENLL	sco	na	EDGYLFDKRRKRTLACIDKTGDKNCDRNCKKE--GGSFGHCSYSACWCKGLPGITPISRTPGKTC---	63

: . ** . * * . * :

APÊNDICE 4: Grupos estruturais para SiliMod – árvores NJ

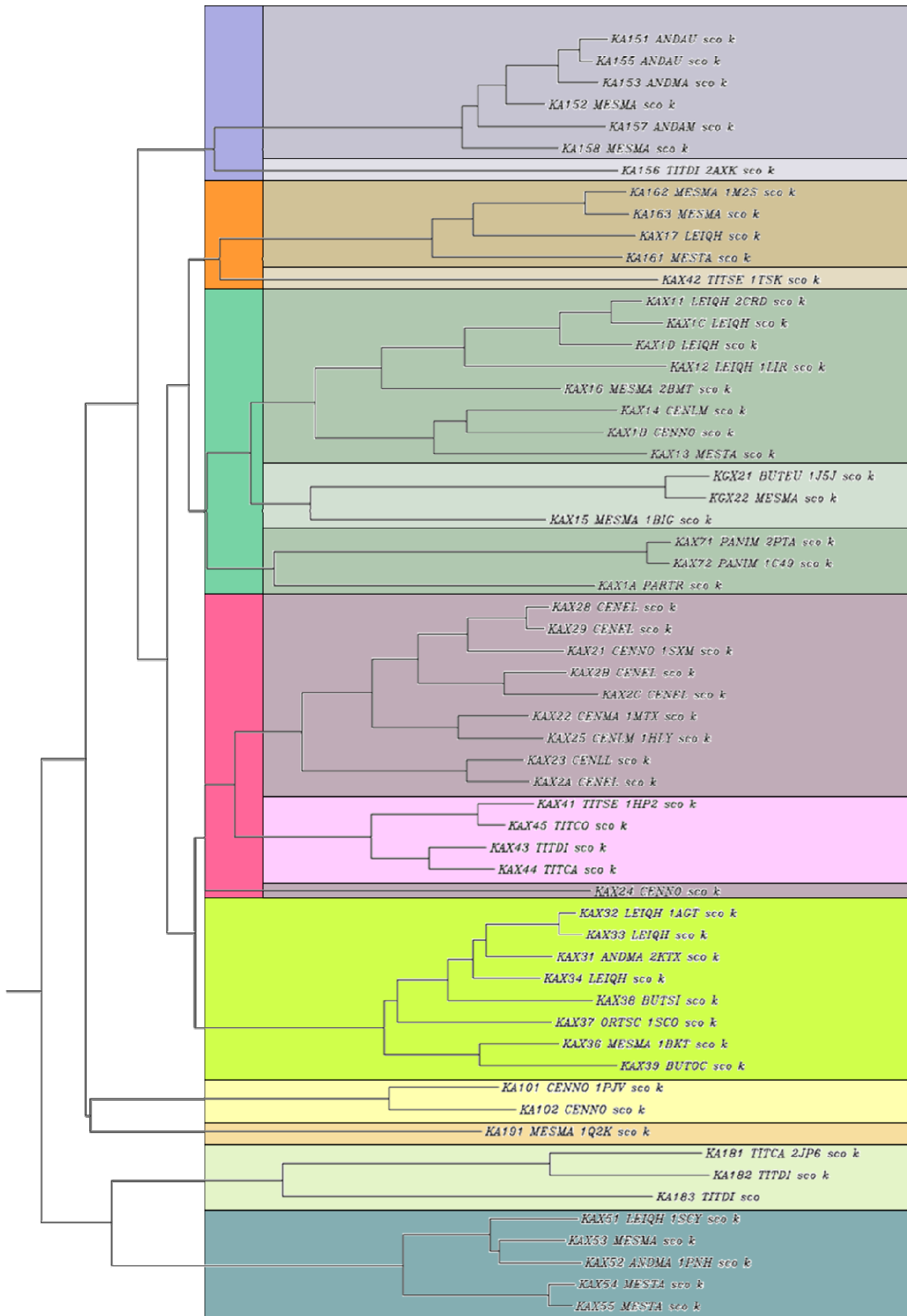


Figura i. Árvore de Neighbor-Joining das sequências de toxinas aracnídeas membros do grupo de motivo cisteína 6C_{Nx.3.x.4.1} (calculado por ClustalW), e resultado do sub-clustering do grupo estrutural.

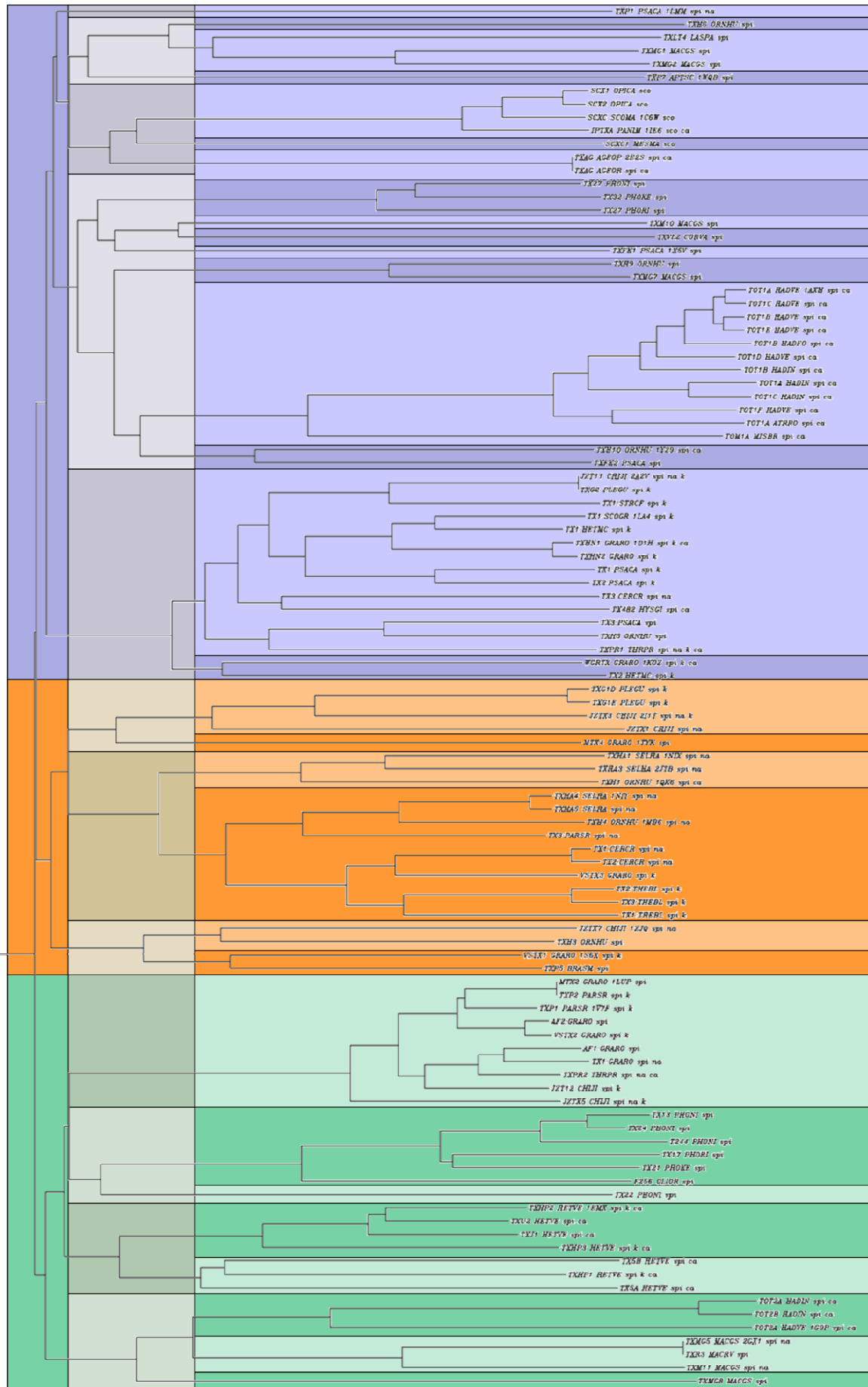


Figure ii. Árvore de Neighbor-Joining das sequências de toxinas aracnídicas membros do grupo de motivo cisteína Kozclass1.1.0.2.

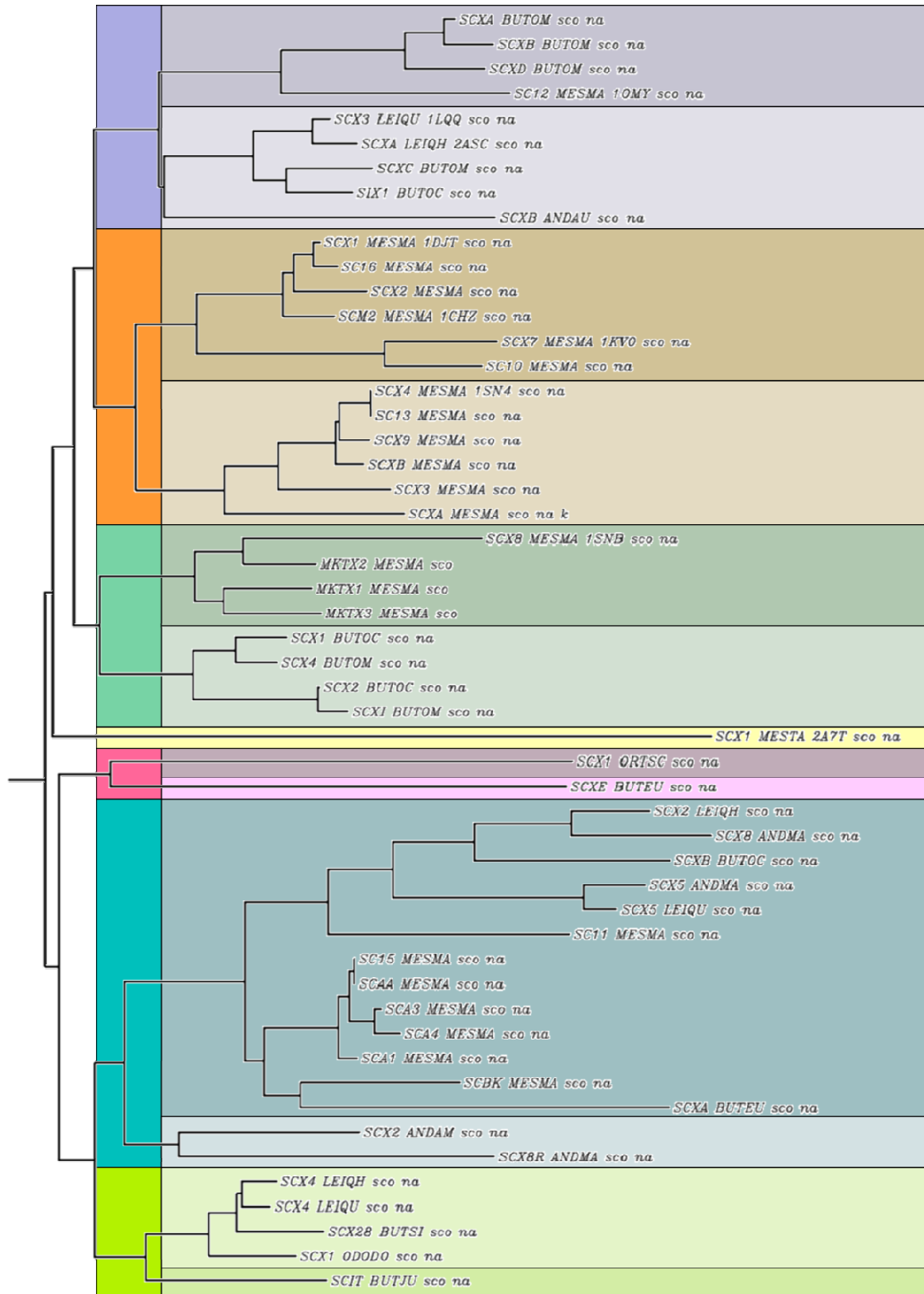


Figure iii. Árvore de Neighbor-Joining das sequências de toxinas aracnídicas membros do grupo de motivo cisteína 8C_N3.5.3.

Research article

Open Access

PEPOP: Computational design of immunogenic peptides

Violaine Moreau, Cécile Fleury, Dominique Piquier, Christophe Nguyen, Nicolas Novali, Sylvie Villard, Daniel Laune, Claude Granier[†] and Franck Molina^{*†}

Address: CNRS FRE 3009, SysDiag, CAP DELTA, 1682 Rue de la Valsière, CS 61003, 34184 Montpellier Cedex 4, France

Email: Violaine Moreau - violaine.moreau@sysdiag.cnrs.fr; Cécile Fleury - cecile.fleury@sysdiag.cnrs.fr;

Dominique Piquier - dominique.piquier@sysdiag.cnrs.fr; Christophe Nguyen - christophe.nguyen@sysdiag.cnrs.fr;

Nicolas Novali - nicolas.novali@sysdiag.cnrs.fr; Sylvie Villard - sylvie.villard@sysdiag.cnrs.fr; Daniel Laune - daniel.laune@sysdiag.cnrs.fr;

Claude Granier - claudes.granier@sysdiag.cnrs.fr; Franck Molina^{*} - franck.molina@sysdiag.cnrs.fr

^{*} Corresponding author [†]Equal contributors

Published: 30 January 2008

Received: 16 May 2007

BMC Bioinformatics 2008, **9**:71 doi:10.1186/1471-2105-9-71

Accepted: 30 January 2008

This article is available from: <http://www.biomedcentral.com/1471-2105/9/71>

© 2008 Moreau et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Most methods available to predict protein epitopes are sequence based. There is a need for methods using 3D information for prediction of discontinuous epitopes and derived immunogenic peptides.

Results: PEPOP uses the 3D coordinates of a protein both to predict clusters of surface accessible segments that might correspond to epitopes and to design peptides to be used to raise antibodies that target the cognate antigen at specific sites. To verify the ability of PEPOP to identify epitopes, 13 crystallographically defined epitopes were compared with PEPOP clusters: specificity ranged from 0.75 to 1.00, sensitivity from 0.33 to 1.00, and the positive predictive value from 0.19 to 0.89. Comparison of these results with those obtained with two other prediction algorithms showed comparable specificity and slightly better sensitivity and PPV. To prove the capacity of PEPOP to predict immunogenic peptides that induce protein cross-reactive antibodies, several peptides were designed from the 3D structure of model antigens (IA-2, TPO, and IL8) and chemically synthesized. The reactivity of the resulting anti-peptides antibodies with the cognate antigens was measured. In 80% of the cases (four out of five peptides), the flanking protein sequence process (sequence-based) of PEPOP successfully proposed peptides that elicited antibodies cross-reacting with the parent proteins. Polyclonal antibodies raised against peptides designed from amino acids which are spatially close in the protein, but separated in the sequence, could also be obtained, although they were much less reactive. The capacity of PEPOP to design immunogenic peptides that induce antibodies suitable for a sandwich capture assay was also demonstrated.

Conclusion: PEPOP has the potential to guide experimentalists that want to localize an epitope or design immunogenic peptides for raising antibodies which target proteins at specific sites. More successful predictions of immunogenic peptides were obtained when a peptide was continuous as compared with peptides corresponding to discontinuous epitopes. PEPOP is available for use at <http://diagtools.sysdiag.cnrs.fr/PEPOP/>.

Background

In antibody-antigen (Ab-Ag) interactions, the paratope of the Ab binds to the epitope of the Ag. The identification of epitopes is an important step for understanding molecular recognition rules and is also helpful for diagnosis of diseases and for drug and vaccine design. The ultimate method to precisely define an epitope is to solve the 3D structure of the Ab-Ag complex either by X-ray crystallography or NMR [1]. These techniques are, however, demanding and generally time-consuming. Faster epitope identification methods have been described such as site-directed mutagenesis of the Ag [2,3]. Another popular approach to map an epitope is parallel peptide synthesis [4,5], based on the synthesis of overlapping peptides covering the entire Ag sequence. In this case, mainly continuous (sequential or linear) epitopes can be identified. Screening chemical or biological combinatorial libraries [6] for Ab binders allows selection of peptides also called mimotopes [7], mimicking more or less faithfully the epitope. Bioinformatics tools have been developed to help experimentalists in localizing the epitope by the sequence analysis of the selected mimotopes [8,9].

Synthetic peptides are commonly used as immunogens to raise anti-peptide Abs that may cross-react with proteins [10], thus allowing their detection and quantification. These peptides are generally designed by using methods that attempt to predict antigenic determinants of a protein. Numerous algorithms have been developed over the past 25 years. They are based on different theoretical physicochemical characteristics of the target protein such as hydrophilicity, flexibility, accessibility, and secondary structure, especially turns [11]. Other methods are combinations of the latter approaches [12], the most recent [13] being an extension and combination of the methods of Parker *et al.* [14] and Jameson and Wolf [15]. Likewise, Welling *et al.* [16] developed an antigenicity scale, with the aim of predicting antigenic regions and synthesizing the corresponding antigenic peptides to elicit Abs reactive with the intact protein. All these algorithms have led to the development of several softwares or web interfaces that make the use of such methods very easy. It is, however, difficult to assess the efficacy of all predictive methods. A comparative study published some years ago [11,17] indicated that the most accurate predictive method at that time is based on the prediction of turns. This method was implemented in BEPITOPE [18]. A more recent and more exhaustive comparative study [19] concluded that the methods based on sequence analysis do not predict epitopes better than chance.

All these methods predict antigenic determinants from the protein sequence alone, neglecting 3D structure information. This is surprising because the 3D structure of an increasing number of proteins has been solved by X-ray

crystallography or NMR, and predictive modeling methods are available that show increasing accuracy [20]. Recently, however, a few recent studies [21-24] propose bioinformatics tools based on 3D information to predict epitopes.

In this article, we describe PEPOP, an algorithm that makes use of the 3D information of a protein to predict peptides which could serve as immunogens to raise site-specific anti-protein Abs. Clusters of surface accessible segments of the protein are first identified by PEPOP, and this information is further used to design the peptides. We analyzed how PEPOP clusters corresponded to structurally defined epitopes (dataset of 13 epitopes on 8 antigens) and how Abs raised against peptides designed by PEPOP reacted with the parent protein.

Results

Clustering of exposed segments of the Ag

A- PEPOP features and outputs for clustering

We developed the PEPOP algorithm as a new method intended to identify peptide sequences that, when injected into animals, induce the production of Abs that should recognize specific areas of a protein. From the 3D structure of a protein, PEPOP first identifies segments composed of accessible and sequence contiguous amino acids. Then, these segments are clustered according to their spatial distances (Figure 1A). Clusters and their segments are then further used to design immunogenic peptides. The PEPOP interface was designed so as to provide both detailed information (atomic coordinates, distance matrix, etc.) and modifiable views of the cluster(s) in the 3D context of the protein (Figure 1B).

B- Experimental evaluation of PEPOP capacity to predict antigenic epitopes

Since there is a documented relationship between surface accessibility and antigenicity [25,26], we first evaluated the correspondence between the surface clusters identified by the PEPOP algorithm on given proteins and their known epitopes. To this end, 13 Ab-Ag complexes for which the amino acids composing the discontinuous epitope have been identified precisely were studied (selected from the literature and from the PDB [27]). A detailed analysis was first carried out on hen egg lysozyme (HEL) because several structurally, well-defined epitopes have been identified (Table 1) by X-ray crystallography (resolutions from 1.5 to 3.22 Å). The PEPOP algorithm was run on the 3D structure of HEL [PDB: [1SFG](#)]. PEPOP identified 23 segments of accessible residues which were automatically grouped into four clusters on the surface of HEL (Table 1). Cluster 1 contains antigenic residues belonging to three epitopes (D11.15, HyHEL-10, and HyHEL-63). Cluster 2 exactly matches the D1.3 epitope and 9 of 10 residues of the D11.15 epitope; it also predicts

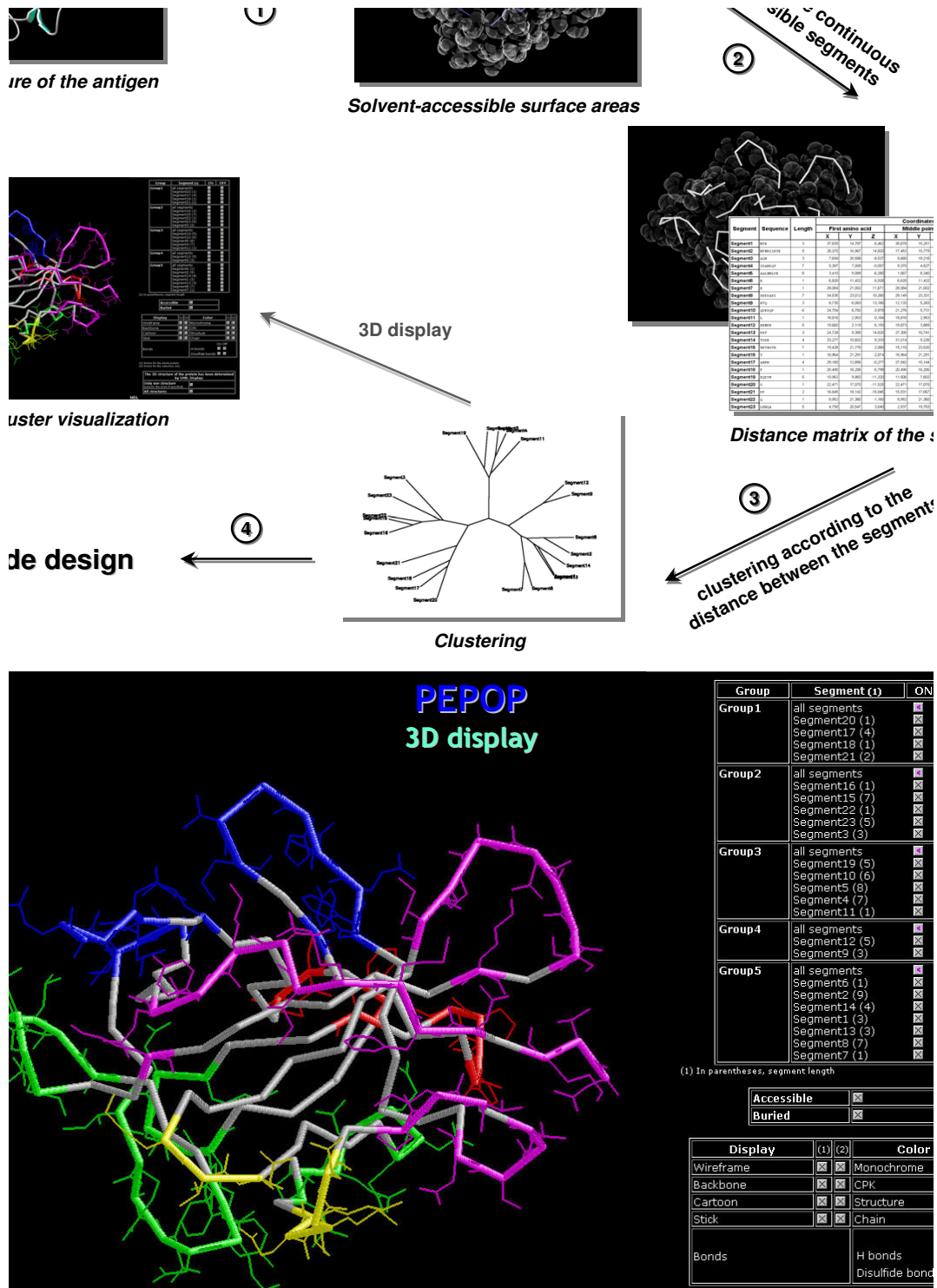


Figure 1
PEPOP. (A) Description of PEPPOP. Step 1: The solvent accessible surface area is calculated from the 3D structure of a protein. Step 2: Segments of accessible and contiguous amino acids are listed, and the shortest Euclidean distance matrix between segments is calculated. Step 3: The segments are clustered according to the matrix (clusters can be displayed on the 3D structure of the Ag). Step 4: Peptides are designed according to the clustering analyses. (B) The PEPPOP interface. Segments and clusters can interactively be displayed on a 3D view of the Ag.

Table 1: Distribution of the residues of the HEL epitopes in the clusters identified by PEPOP

PEPOP cluster	PEPOP segment	Position ([PDB:LSFG] chain A)	mAb ¹					
			D1.3	D11.15	D44.1	HyHEL-5	HyHEL-10	HyHEL-63
Cluster1	Segment17	93–94		1/10 (103)			7/16 (93, 96, 97, 100, 101, 102, 103)	7/23 (93, 96, 97, 100, 101, 102, 103)
	Segment18	96–97						
	Segment19	100–103						
Cluster2	Segment2	10–16	10/10 (13, 14, 19, 21, 22, 24, 117, 119, 121, 125)	9/10 (21, 23, 106, 112, 113, 116, 117, 118, 119)			5/16 (15, 16, 19, 20, 21)	8/23 (13, 14, 15, 16, 18, 19, 20, 21)
	Segment20	106–109						
	Segment21	111–114						
	Segment22	116–119						
	Segment23	121–129						
	Segment3	18–24						
	Segment4	27–28						
Cluster3	Segment10	56–57			11/11 (41, 45, 46, 47, 49, 53, 67, 68, 70, 81, 84)	13/14 (41, 43, 44, 45, 46, 47, 63, 73, 75, 48, 49, 89) 53, 67, 68, 70, 84)	4/16 (63, 73, 75, 89)	6/23 (62, 63, 73, 75, 77, 89)
	Segment11	59						
	Segment12	61–63						
	Segment13	65–68						
	Segment14	70–79						
	Segment15	81–82						
	Segment16	84–90						
	Segment8	41–49						
	Segment9	53						
Cluster4	Segment1	1–8						
	Segment5	33–35						
	Segment6	37						
	Segment7	39						
	amino acids not accessible						1/14 (69)	2/23 (98, 99)

¹ Ratio of the number of amino acids in the epitope included in the cluster to the total number of amino acids in the epitope. The epitopes were defined as in the original publication (D1.3 [54], D11.15 [55], D44.1 [56], HyHEL-5 [57], HyHEL-10 [58], HyHEL-63 [59]). Positions of the amino acids in the epitope included in the cluster are indicated in parentheses. The ratios are in bold for the major clusters.

some of the antigenic residues of the HyHEL-10 and HyHEL-63 epitopes. Cluster 3 perfectly matches the D44.1 epitope and almost perfectly the HyHEL-5 epitope (13 of 14 residues); it also predicts part of the HyHEL-10 and HyHEL-63 epitopes. Cluster 4 does not correspond to any X-ray described epitope of the lysozyme.

Next, seven additional Ab-Ag complexes were analyzed to extend these observations to a database of 13 crystallographic epitopes (Table 2). PEPOP analyses were performed on the 3D structures of the Ag alone, i.e., not complexed with the mAb specific for the epitope studied because this is the most frequently encountered case. The distribution of the amino acids in the epitope among the clusters identified by PEPOP was analyzed, and the number of the amino acids in the epitope in the major cluster (cluster containing the greatest number of amino acids of the epitope) was calculated. The results show that the experimentally identified amino acids in the epitope

belong to a single cluster for the D1.3, D44.1, and HyHEL-5 epitopes in HEL and the 5G9 epitope in tissue factor, to two clusters for the D11.15 epitope in HEL and the Jel42 epitope in the histidine-containing phosphocarrier protein HPr or to three clusters for the complexes BH151 – hemagglutinin, Bo2C11 – C2 domain of FVIII, NC41 – neuraminidase, F9.13.7 – guinea fowl lysozyme (GEL), HyHEL-10 – HEL, HyHEL-63 – HEL, and N10 – staphylococcal nuclease (SN). For example, all the amino acids of the epitope on HEL recognized by mAb D44.1 are included in a single cluster identified by PEPOP (Figure 2A), and 13 amino acids out of 16 of the epitope on HPr recognized by mAb Jel42 are included in the major cluster (Figure 2B). Table 2 shows that the specificity of the method ranged from 0.75 to 1.00 (median value 0.87) and the sensitivity varied within a broader range (0.33 to 1.00; median value: 0.63). The positive predictive value (PPV) varied from 0.19 to 0.89 (mean value 0.43; median value 0.33). The same database of 13 crystallographically

Table 2: Evaluation and comparison of the performances of PEPOP

Ab – Ag complex ¹	PDB	Epitope (number of amino acids)	PEPOP			CEP			DiscoTope				
			Number of clusters containing epitopic residues/total number of clusters	Number of predicted residues of the epitope	Sp	Se	PPV	Sp	Se	PPV	Sp	Se	PPV
D1.3 – hen egg lysozyme [54]	<u>ISFG_A</u>	10	1/4	10	0.77	1.00	0.27	0.76	1.00	0.26	0.80	0.40	0.14
D44.1 – hen egg lysozyme [56]	<u>IT6V_M</u>	11	1/5	11	0.81	1.00	0.33	0.78	0.91	0.28	0.86	0.55	0.27
HyHEL-5 – hen egg lysozyme [57]	<u>IVDP_B</u>	14	1/5	13	0.83	0.93	0.41	0.88	0.00	0.00	0.78	0.00	0.00
Jel42 – HPr [60]	<u>IPOH</u>	16	2/8	13	0.93	0.81	0.72	0.90	0.75	0.63	0.97	0.06	0.33
D11.15 – hen egg lysozyme [55]	<u>IHEL</u>	10	2/5	8	0.95	0.80	0.57	0.90	0.10	0.08	0.82	0.30	0.13
5G9 – tissue factor [61]	<u>IWV7_T</u>	18	1/3	14	0.75	0.78	0.25	0.84	0.28	0.15	0.84	0.17	0.10
BH151 – hemagglutinin [62]	<u>5HMG_C</u>	19	3/7	12	0.83	0.63	0.19	0.97	0.32	0.40	0.75	0.11	0.03
Bo2C11 – FVIII C2 domain [63]	<u>ID7P_M</u>	15	3/5	8	0.92	0.53	0.42	0.94	0.47	0.44	0.83	0.53	0.25
NC41 – neuraminidase [64]	<u>INMC_N</u>	22	3/4	11	0.90	0.50	0.23	0.99	0.59	0.81	0.85	0.64	0.20
F9.13.7 – guinea fowl lysozyme [65]	<u>IHHL</u>	10	3/5	5	0.98	0.50	0.71	0.92	0.50	0.33	0.83	0.40	0.17
HyHEL-10 – hen egg lysozyme [58]	<u>IUCO_A</u>	16	3/5	8	0.99	0.50	0.89	0.96	0.56	0.64	0.87	0.44	0.32
HyHEL-63 – hen egg lysozyme [59]	<u>IVFB_C</u>	23	3/4	9	0.82	0.39	0.32	0.73	0.30	0.19	0.87	0.39	0.39
N10 – staphylococcal nuclease [66]	<u>I EYO_A</u>	18	3/6	6	0.87	0.33	0.29	0.88	0.06	0.07	0.66	0.33	0.13
Average					0.87	0.67	0.43	0.88	0.45	0.33	0.83	0.33	0.19
Median					0.87	0.63	0.33	0.90	0.47	0.28	0.83	0.39	0.17

¹ The corresponding reference is given in parentheses
nd: not determined (no results returned from the server)
Sp: specificity; Se: sensibility; PPV: positive predictive value (see Materials and methods)

defined epitopes was used with two freely available web tools that also make use of the 3D information of the protein to predict epitopes (DiscoTope [23,28] and CEP [24,29]). The results in Table 2 show that both methods have similar specificity (median value for DiscoTope: 0.83; median value for CEP: 0.90). Their sensitivities, however, are slightly lower than that of PEPOP (median value for DiscoTope: 0.39; median value for CEP: 0.47) and the PPV are again lower than that found for PEPOP predictions (median value for DiscoTope: 0.17; median value for CEP: 0.28) (Table 2). Thus, performances of PEPOP compare well with similar, but not identical, algorithms.

Design of peptides from clusters

A- PEPOP features and outputs for the design of immunogenic peptides

Based on PEPOP prediction of clusters of surface accessible segments, the design of peptides suitable for raising Abs potentially cross-reactive with the target protein can be achieved by different methods. To build *in silico* a candidate peptide, a segment, called a "reference segment", is first chosen, and its sequence is extended with a "method

of extension" in a specified "area of extension" on the protein to yield the designed peptide. Each of the reference segments, the method, and the area of extension can be selected among several possibilities. Figure 3 illustrates the method of adding the nearest neighbor (NN) segments to the reference segment to obtain a suitable peptide. Peptides #5 and #6 in Table 3 were constructed from the longest segment to which the nearest neighbor segments were added, according to two different methods (respectively, NN and NNd, for segments synthesized using D-amino acids). The software was developed in such a way that, at each of the three steps of the design of a given peptide, the user can choose the parameter (for example, inclusion or not of a part of the protein in the peptide) or let the algorithm automatically do it. By default, the reference segment is the longest segment, the method of extension is the addition of the protein sequence and in this case it is not necessary to select an area of extension. The peptide is therefore extended until the default minimal length of 20 amino acids is reached.

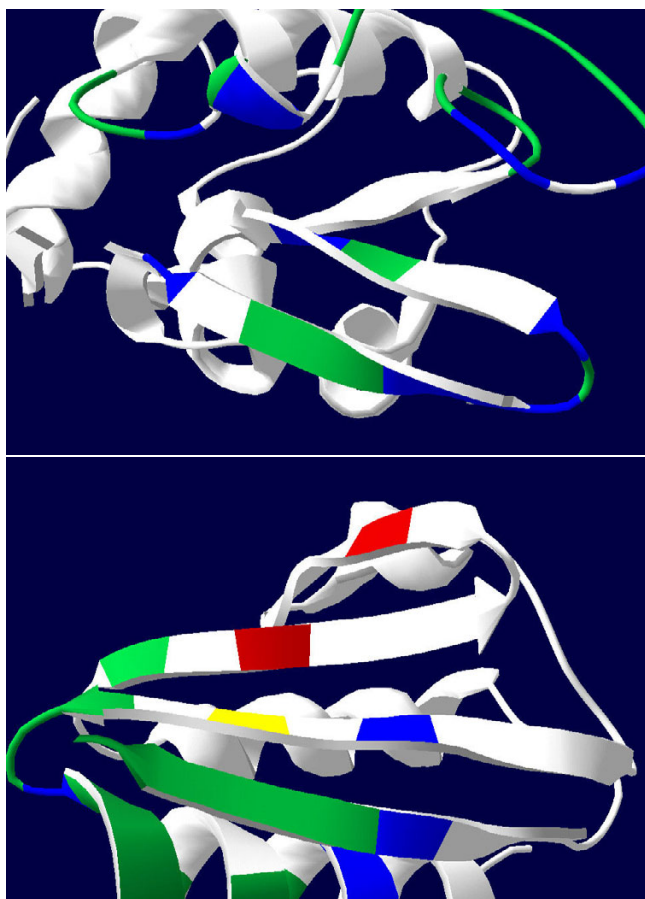


Figure 2
Comparison of known epitopes with the predicted major cluster. Amino acids in the epitope and in the major cluster are in green; those outside of the major cluster are in red; those non-accessible are in yellow; amino acids in the major cluster but not in the epitope are in blue. **(A)** Epitope on HEL recognized by mAb D44.1 **(B)** Epitope on HPr recognized by mAb Jel42.

B- Experimental evaluation of the immunogenicity of designed peptides

To prove the capacity of PEPOP to successfully predict peptides that are able to induce Abs able to recognize the cognate protein, several peptides were designed from the 3D structure of three model Ags and chemically synthesized, then rabbits were immunized with these peptides (conjugated to a carrier protein), and the reactivity of the resulting Abs with the cognate Ags was measured. From the 3D structures of the PTP domain of IA-2, the MPO-like domain of TPO, and IL8, PEPOP was run to predict clusters of surface accessible residues and to design peptides from them. Several of the possible methods to construct peptides *in silico* were used although not all could be tested on the three Ags, thus precluding a strict comparison of methods. Nevertheless, five peptides (peptides

#4–8) were designed from the IA-2 model, four from the TPO model (peptides #13–16), and two from the 3D structure of IL8 (peptides #17 and #18). All but three of the peptides predicted by PEPOP have the longest identified segment as reference segment. The other three peptides started from the top-ranked segment either of the top-ranked cluster (peptide #15) or of a user-selected cluster (peptide #8 and #18). To predict the peptides, the simplest method [the flanking protein sequence (FPS) process] was first tested (peptides #4, #8, #13, #17, and #18 in Table 3). Then, the ability of the software to predict immunogenic peptides corresponding to a discontinuous epitope of the protein was experimented with the use of the NN method to design peptides #5, #6, #14 and #15. The design process is illustrated in Figure 3. Some other peptides were designed "manually", i.e., by using user-defined information (peptides #7 and #16 in Table 3). To evaluate the method, seven peptides (peptides #1, #2, #3, #9, #10, #11, and #12) were predicted from the amino acid sequence of TPO and IA-2 by using the standard algorithm BEPITOPE [11]. This method predicts peptides from the protein sequence, the corresponding epitopes thus being continuous.

Synthetic peptides were prepared according to the designed sequences, then coupled to KLH and used to immunize rabbits. To satisfy the predictive goal, anti-peptide polyclonal Abs must recognize the predicted peptide and cross-react with the cognate protein. Table 3 presents the reactivity with the cognate Ag of the Abs obtained by immunization with the different peptides (all Abs reacted strongly with the cognate peptide; results not shown). The best results (highest ELISA reactivity with the Ag) were obtained with peptides designed by the FPS method since four out of five such peptides led to the production of Abs that reacted strongly with the protein Ag (peptides #4 and #8 with IA-2, peptide #13 with TPO, and peptide #18 with IL8). Only peptide #17 failed to recognize the IL8 Ag. Peptides designed by the FPS method of PEPOP led more frequently to strongly reactive sera than peptides selected according to the predictions of BEPITOPE: two control peptides (#1 and #9) induced an Ab response of the same quality (as measured by ELISA), whereas the five other control peptides (peptide #2; #3, #10, and #12) gave rise to medium-range reactivities (Table 3), and peptide #11 did not lead to any significant Ab response. The two peptides (#7 and #16) that were "manually designed" as well as the retro-inverso peptide (#6) did not lead to any significant Ab response. Two peptides out of the four designed by using the NN method of PEPOP gave intermediate cross-reactivity with the cognate Ag reactivity (peptides #14 and #15 on TPO). In summary, the simple FPS method (which provides continuous sequences) proved to be very efficient to yield peptides which, when conveniently coupled to a suitable carrier, induced a

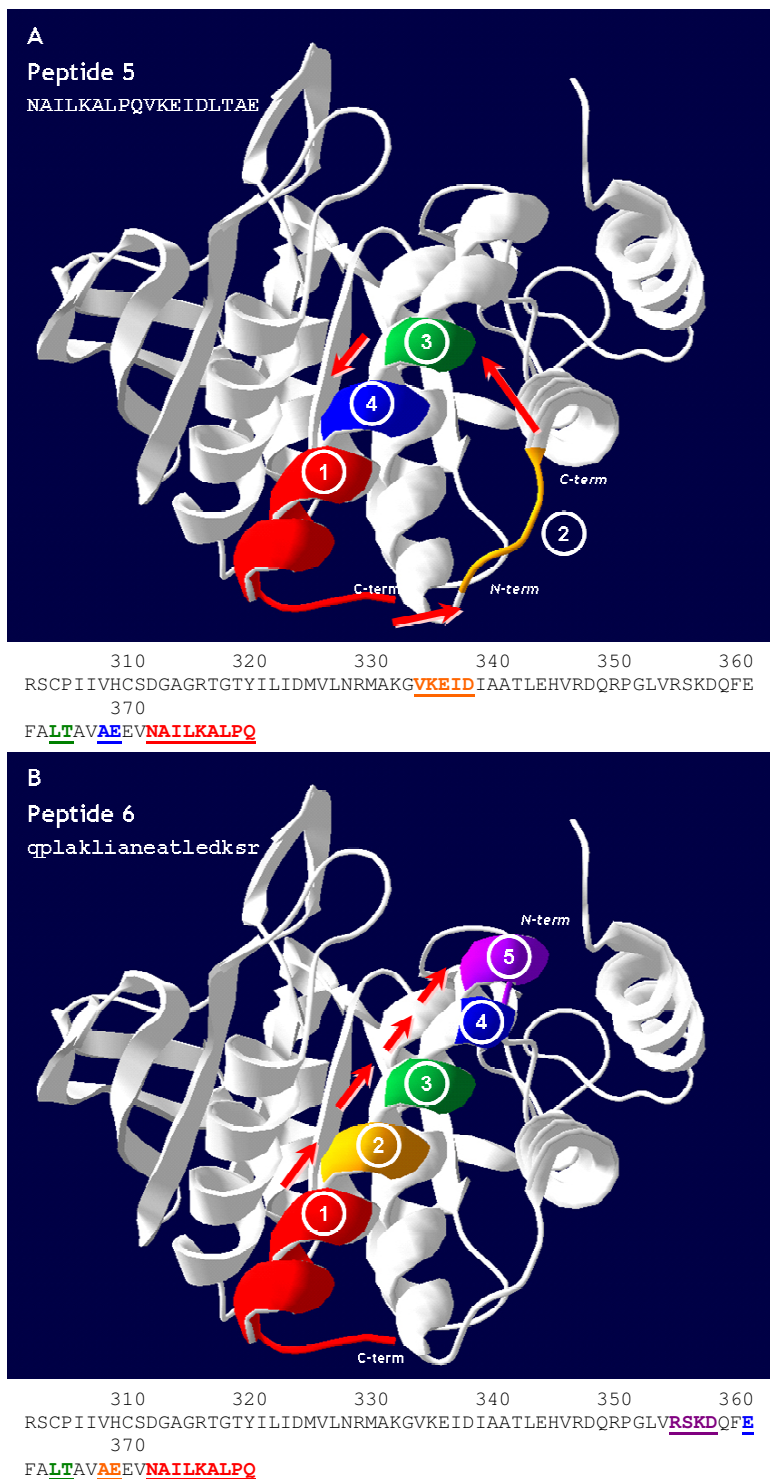


Figure 3
Example of peptides designed from the 3D structure of IA-2. The two peptides, #5 (A) and #6 (B) of Table 3, were constructed from the longest segment (segment 1: reference segment) to which were added the nearest neighbor segments (segment 2 to respectively segment 4 and segment 5), according to two different methods (respectively method NN and NNd). Segments indicated in the IA-2 sequence are in the same color as in the 3D structure.

Table 3: Reactivity with the cognate Ag of the Abs obtained by immunization with the peptides

Ag	Peptide #	Peptide sequence ¹	Reference segment ³	Method ² Method of extension ⁴	Area of extension ⁵	A _{450 nm} for serum diluted 1:10 000 ⁶
IA-2	1	VSSQFSDAAQASPS		B		1.75
	2	SPSSHSSTPSWCEEP		B		0.30
	3	RYWPDGASLYHVVEV		B		0.36
	4	ALTAVAEVNAIKALPQ	L	FPS	-	3.62
	5	NAILKALPQVKEIDLTAE	L	NN	CC	0.14
	6	qplaklianeatledksr	L	NNd	CC	0.02
	7	diekvgaGPGNAIKALPQ	L	MD	CC	0.02
	8	ARIKLVESPSRSDYIN	T	FPS	SC	1.83
TPO	9	VVTDDDRYSDLLMAWGQ		B		3.76
	10	ARLRDSGRAYLPFVPP		B		0.46
	11	PYEGYDSTANPTVSNVF		B		0.06
	12	EDFESCDSIPGMNLEA		B		0.39
	13	RRLDASFQEHDPDLPGWLWH	L	FPS	-	3.75
	14	ASFQEHDPDLPGRLWQFVLS	L	NN	CC	0.33
	15	APEPGIPGERPCPPRAPAA	T	NN	TC	0.47
	16	ASFQEHDPDLGRAPEPGIPGE	L	MD	CC	0.14
IL8	17	PKFIKELRVIESGPHCANT	L	FPS	-	0.16
	18	ENWVQRVVEKFLKRAENS	T	FPS	SC	2.25

¹ Amino acids in lower case are D-amino acids. ² Algorithm used to predict the sequence of the immunogenic peptide: B for BEPITOPE otherwise PEPOP. ³ The reference segment chosen is the longest (L: longest) or the segment at the top (T) of the list of all the segments ranked according to the five characteristics indicated in the text. ⁴ The method of extension chosen is either elongation with the flanking protein sequence (FPS); addition of the nearest neighbor (NN) segment(s) regardless of the sense of the segment added: if the segment is in the unnatural sense, i.e., from the C-terminus to the N-terminus, it is synthesized using D-amino acids (NNd); or manual design (MD). ⁵ The area of extension is the cluster top of the list (TC: top cluster) of the clusters ranked according to the five characteristics (see the text), or the cluster containing the selected reference segment (CC: containing cluster), or a user selected cluster (SC: selected cluster), or none when the reference segment is the longest and the method is the FPS. ⁶ Positive reactivities in the indirect ELISA are values over 0.2, and strong reactivities are values over 1.0. Control experiments done by using the preimmune serum gave values ranging from 0.0020 to 0.0028. The given values correspond to specific absorbances (A_{450 nm} immune serum – A_{450 nm} preimmune serum).

strong Ab response against the corresponding protein Ag. Except in a few cases, more sophisticated design methods (NN, NNd, and MD) that tried to reconstitute discontinuous epitopes failed to propose peptides which had useful immunogenic properties. As compared with peptides designed from the standard method (BEPITOPE), PEPOP performed equally well or even better.

C- Use of the PEPOP algorithm for sandwich immunoassay design

Since PEPOP is able to predict immunogenic peptides localized on the surface of the 3D structure of a target protein, it could conceivably be used to select two candidate peptides that are structurally appropriately separated in the 3D model such that they would *a priori* generate Abs able to react with the protein in a sandwich assay (i.e., an assay in which the protein Ag, in solution, is captured by two different Abs). As an example of such an application of PEPOP, Figure 4A shows the localization of peptide #1 and peptide #4 on the 3D structure of IA-2. Peptide #1 corresponds to an α -helical part of the PTP domain of IA-2, whereas peptide #4 maps to a hairpin in the juxta-membrane domain. These peptides clearly belong to spatially opposite regions on the protein. The resulting anti-peptide Abs were used in a sandwich ELISA to validate their

capacity to simultaneously bind the cognate Ag. The results (Figure 4B) demonstrate the possibility to use PEPOP to target specific regions of the protein so as to obtain a pair of Abs able to capture and quantify the protein of interest in solution.

Discussion

This work was conducted in the context of an abundance of epitope predictive methods based on protein sequence information and a relative scarcity of methods using 3D information. Since the majority of Abs against protein Ags binds to discontinuous sites, predictive methods should take into account structural information and try to guess the identity of discontinuous epitopes. This idea has recently been taken into account by a few groups of investigators [21-24].

PEPOP is a two-purpose algorithm. On the one hand, it allows the identification of clusters of accessible surface residues and segments that might correspond to antigenic epitopes, and on the other hand, it is able to design peptides that can be used to prepare Abs that target the cognate Ag at specific sites on its molecular surface. From the 3D structure of a protein, PEPOP identifies segments com-

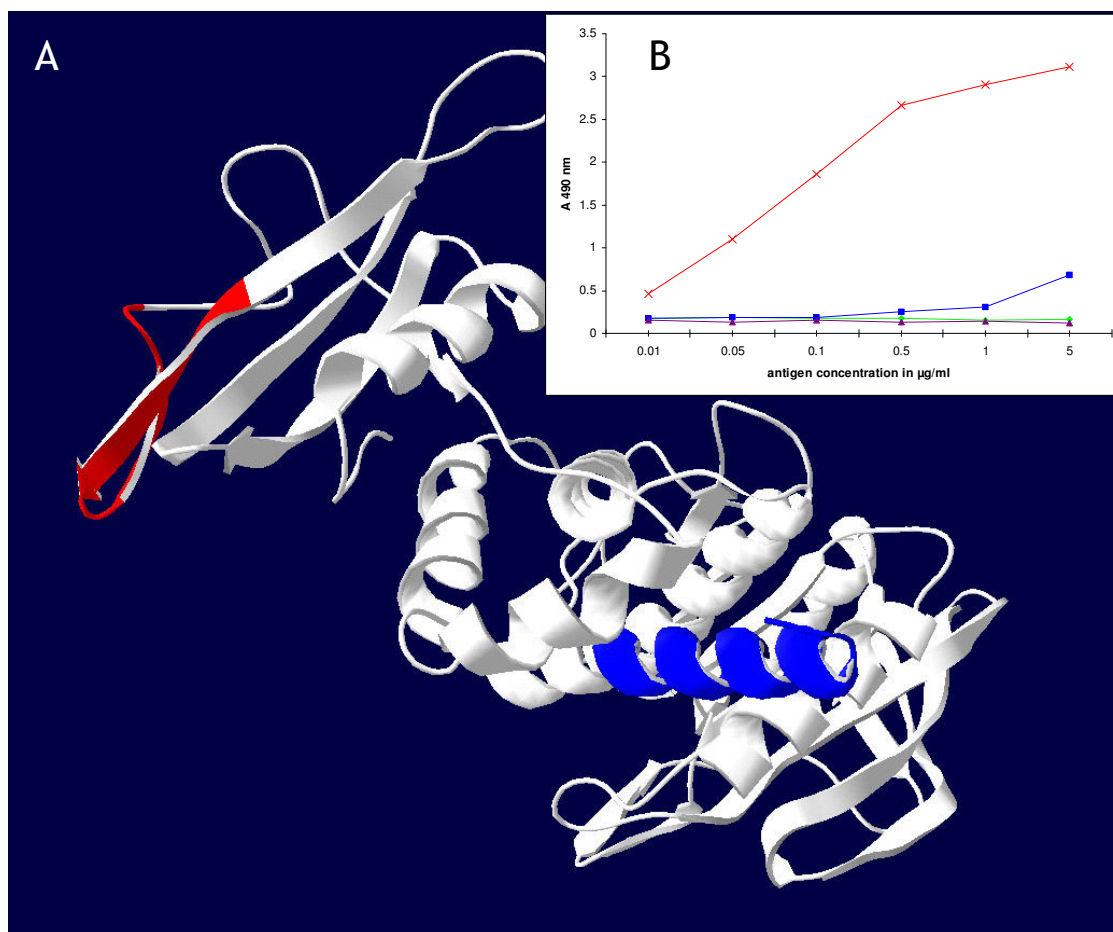


Figure 4

Sandwich immunoassay design for IA-2. (A) 3D view. Peptide #1 (in red) and peptide #4 (in blue) are represented on the 3D structure of IA-2. **(B)** Graphic representation. The graph indicates the reactivity of the polyclonal (Abs) with IA-2, in a sandwich-type ELISA format. The Ag, captured by the immobilized Ab, is detected by a biotinylated Ab, followed by addition of streptavidin-peroxidase. Reactivity of x, anti-peptide #4 with IA-2, detected by anti-peptide #1; ■, anti-peptide #4 with TPO, detected by anti-peptide #1; ▲, anti-peptide #15 with IA-2, detected by anti-peptide #1; ◆, anti-peptide #15 with TPO, detected by anti-peptide #1.

posed of one to n continuous surface accessible residues. These segments are clustered according to their spatial vicinity: thus, the clusters are proposed to be putative discontinuous epitopes. Segments and clusters are further used to design putatively immunogenic peptides. Antigenicity is the ability of a protein to bind specifically to the paratope of an Ab. Immunogenicity is the ability of an Ag to induce an immune response in a suitable host. Thus, antigenicity is just a molecular recognition phenomenon, whereas immunogenicity depends on multiple factors extrinsic to the sole Ag [30]. In developing and using PEPOP, we have clearly distinguished these two properties and we show that the outputs of PEPOP could be used in both settings. As concerns antigenicity, one has to keep in mind that epitopes are in essence unpredictable since

epitopes are not intrinsic features of proteins but exist only in the context of the Ab that binds to it [31]. Thus, what one generally predicts is at best a putative epitope to which an (undefined) Ab might bind. Our method, as all others in the same vein, is a tool that helps experimentalists to focus on a defined part of a protein for experimentally assessing its real antigenic character by measuring antibody binding to this particular part.

A key step in proposing a new predictive tool is the demonstration of its efficacy [32]. To evaluate the performance of an algorithm to predict antigenicity seems rather easy since it consists of comparing the predictions with known epitopes. The gold standard for comparison is an epitope, as defined by X-ray crystallography of an Ab-Ag complex

since this method provides a comprehensive topological view of the epitope. However, it certainly counts as epitope residues, some residues which are not energetically involved in the interaction with antigen (Novotny, 1989), and this may bias the comparison between predicted and epitopes defined by X-ray crystallography. The authors of the three reported predictive algorithms that make use of 3D information used different methods for evaluation of the algorithm performances. Batori *et al.* [22] reported results on a single Ab-Ag model, but they rigorously compared the performances of their software with that of sequence-based algorithms. As expected, their EMT algorithm performed better than conventional methods [22,33]. To assess the performance of their DiscoTope tool, Haste Andersen *et al.* [23] compared its predictions with those obtained by the method of Parker *et al.* [14] and by a surface accessibility calculation. They also reported epitope prediction for a single antigen. The evaluation strategy developed by Kulkarni-Kale *et al.* [24] was similar to the one we undertook. Their rationale for prediction was also very similar to ours. It seems, however, that their CEP algorithm tends to predict much larger antigenic determinants than PEPOP. Authors of the fourth tool, the MEPS server [21], did not present any validation data. The ability of PEPOP to identify putative antigenic regions on proteins was evaluated by comparing 13 structurally defined epitopes on a total of 8 different protein Ags with the clusters calculated by the algorithm. The identified clusters were shown to correspond to a large (but not complete) extent to structurally defined epitopes. In 11 out of the 13 Ab-Ag complexes analyzed, at least 50% of the structural epitope residues were found to be included in a single PEPOP cluster. In the last two Ab-Ag complexes [PDB: [1VFB](#), [1EY0](#)], the predicted residues almost equally distribute into three clusters, thus yielding the lowest sensitivity numbers. It should be noted that our method being based on detection of exposed residues, residues buried in the structure that may participate to epitopes are missed. CEP, EMT and DiscoTope also make use of the accessibility parameter in their calculations. The comparative assessment of sensitivity, specificity and positive predictive value of PEPOP, CEP, and DiscoTope showed that PEPOP achieves similar or slightly better performances than the other two algorithms.

As regards to immunogenicity, assessing the performance of a tool to predict immunogenic peptides is more difficult because it implies time consuming experiments, notably immunizing animals. As others [34,35], we have made this effort. To demonstrate the ability of PEPOP to successfully predict immunogenic peptides, several peptides were designed from the 3D structure of three model Ags using different methods; they were then chemically synthesized and used to elicit Abs in rabbits. Since all peptides induced a strong anti-peptide antibody response

(data not shown) and not all antibodies recognized the cognate protein, it was assumed that antibodies did not cross-react with a denatured antigen. The capacity of the Abs to cross-react with the target Ag was taken as proof of a successful prediction. The FPS method of peptide design was found to be efficient to yield peptides able to induce a strong Ab response against the cognate protein Ag since an 80% success rate was achieved. The FPS method provides a peptide with a sequence made from contiguous residues of the protein, thus the 3D information is lost. However, as compared with peptides designed by using BEPITOPE [18], the FPS method of PEPOP performed better. Both PEPOP and BEPITOPE use the β -turn criteria to predict epitope (assignment in PEPOP, propensity of the antigen sequence in BEPITOPE). PEPOP, however, uses four additional parameters, the length of the segment, its hydrophobicity, its accessibility as well as the occurrence of R, W, Y, and P, that have been found to be overrepresented in protein-protein interfaces [36]. It is possible that this aggregated information might add some value to the prediction simply based on β -turn propensity. Note, however, that the performances of BEPITOPE have recently been surpassed by machine learning approaches [37]. A disappointing observation was that our attempts to design peptides mimicking discontinuous regions of the Ags were not fully successful, in that the corresponding Abs were poorly or not at all cross-reactive with the cognate protein. Cross-reactive antibodies were obtained with several peptides (designed notably using the NN method), but their reactivity was low as compared with FPS-based antibodies. Chemically mimicking discontinuous protein surfaces have been a long-standing scientific challenge [38]. We are nevertheless confident that this could be amenable since many mimotopes have been identified that are considered as low molecular weight functional replicas of discontinuous epitopes [39]. As the PEPOP software has only recently been developed, many parameters can probably be improved to better predict discontinuous peptides. For example, the scoring and ranking of the segments can be adjusted and the combination of one to five of the characteristics deserves to be thoroughly tested. Although PEPOP has proved its capacity to predict relevant clusters and immunogenic peptides, it still contains a great potential for improvement, particularly if a prediction \rightarrow experimental validation \rightarrow optimization loop is implemented. PEPOP is flexible and can be a useful tool for different purposes. For example, in an antigenic use, PEPOP can help to map an epitope by building up a small virtual peptide library that can then be tested for binding to the Ab; should the Ab recognize a peptide, the epitope is localized. PEPOP has also the potential to predict mimotopes, i.e. peptides without sequence similarity with the Ag sequence which are recognized by the mAb. Mimotopes can have many applications and lead to pharmacological target identification (interesting for drug

design or guided docking), to protein engineering, vaccine design, identification of protein function, etc. As concerns immunogenicity, PEPOP can be used to target a specific region of a protein or to obtain Abs that capture the Ag. The need for mAbs to characterize proteins identified by large-scale proteomic studies is ever increasing. Since the protein itself is not always available, the interest in developing a method using peptides as surrogate Ags may have great potential. An interesting feature of the PEPOP algorithm is its possibility to propose putative immunogenic peptides that could yield Abs suitable for a sandwich capture assay of any protein Ag that can be modeled. With the advent of large-scale proteomic studies and Ab arrays, there is an increasing need for such immunoassays [40,41]. An ideal pipeline to fit these requirements would rely on novel high-throughput modeling capacities [42] and bioinformatics tools like PEPOP to select peptides so as to obtain in a straightforward way pairs of surface-targeted Abs for developing sandwich assays for diagnostic or discovery purposes.

Conclusion

PEPOP can identify epitopes at the surface of proteins with accuracy comparable to similar tools available through web interface. Moreover, PEPOP can also be used to design immunogenic peptides from the 3D coordinates of the protein.

Methods

PEPOP algorithm

Clustering of the accessible surface segments

From the 3D structure of a protein, the solvent accessible surface area is calculated (Figure 1A, step 1). The surface accessibility of amino acids is determined by using DSSP [43] with the default parameters. Segments composed of accessible and contiguous amino acids are then listed (a segment can be constituted by a single amino acid) (Figure 1A, step 2). Each segment is then approximated to a geometric segment represented by three points: the first two points are the $C\alpha$ of the N-terminal and C-terminal residues of the segment, the third point is calculated as the mid-point between the other two. Hence, a segment is represented by the 3D coordinates (X, Y, Z) of the three characteristic points. A distance matrix is then calculated in which the comparison between two segments produces nine values since a segment is represented by three points. A matrix of the shortest distances is then derived in which the distance between two segments is represented by the lowest value among the nine previously calculated. This matrix is used to cluster the segments (Figure 1A, step 3). The clustering of the segments is performed by Kitsch (from PHYLIP package v3.6) [44]. At this stage, a set of accessible surface segments is identified and segments clustered based on the shortest distance matrix.

Scoring the segments

A score is attributed to the segments of a given list (either the segments of a cluster or all the segments identified) for each selected property (segment length, segment accessibility, segment hydrophobicity, occurrence of particular amino acids, and occurrence of residues in β -turns). The length score is the number of amino acids in a segment. The accessibility score is the average accessibility of the amino acids composing the segment (values from DSSP). The hydrophobicity score is the number of hydrophobic amino acids (Y, W, F, L, V, I, C, P, M) in the segment. The particular amino acid score is the number of W, R, Y and P in the segment. The β -turn score is the number of amino acids involved in a β -turn (DSSP assignment).

Scoring the clusters

Each cluster is scored for the five properties according to the score of the composing segments. The length score of a cluster is the length of the longest segment of this cluster. The accessibility score is the number of segments of the cluster being part of the first quartile of the most accessible segments. The hydrophobicity score is the sum of the hydrophobic amino acids of the segments contained in the cluster. The particular amino acid score is the sum of W, R, Y, and P of the segments contained in the cluster. The β -turn score is the sum of amino acids implied in a β -turn of the segments contained in the cluster.

Ranking the segments or the clusters

The segments or the clusters are ranked for each of the five properties according to the assigned score. The five ranks of a segment or a cluster are summed, and the segments or clusters are finally ranked according to these sums.

Methods to design immunogenic peptides

To design a peptide, a "reference segment" is chosen from the set of identified segments and its sequence extended with residues selected by a "method of extension" in a specified "area of extension". The reference segment can be manually selected among all the segments if a particular region of the protein is desired to be targeted and so present in the final peptide. Otherwise, the PEPOP algorithm is devised to automatically select the first ranked segment according to criteria selected among the five physicochemical or structural properties of the amino acids. These parameters were chosen in agreement with different analyses of Ab-Ag and/or protein-protein interactions (among them [45,46]) that have shown that they play a role in or they favorably influence the antigenicity and even the immunogenicity of a protein, or they are favorably associated with the binding between proteins. In the PEPOP software, these five characteristics can be chosen one by one or combined. The segments are ranked according to the chosen characteristics within each segment: the top scored segment is automatically selected as

the reference segment. In a second step, the reference segment is elongated to yield a peptide of a suitable molecular size. Three methods of extension have been implemented. The simplest is the extension of both sides of the sequence of the reference segment with the flanking sequence of the cognate protein (called FPS for Flanking Protein Sequence). In the second method (called NN for nearest neighbours), the segments having the shortest distance from the reference segment are searched for by the algorithm. The sequence of the segment nearest to the C-terminus of the reference segment is added C-terminally (Figure 3). If necessary, a further extension is conducted in the same way, from the last segment added until a defined total minimal length of the peptide sequence is reached. The third method is the search for an optimized path between the segments composing a peptide (called ONN for optimized nearest neighbor path). From a set of segments that composes a peptide, all the possible combinations are explored to select the one for which the global distance between all the segments is the shortest. The last element for the design of the peptide is the choice of the area of extension. The elongation of the peptide can take into account either the whole set of accessible segments from the protein or only a previously determined cluster of accessible segments. This cluster can be chosen among all clusters by the user, or it is automatically selected by the algorithm in the same way as the reference segment was automatically selected. Thus, the clusters are ranked according to the same previously described five characteristics that can be combined: accessibility, length, hydrophobicity, number of R, W, Y, P, and number of β -turns. The cluster having the best rank according to the selected parameters is selected as the area of extension.

Implementation details

PEPOP has been implemented on a Linux server (Dell PE2250 virtualized server with the Mandriva 2007 OS distribution) running the Apache web server version 2.0. The algorithm has been implemented in object oriented PHP (version 5), which allows the simultaneous development of the web interface. The segments and clusters identified by PEPOP can be directly visualized on the 3D structure of the Ag thanks to the plug-in Chime [47]. Further development will allow the prediction and display of several peptides at the same time.

Comparison of the PEPOP performances with that of other available epitope prediction tools

The performances of PEPOP were compared with the two other available epitope tools, i. e. DiscoTope and CEP by calculating specificity, sensibility, and positive predictive value derived from a two-by-two contingency table [48,49]:

$$Sp = \frac{TN}{(TN+FP)} \quad Se = \frac{TP}{(TP+FN)} \quad PPV = \frac{TP}{(TP+FP)}$$

where Sp is the specificity, Se the sensibility, PPV the Positive Predictive Value, and TN the number of amino acids not predicted and actually not part of the epitope, TP the number of predicted amino acids that are actually part of the epitope, FP the number of predicted amino acids not part of the epitope, and FN the number of amino acids not predicted but part of the epitope. Therefore, the specificity evaluates the capacity of the tool to exclude those amino acids that are not part of the epitope, and the sensibility measures the capacity of the tool to identify the amino acids of the epitope. The positive predictive value is the proportion of predicted amino acids that are truly part of the epitope.

Note that the outputs of the three epitope prediction tools are different. DiscoTope yields a single prediction, identifying along the protein sequence the amino acids that might belong to an epitope. Instead, PEPOP and CEP suggest several potential epitopes. By grouping the segments, PEPOP identifies a few exclusive clusters, i.e., the potential epitopes are not overlapping. CEP proposed several CE (conformational epitope) that may or may not be partially overlapping. For each of the 13 epitope predictions, the Sp, Se, and PPV were calculated on the single proposition of DiscoTope, and on the cluster (for PEPOP) and the CE (for CEP) giving the best values.

3D structures of IA-2, TPO, and IL8

The intracytoplasmic part of insulinoma associated antigen 2 (IA-2) ([Swiss-Prot: [Q16849](#)] residues 601–979) is composed of two domains: the juxta-membrane domain (residues 601–690) and the PTP domain (residues 691–979). A theoretical model of the PTP domain of IA-2 was previously published by Dromey *et al.* [50] but was not made available in the PDB [51]. Consequently, a new theoretical model was calculated (Moreau, Valera *et al.*, in preparation) to determine the 3D coordinates of the structure. The human thyroid peroxidase (TPO) [Swiss-Prot: [P07202](#)] contains a large extracytoplasmic domain (residues 15–846) composed of four domains: a structurally undefined domain (residues 15–140), a MPO-like domain (141–740), a CCP-like domain (residues 741–795), and an EGF-like domain (residues 796–839). A theoretical model of the MPO-like domain has been reported [52] and was made available to us. The interleukin-8 (IL8) [Swiss-Prot: [P10145](#)] is composed of a single domain of about 70 amino acids, depending on the variant. Its 3D structure is available in PDB [PDB: [3IL8](#)].

Peptide predictions

The PEPOP predictions were made with the 3D structures of the PTP domain of IA-2, the MPO-like domain of TPO and the entire IL8. BEPITOPE was used with the default parameters on the Pellequer's TURN33 addition method. The predicted epitopes proposed were ranked according to the average value. The overlapping regions are considered as a unique predicted epitope. The sequences used were the intracytoplasmic part of IA-2, the extracytoplasmic part of TPO, and the entire sequence of IL8.

Peptide synthesis and purification

Peptides were prepared by Fmoc solid-phase synthesis using an AMS 422 robot or a MultiPep synthesizer (INTAVIS AG, Germany). Peptides were tagged by adding the tripeptide Cys-Tyr-Gly N- or C- terminally to the target sequence to facilitate monitoring at 280 nm during purification and to provide a thiol handle for coupling to a protein carrier. The standard synthesis protocol [53] was used throughout. Mass spectrometry (MALDI-Tof Voyager DE, Applied Biosystems) was used to confirm the identity of the synthetic peptide with the target sequence. Peptides were coupled to keyhole limpet hemocyanin (KLH) by using the heterobifunctional coupling reagent sulfo-SMCC, according to the manufacturer's instructions.

Immunizations

New Zealand white rabbits (Centre Lago, France) were immunized three times via the intradermic route at 14-day intervals using 50 µg of Ag (KLH-coupled peptides) and Freund's complete or incomplete adjuvant. Rabbits were bled 10 days after the third immunization and the serum titer measured by ELISA. The rabbits were boosted twice subcutaneously at 21-day intervals with 50 µg of Ag prior to terminal blood collection (by cardiac puncture).

Ab reactivity with the cognate protein

The entire IL8 (recombinant protein produced in *E. coli*) was purchased from Peprotech.

Extracytoplasmic TPO was purchased from HyTest Ltd. The intracytoplasmic part of IA-2 (residues 601–979) was produced in the baculovirus-insect cell system in our laboratory.

For indirect ELISA, maxisorp 96-well plates were coated with 2 µg/ml of Ag in phosphate-buffered saline (PBS) (overnight at 4°C). The plates were washed in PBS containing 0.1% Tween-20 (PBS-T) and blocked with 2% nonfat milk in PBS-T (1 h at 37°C). One hundred microlitres of serial dilutions of the final bleeds (1:5 000 to 1:1 000 000) in 2% nonfat milk PBS-T was added to each well (2 h at 37°C). After three washes in PBS-T, the plates were incubated with a peroxidase-conjugated anti-rabbit Ab (Sigma), diluted 1:3 000 in PBS-T 2% nonfat milk (1 h at

37°C). Plates were washed four times in PBS-T and incubated with OPD. After 20 min at room temperature, the absorbance at 450 nm was measured.

For sandwich ELISA, maxisorp 96-well plates were coated with 2 µg/ml of Protein A purified Ab (overnight at 4°C). After blocking with 3% bovine serum albumin (BSA) in PBS-T (1 h at 37°C), 100 µl of serial dilutions of the Ag (0.01 µg/ml to 5 µg/ml in 3% BSA-PBS-T) was added (2 h at 37°C) to each well. After three washes in PBS-T, the plates were incubated with 2 µg/ml of a biotinylated-conjugated purified polyclonal Ab in 3% BSA-PBS-T (2 h at 37°C). After three washes in PBS-T, the plates were incubated with peroxidase-conjugated streptavidin (Amersham), diluted 1:3 000 in 3% BSA-PBS-T (1 h at 37°C). Plates were washed four times in PBS-T and incubated with OPD substrate for 20 min at room temperature. The reaction was stopped by adding 50 µl of 4 N H₂SO₄ and the absorbance at 490 nm was measured. In control experiments, rabbit preimmune sera were tested in the above conditions. The absorbance value noticed for each dilution was subtracted from the absorbance value given by the corresponding dilution of the rabbit immune serum.

Availability and requirements

PEPOP is a server web based application and is usable as a Sysdiag Service at <http://diagtools.sysdiag.cnrs.fr/PEPOP/>. This service is platform independent, fully tested with Windows 2000™, Windows XP™. Programming language: PHP; required the installation of the plugin Chime (not compatible with all the navigators) to visualize the results in the 3D structure of the protein.

Authors' contributions

VM and FM developed the concept of the method, VM and CF wrote the computer code and implemented the Web interface. VM performed the predictions for the evaluation tests. For the antigenic evaluation, VM and CF collected the data; VM, CG, FM, and CF analyzed the results. For the immunogenic evaluation, CN synthesized the peptides; DP, SV and NN carried out the immunizations and tested the binding reactivities. CG, VM, DL, and FM analyzed the results. CG, VM, DL and FM participated in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Pr. B. J. Sutton for permission to use the three-dimensional model of TPO. We thank C. Frauer for participation in the development of the PEPOP algorithm and Dr. P. Mazière and D. Jean for computing assistance. We are grateful to Drs. C. Bareil, D. Bresson, and S. Peraldi-Roux for helpful discussions. The expert presubmission editorial assistance of Dr. S. L. Salhi is acknowledged.

References

1. Edwards AM, Arrowsmith CH, Christendat D, Dharamsi A, Friesen JD, Greenblatt JF, Vedadi M: **Protein production: feeding the crystallographers and NMR spectroscopists.** *Nat Struct Biol* 2000, **7(Suppl)**:970-972.
2. Carter P: **Site-directed mutagenesis.** *Biochem J* 1986, **237(1)**:1-7.
3. Szklarz GD, Halpert JR: **Use of homology modeling in conjunction with site-directed mutagenesis for analysis of structure-function relationships of mammalian cytochromes P450.** *Life Sci* 1997, **61(26)**:2507-2520.
4. Frank R: **The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports – principles and applications.** *J Immunol Methods* 2002, **267(1)**:13-26.
5. Reineke U, Kramer A, Schneider-Mergener J: **Antigen sequence- and library-based mapping of linear and discontinuous protein-protein-interaction sites by spot synthesis.** *Curr Top Microbiol Immunol* 1999, **243**:23-36.
6. Eichler J: **Synthetic peptide arrays and peptide combinatorial libraries for the exploration of protein-ligand interactions and the design of protein inhibitors.** *Comb Chem High Throughput Screen* 2005, **8(2)**:135-143.
7. Geysen HM, Barteling SJ, Meloen RH: **Small peptides induce antibodies with a sequence and structural requirement for binding antigen comparable to antibodies raised against the native protein.** *Proc Natl Acad Sci USA* 1985, **82(1)**:178-182.
8. Halperin I, Wolfson H, Nussinov R: **SiteLight: binding-site prediction using phage display libraries.** *Protein Sci* 2003, **12(7)**:1344-1359.
9. Moreau V, Granier C, Villard S, Laune D, Molina F: **Discontinuous epitope prediction based on mimotope analysis.** *Bioinformatics* 2006, **22(9)**:1088-1095.
10. Wilson IA, Niman HL, Houghten RA, Chersonson AR, Connolly ML, Lerner RA: **The structure of an antigenic determinant in a protein.** *Cell* 1984, **37(3)**:767-778.
11. Pellequer JL, Westhof E, Van Regenmortel MH: **Epitope prediction from primary structure of proteins.** In *Peptide Antigens: A Practical Approach* Edited by: Wisdow GB. Oxford: IRL Press; 1994:7-25.
12. Kolaskar AS, Tongaonkar PC: **A semi-empirical method for prediction of antigenic determinants on protein antigens.** *FEBS Lett* 1990, **276(1-2)**:172-174.
13. Alix AJ: **Predictive estimation of protein linear epitopes by using the program PEOPLE.** *Vaccine* 1999, **18(3-4)**:311-314.
14. Parker JM, Guo D, Hodges RS: **New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites.** *Biochemistry* 1986, **25(19)**:5425-5432.
15. Jameson BA, Wolf H: **The antigenic index: a novel algorithm for predicting antigenic determinants.** *Comput Appl Biosci* 1988, **4(1)**:181-186.
16. Welling GV, Weijer WJ, van der Zee R, Welling-Wester S: **Prediction of sequential antigenic regions in proteins.** *FEBS Lett* 1985, **188(2)**:215-218.
17. Van Regenmortel MH, Pellequer JL: **Predicting antigenic determinants in proteins: looking for unidimensional solutions to a three-dimensional problem?** *Pept Res* 1994, **7(4)**:224-228.
18. Odorico M, Pellequer JL: **BEPITOPE: predicting the location of continuous epitopes and patterns in proteins.** *J Mol Recognit* 2003, **16(1)**:20-22.
19. Blythe MJ, Flower DR: **Benchmarking B cell epitope prediction: underperformance of existing methods.** *Protein Sci* 2005, **14(1)**:246-248.
20. Moulton J: **A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction.** *Curr Opin Struct Biol* 2005, **15(3)**:285-289.
21. Castrignano T, De Meo PD, Carrabino D, Orsini M, Floris M, Tramontano A: **The MEPS server for identifying protein conformational epitopes.** *BMC Bioinformatics* 2007, **8(Suppl 1)**:S6.
22. Batori V, Friis EP, Nielsen H, Roggen EL: **An in silico method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context.** *J Mol Recognit* 2006, **19(1)**:21-29.
23. Haste Andersen P, Nielsen M, Lund O: **Prediction of residues in discontinuous B-cell epitopes using protein 3D structures.** *Protein Sci* 2006, **15(11)**:2558-2567.
24. Kulkarni-Kale U, Bhosle S, Kolaskar AS: **CEP: a conformational epitope prediction server.** *Nucleic Acids Res* 2005:W168-171.
25. Novotny J, Handschumacher M, Haber E, Brucoleri RE, Carlson WB, Fanning DW, Smith JA, Rose GD: **Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains).** *Proc Natl Acad Sci USA* 1986, **83(2)**:226-230.
26. Thornton JM, Edwards MS, Taylor WR, Barlow DJ: **Location of 'continuous' antigenic determinants in the protruding regions of proteins.** *Embo J* 1986, **5(2)**:409-413.
27. **The Protein DataBase** [<http://www.rcsb.org/pdb/home/home.do>]
28. **DiscoTope** [<http://www.cbs.dtu.dk/services/DiscoTope/>]
29. **CEP** [<http://202.41.70.74:8080/cgi-bin/cep.pl>]
30. Van Regenmortel MH: **Antigenicity and immunogenicity of synthetic peptides.** *Biologicals* 2001, **29(3-4)**:209-213.
31. Van Regenmortel MH: **Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism.** *EMBO Rep* 2004, **5(11)**:1016-1020.
32. Greenbaum JA, Haste Andersen P, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar AS, Lund O, Morrison S, et al.: **Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools.** *J Mol Recognit* 2007, **20(2)**:75-82.
33. Roggen EL: **Recent developments with B-cell epitope identification for predictive studies.** *Journal of Immunotoxicology* 2006, **3**:1-13.
34. Di Giambattista M, Branckaert T, Hougard V, Kembell-Cook G, Laub R: **In silico prediction of FVIII epitopes recognised by natural autoantibodies in polyvalent immunoglobulin concentrates.** *Mol Immunol* 2007, **44(8)**:1903-1913.
35. Renukaradhya GJ, Mitra-Kaushik S, Sinnathamby G, Rajasekhar M, Shaila MS: **Mapping of B-cell epitopes of hemagglutinin protein of rinderpest virus.** *Virology* 2002, **298(2)**:214-223.
36. Young L, Jernigan RL, Covell DG: **A role for surface hydrophobicity in protein-protein recognition.** *Protein Sci* 1994, **3(5)**:717-729.
37. Sollner J, Mayer B: **Machine learning approaches for prediction of linear B-cell epitopes on proteins.** *J Mol Recognit* 2006, **19(3)**:200-208.
38. Reineke U, Sabat R, Misselwitz R, Welfle H, Volk HD, Schneider-Mergener J: **A synthetic mimic of a discontinuous binding site on interleukin-10.** *Nat Biotechnol* 1999, **17(3)**:271-275.
39. Meloen RH, Puijk WC, Slootstra JW: **Mimotopes: realization of an unlikely concept.** *J Mol Recognit* 2000, **13(6)**:352-359.
40. Kusnezow W, Hoheisel JD: **Antibody microarrays: promises and problems.** *Biotechniques* 2002:14-23.
41. Michaud GA, Salcius M, Zhou F, Bangham R, Bonin J, Guo H, Snyder M, Predki PF, Schweitzer BI: **Analyzing antibody specificity with whole proteome microarrays.** *Nat Biotechnol* 2003, **21(12)**:1509-1512.
42. Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A, Marti-Renom M, Karchin R, Webb BM, Eramian D, et al.: **MODBASE: a database of annotated comparative protein structure models and associated resources.** *Nucleic Acids Res* 2006:D291-295.
43. Kabsch W, Sander C: **Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.** *Biopolymers* 1983, **22(12)**:2577-2637.
44. Felsenstein J: **PHYLIP – Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5**:164-166.
45. Bogan AA, Thorn KS: **Anatomy of hot spots in protein interfaces.** *J Mol Biol* 1998, **280(1)**:1-9.
46. Chakrabarti P, Janin J: **Dissecting protein-protein recognition sites.** *Proteins* 2002, **47(3)**:334-343.
47. **Chime** [<http://www.mdl.com/>]
48. Pretty IA, Maupome G: **A closer look at diagnosis in clinical dental practice: part 2. Using predictive values and receiver operating characteristics in assessing diagnostic accuracy.** *J Can Dent Assoc* 2004, **70(5)**:313-316.
49. Swets JA: **Measuring the accuracy of diagnostic systems.** *Science* 1988, **240(4857)**:1285-1293.
50. Dromey JA, Weenink SM, Peters GH, Endl J, Tighe PJ, Todd I, Christie MR: **Mapping of epitopes for autoantibodies to the type I diabetes autoantigen IA-2 by peptide phage display and molec-**

- ular modeling: overlap of antibody and T cell determinants. *J Immunol* 2004, **172(7)**:4084-4090.
51. Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, Abola EE: **Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules.** *Acta Crystallogr D Biol Crystallogr* 1998, **54(Pt 6 Pt 1)**:1078-1084.
 52. Hobby P, Gardas A, Radomski R, McGregor AM, Banga JP, Sutton BJ: **Identification of an immunodominant region recognized by human autoantibodies in a three-dimensional model of thyroid peroxidase.** *Endocrinology* 2000, **141(6)**:2018-2026.
 53. Laune D, Molina F, Ferrieres G, Villard S, Bes C, Rieunier F, Charde T, Granier C: **Application of the Spot method to the identification of peptides and amino acids from the antibody paratope that contribute to antigen binding.** *J Immunol Methods* 2002, **267(1)**:53-70.
 54. Amit AG, Mariuzza RA, Phillips SE, Poljak RJ: **Three-dimensional structure of an antigen-antibody complex at 6 Å resolution.** *Nature* 1985, **313(5998)**:156-158.
 55. Chitarra V, Alzari PM, Bentley GA, Bhat TN, Eisele JL, Houdusse A, Lescar J, Souchon H, Poljak RJ: **Three-dimensional structure of a heteroclitic antigen-antibody cross-reaction complex.** *Proc Natl Acad Sci USA* 1993, **90(16)**:7711-7715.
 56. Braden BC, Souchon H, Eisele JL, Bentley GA, Bhat TN, Navaza J, Poljak RJ: **Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1.** *J Mol Biol* 1994, **243(4)**:767-781.
 57. Sheriff S, Silverton EVW, Padlan EA, Cohen GH, Smith-Gill SJ, Finzel BC, Davies DR: **Three-dimensional structure of an antibody-antigen complex.** *Proc Natl Acad Sci USA* 1987, **84(22)**:8075-8079.
 58. Kondo H, Shiroishi M, Matsushima M, Tsumoto K, Kumagai I: **Crystal structure of anti-Hen egg white lysozyme antibody (HyHEL-10) Fv-antigen complex. Local structural changes in the protein antigen and water-mediated interactions of Fv-antigen and light chain-heavy chain interfaces.** *J Biol Chem* 1999, **274(39)**:27623-27631.
 59. Li Y, Li H, Smith-Gill SJ, Mariuzza RA: **Three-dimensional structures of the free and antigen-bound Fab from monoclonal antilysozyme antibody HyHEL-63.** *Biochemistry* 2000, **39(21)**:6296-6309.
 60. Prasad L, Waygood EB, Lee JS, Delbaere LT: **The 2.5 Å resolution structure of the jcl42 Fab fragment/HPr complex.** *J Mol Biol* 1998, **280(5)**:829-845.
 61. Huang M, Syed R, Stura EA, Stone MJ, Stefanko RS, Ruf W, Edgington TS, Wilson IA: **The mechanism of an inhibitory antibody on TF-initiated blood coagulation revealed by the crystal structures of human tissue factor, Fab 5G9 and TF.G9 complex.** *J Mol Biol* 1998, **275(5)**:873-894.
 62. Fleury D, Daniels RS, Skehel JJ, Knossow M, Bizebard T: **Structural evidence for recognition of a single epitope by two distinct antibodies.** *Proteins* 2000, **40(4)**:572-578.
 63. Spiegel J, P C, Jacquemin M, Saint-Remy JM, Stoddard BL, Pratt KP: **Structure of a factor VIII C2 domain-immunoglobulin G4kappa Fab complex: identification of an inhibitory antibody epitope on the surface of factor VIII.** *Blood* 2001, **98(1)**:13-19.
 64. Tulip WR, Varghese JN, Webster RG, Laver WG, Colman PM: **Crystal structures of two mutant neuraminidase-antibody complexes with amino acid substitutions in the interface.** *J Mol Biol* 1992, **227(1)**:149-159.
 65. Lescar J, Pellegrini M, Souchon H, Tello D, Poljak RJ, Peterson N, Greene M, Alzari PM: **Crystal structure of a cross-reaction complex between Fab F9.13.7 and guinea fowl lysozyme.** *J Biol Chem* 1995, **270(30)**:18067-18076.
 66. Bossart-Whitaker P, Chang CY, Novotny J, Benjamin DC, Sheriff S: **The crystal structure of the antibody N10-staphylococcal nuclease complex at 2.9 Å resolution.** *J Mol Biol* 1995, **253(4)**:559-575.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Structure and Activity Analysis of Two Spider Toxins That Alter Sodium Channel Inactivation Kinetics[†]

Alessandra Matavel,^{‡,§} Cécile Fleury,^{‡,§,||} Leida C. Oliveira,[‡] Franck Molina,^{||} Maria Elena de Lima,[‡] Jader S. Cruz,[‡] Marta N. Cordeiro,[⊥] Michael Richardson,[⊥] Carlos H. I. Ramos,[#] and Paulo S. L. Beirão^{*‡}

Department of Biochemistry and Immunology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, CEP 31270-901 Belo Horizonte, MG, Brazil, Centro de Pesquisa e Desenvolvimento Carlos Ribeiro Diniz, Fundação Ezequiel Dias, Belo Horizonte, Brazil, Sysdiag, CNRS FRE 3009, 1682 rue de la Valsière, CS 61003, 34184 Montpellier CEDEX 4, France, and Department of Organic Chemistry, Instituto de Química, Universidade Estadual de Campinas, Campinas, Brazil

Received November 21, 2008; Revised Manuscript Received February 17, 2009

ABSTRACT: In this work, *Phoneutria nigriventer* toxins PnTx2–5 and PnTx2–6 were shown to markedly delay the fast inactivation kinetics of neuronal-type sodium channels. Furthermore, our data show that they have significant differences in their interaction with the channel. PnTx2–6 has an affinity 6 times higher than that of PnTx2–5, and its effects are not reversible within 10–15 min of washing. PnTx2–6 partially (59%) competes with the scorpion α -toxin AaHII, but not with the scorpion β -toxin CssIV, thus suggesting a mode of action similar to that of site 3 toxins. However, PnTx2–6 is not removed by strong depolarizing pulses, as in the known site 3 toxins. We have also established the correct PnTx2–5 amino acid sequence and confirmed the sequence of PnTx2–6, in both cases establishing that the cysteines are in their oxidized form. A structural model of each toxin is proposed. They show structures with poor α -helix content. The model is supported by experimental and theoretical tests. A likely binding region on PnTx2–5 and PnTx2–6 is proposed on the basis of their different affinities and sequence differences.

Peptide toxins were selected over millions of years to act on vital processes to kill or paralyze prey or predators. This natural selection provided an assortment of molecules capable of acting at low concentrations on ion channels, either inhibiting their conduction pathway or altering their kinetics (1). The latter type of toxins is proving to be a useful tool for investigating the molecular mechanisms of ion channel gating and has provided valuable information about this poorly understood process (see ref 2 as an example). At least six toxin binding sites have been characterized in the mammalian sodium channel, most of them affecting its kinetics. Site 3 is known to be occupied by peptide toxins that originated from scorpion venom (scorpion α -toxins) and from sea anemone, whereas site 4 is occupied by scorpion β -toxins (3). More recently, some spider toxins have also been shown to alter the kinetics of sodium channels. In contrast with the scorpion and anemone toxins, little is known about their molecular mechanism and binding sites (4).

In this paper, we have studied two homologous toxins from the Brazilian spider *Phoneutria nigriventer*, whose effects can account for the major symptoms of envenomation caused

by that spider, which comprise hyperexcitation, salivation, lachrymation, and priapism. At the cellular level, these toxins have been shown to have complex effects on sodium channel kinetics, inhibiting its inactivation and shifting the activation voltage dependence toward negative potentials (4, 5), effects that correspond to the action of scorpion α - and β -toxins, respectively. Those two toxins, named *P. nigriventer* toxins Tx2–5 (PnTx2–5)¹ and Tx2–6 (PnTx2–6), are similar in primary sequence, and their in vivo activities also resemble those produced by scorpion toxins, except for the conspicuous priapism. However, more detailed studies at the cellular level were still required to characterize their mechanisms of action. In this work, we present a detailed and ample set of results that indicates that these toxins act on neuronal sodium channels by binding to the overlapping site with scorpion α -toxins and by a similar mechanism.

MATERIALS AND METHODS

Electrophysiology. The GH3 (ATCC, Manassas, VA) cell line was used for the electrophysiology assays. The cells were kept in Dubelco's modified Eagle's medium (DMEM, Gibco)

[†] This work was supported by the Brazilian Research Council (CNPq), FAPEMIG, and CAPES/COFECUB exchange program. This work is part of the INCTTOX/CNPq.

* To whom correspondence should be addressed. Phone: +55 31 3409-2663. Fax: +55 31 3409-2614. E-mail: pslb@ufmg.br.

[‡] Universidade Federal de Minas Gerais.

[§] These authors contributed equally to this work.

^{||} CNRS FRE 3009.

[⊥] Fundação Ezequiel Dias.

[#] Universidade Estadual de Campinas.

¹ Abbreviations: AaHII, *Androctonus australis* Hector toxin II; BSA, bovine serum albumin; CD, circular dichroism; Css IV, *Centruroides suffusus suffusus* toxin IV; EGTA, [ethylene bis(oxymethylenetriethyl)ammonium]tetraacetic acid; FPLC, fast protein liquid chromatography; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HPLC, high-performance liquid chromatography; ICK motif, inhibitory cystine knot motif; PnTx2–5, *P. nigriventer* toxin Tx2–5; PnTx2–6, *P. nigriventer* toxin Tx2–6; SEM, standard error of the mean; TEA, tetraethylammonium; UV, ultraviolet.

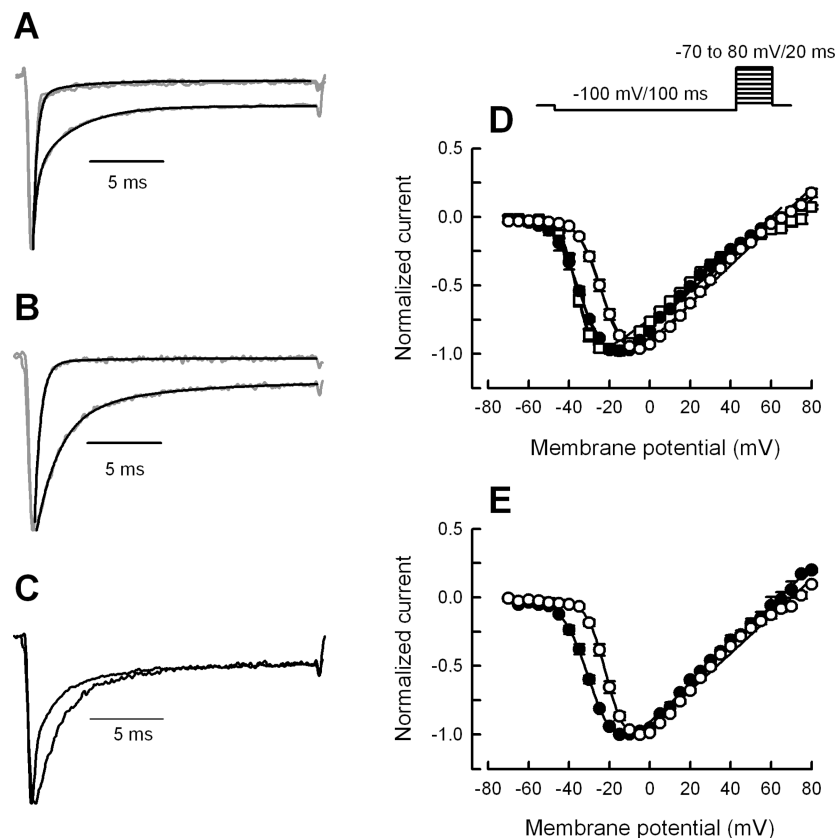


FIGURE 1: Effect of PnTx2-5 and PnTx2-6. (A) Superimposed records (gray lines) obtained from the same cell at 0 mV in the absence or presence of 400 nM PnTx2-5 and 5 min after washing. The latter record is almost identical with the control. The black line shows the curve obtained by fitting the inactivation kinetics, as described in Materials and Methods. (B) Same as panel A, with and without 100 nM PnTx2-6. (C) The records in the presence of PnTx2-5 and PnTx2-6 are superimposed to stress the difference in their effects. (D) Normalized $I-V$ curves in the absence (○) or presence of 400 nM PnTx2-5 (●) and 5 min after washout (□). (E) Normalized $I-V$ curves in the absence (○) and presence of 100 nM PnTx2-6 (●). The pulse protocol of the $I-V$ curves is shown above panel D. The $I-V$ curve was fitted with the equation $I_{Na} = g_{Na(Max)} \times (V_M - V_R) / \{1 + \exp[(V_g - V_M)/K_g]\}$, where I_{Na} is the current at each point, $g_{Na(Max)}$ is the fitted maximal conductance, V_M is the membrane potential, V_R is the reversal potential, V_g is the voltage which activates half of the maximal conductance, and K_g is the slope factor.

supplemented with 10% fetal bovine serum (Cultilab), 10 units/mL penicillin, and 10 μ g/mL streptomycin. The cells were plated 2–5 days before the experiment, and 24–48 h before the experiment, the concentrations of KCl and CaCl₂ in the culture medium were increased to 10 and 9.3 mM, respectively, to increase the level of sodium channel expression (6). Macroscopic Na⁺ currents were recorded at room temperature (22–25 °C) in the whole-cell patch clamp configuration (7) using an EPC-9 amplifier (Heka Instruments). Currents were low-pass filtered (Bessel) at a cutoff frequency of 10 KHz, acquired at 20–25 KHz, and digitally stored in a Power-Mac computer. The P/4 protocol was used for linear leak and capacitance subtraction. Patch pipettes were made with soft glass capillaries using a vertical pipet puller (Narishige) and had resistances between 2 and 4 M Ω . The pipettes were filled with internal solution 1 [10 mM NaCl, 20 mM TEA-Cl, 10 mM EGTA, 90 mM CsF, 20 mM CsOH, and 10 mM HEPES (pH 7.2) with CsOH] for experiments on PnTx2-5 or internal solution 2 [10 mM NaCl, 35 mM TEA-Cl, 5 mM EGTA, 100 mM CsF, and 10 mM HEPES (pH 7.2) with CsOH] for experiments with PnTx2-6. TEA-Cl and cesium were used to suppress K⁺ currents. The external solution contained 140 mM NaCl, 5 mM CsCl, 2 mM MgCl₂, 0.1 mM CdCl₂, 1 mM CaCl₂, 5 mM glucose, and 10 mM HEPES (pH 7.4) with NaOH. Cadmium was used to block Ca²⁺ channels. Throughout the

experiment, the cell under investigation was externally perfused via a 0.1 mm diameter pipet that was placed visually next to it. The perfusion was performed with the external solution containing 0.1% bovine serum albumin (BSA), to avoid adsorption of the toxin. BSA had no effect on the Na⁺ current. When appropriate, the perfusing solution was switched to an equal solution containing the toxin at the desired concentration. The effectiveness of the perfusion was checked beforehand by using a Na-free external solution. Under that condition, the inward current completely disappeared within 2 s. All reagents were of analytical grade.

Data Analysis. For the kinetic analysis of Na⁺ current inactivation, the decay of the current at a membrane potential of 0 mV was fitted with an exponential function with one or two time constants:

$$f(x) = a \exp(-T/\tau_f) + c \quad (1)$$

$$f(x) = a \exp(-T/\tau_f) + b \exp(-T/\tau_s) + c \quad (2)$$

where T is the time from the peak of the current and a and b are the proportions of the current that decays with time constants τ_f and τ_s , respectively.

Toxin Purification and Quantification. The PnTx2 fraction was purified by gel filtration chromatography in gel and reversed-phase chromatography in FPLC (8) from a soluble

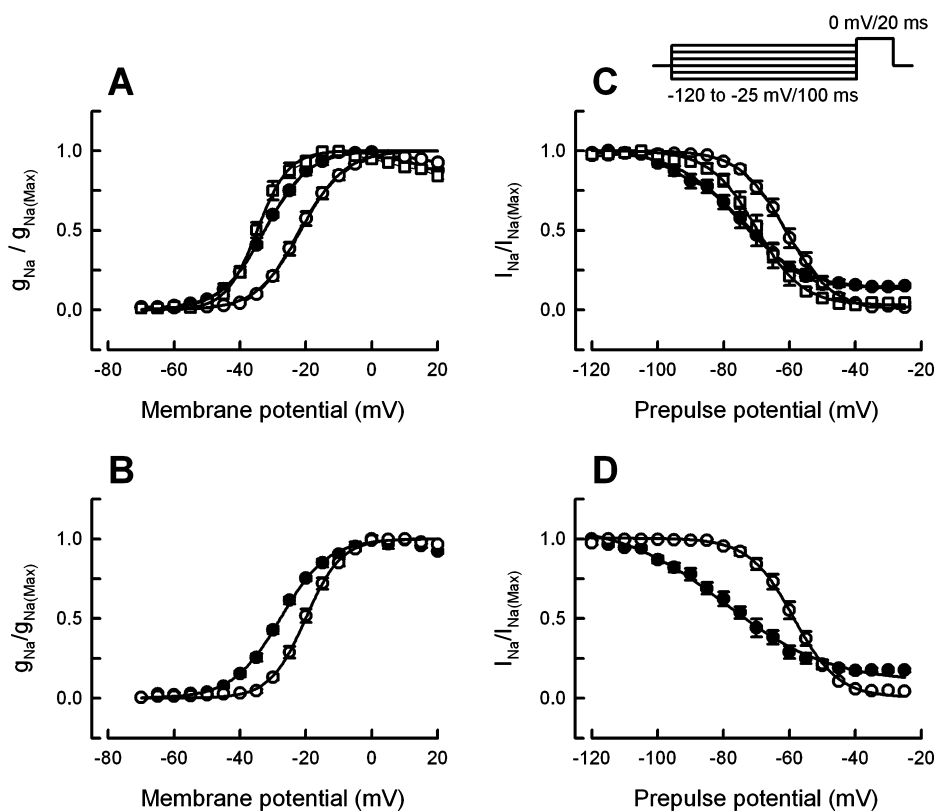


FIGURE 2: Effect on the voltage dependence of the peak conductance of 100 nM PnTx2-6 (A) and 400 nM PnTx2-5 (B). The conductance at each point was calculated using the equation $g_{\text{Na}} = I_{\text{Na}}/(V_{\text{M}} - V_{\text{R}})$ and fitted with a Boltzmann equation: $g_{\text{Na}}/g_{\text{Na}(\text{Max})} = \{1 + \exp[(V_{\text{g}} - V_{\text{M}})/K_{\text{g}}]\}^{-1}$. The right panel shows the voltage dependence of steady-state inactivation with 100 nM PnTx2-6 (C) and 400 nM PnTx2-5 (D). The pulse protocol used to measure the voltage dependence of steady-state inactivation is shown. The steady-state inactivation data were fitted with the equation $h_{\infty} = a\{1 + \exp[(V - V_{\text{h}})/K_{\text{h}}]\}^{-1} + b$, where V is the prepulse membrane potential, V_{h} is the voltage that inactivates half of the maximal current, and K_{h} is the inactivation slope factor. Parameters a and b reflect the proportion of inactivating and noninactivating channels, respectively: control (○), in the presence of the toxin (●), and 5 min after washout (□).

fraction of the total venom of the *P. nigriventer* spider. Isoforms PnTx2-5 and PnTx2-6 were separated by reversed-phase chromatography via HPLC (9). The purity of the toxins was verified by mass spectrometry (Q-ToF, Micromass, Manchester, U.K.). To quantify the toxin concentration, we calculated the molar extinction coefficient at 280 nm of each toxin in its denatured form, on the basis of the respective amino acid content. The ϵ was calculated with the equation $\epsilon_{280} = (\text{number of W}) \times 5690 + (\text{number of Y}) \times 1280 + (\text{number of C-C}) \times 120$ (10). Since it has been shown in native proteins that calculated extinction coefficients were accurate within an average of 5% error (11), we used this value throughout this work.

Tryptophan Fluorescence and Circular Dichroism (CD). Fluorescence records of PnTx2-6 were obtained with an ABL2 Spectronic fluorometer (Applied Biosystems, Foster City, CA) in quartz cuvettes with a path length of 1 cm. The sample was excited at 280 nm and the emission spectrum collected at 300–450 nm at 20 °C. The CD spectrum of PnTx2-6 was obtained with a spectropolarimeter (Jasco J-810) in a quartz cuvette with a path length of 1 cm with UV light from 190 to 260 nm at 23 °C. The toxin was diluted in water. The data were obtained at 0.5 nm intervals at 50 nm/min. All CD spectra are averages of 20 scans. The baseline under each condition was subtracted from the spectrum with the toxin.

Binding Assay. To identify the PnTx2-6 binding site on sodium channels, we performed competition assays using typical scorpion toxins that bind to site 3 (AaHIII from

Androctonus australis Hector) and site 4 (CssIV from *Centruroides suffusus suffusus*). These toxins (5 nmol) were radiolabeled by the lactoperoxidase method with 0.5 mCi of ^{125}I (Amersham Pharmacia Biotech) as described by Rochat and collaborators (12). To perform the experiments, we used either 5 μg of synaptosomes from rat brain (13) or 50 μg of total synaptosomal preparation (P2) (14). The protein concentration was determined by the method of Lowry (15). The tissue was incubated (30–40 min at 37 °C) in a medium containing 10^{-10} M radiolabeled toxin (^{125}I]AaHIII or ^{125}I]CssIV), alone or with a competing toxin. This could be either the same toxin unlabeled at 10^{-7} M, to assess the nonspecific binding, or PnTx2-6 at 10^{-10} – 10^{-6} M, to evaluate if it can compete with the radiolabeled toxin for the binding site. To remove the free ^{125}I -labeled toxin, the tissue was vacuum-filtered (synaptosomes) (12) or centrifuged at 11000g for 5 min (P2) (14). The radioactivity bound to the tissue was measured in a γ radiation counter (1275 mini Gamma Counter, LKB, Wallac).

Molecular Modeling of PnTx2-5 and PnTx2-6 Structures. A search for sequence similarity using NCBI PSI-BLAST (16) was performed, but no major similarity was found with any protein that had its structure experimentally determined. Fold recognition of both toxins, using mGenThreader (17), did not return any significant hit. The cysteine connectivity pattern of PnTx2-5 and PnTx2-6 was deduced by sequence alignment with all short spider toxins that contain at least eight cysteine residues and whose disulfide bridges were experimentally identified (Figure 7A). Accord-

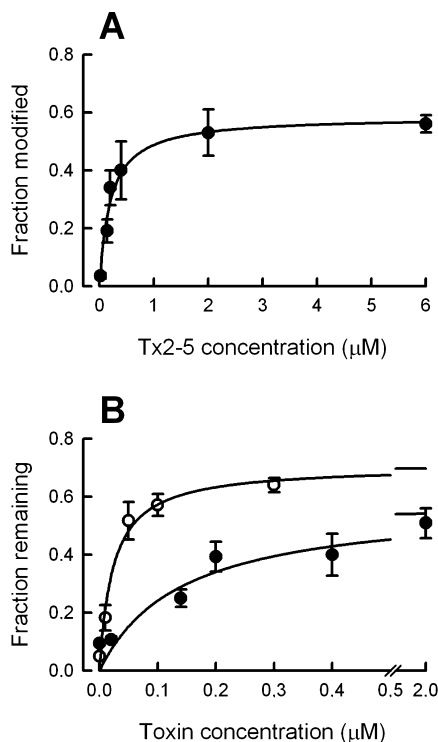


FIGURE 3: Concentration dependence of PnTx2-5 (A) and PnTx2-6 (B). In panel A, the effect was measured by the proportion of the inactivation-modified current calculated by eq 2 (Materials and Methods). The curve is the best fit of the data, with the equation $f(c) = P \times c/(K_{0.5} + c)$, where P is the maximal proportion of modified channels and $K_{0.5}$ is the concentration that modifies half of the channels, having a P of 0.59 and a $K_{0.5}$ of 190 nM. The effect of PnTx2-6 was measured by the current that remained at a time equal to 3-fold the fast time constant of inactivation (○). For the sake of comparison, the same procedure was carried out with the data of PnTx2-5 (●). The curves were obtained by fitting the data with the same equation, with P values of 0.71 and 0.58 and $K_{0.5}$ values of 23 and 140 nM for PnTx2-6 and PnTx2-5, respectively.

ing to SCOP (18), CATH (19), and DBAli (20) databases, they all present the same overall fold, including the so-called inhibitory cystine knot (ICK) structural motif (21, 22), with a relatively high degree of conservation of cysteine residue distribution among the group, in spite of the global high degree of sequence variability (23). Secondary structure predictions were performed on both target protein sequences using algorithms PHD and PROF (24), PSIPRED (17), Jpred (25), and nnPredict (26). Among the group of spider toxins of known structure, template structures were selected according to the following criteria: (i) sequence similarity (using BLAST local alignment algorithm), (ii) secondary structure similarity [by alignment of predicted secondary structures of all toxins, using the SOV algorithm (27)], (iii) cysteine pattern similarity [according to spider neurotoxin classification proposed by Kozlov and Grishin (28)], and (iv) known target ligand similarity (sodium channel gating modifier toxins). The selected template structures are those from δ -atracotoxin-Hv1 (29), δ -palutoxin IT2 (30), toxin AcTx-Hi:OB4219 (31), and μ -agatoxin-1 (32). Their atom coordinates are deposited in the Brookhaven Protein Data Bank (PDB) (33) as entries 1VTX, 1V91, 1EIT, and 1KQH, respectively. Three-dimensional structures of PnTx2-5 and PnTx2-6 were calculated by a homology-based molecular modeling method using Modeler 9v3 (34). Using the Modeler

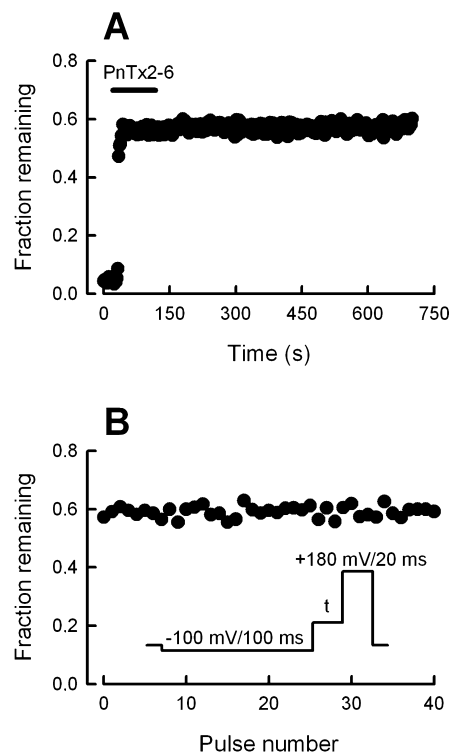


FIGURE 4: Lack of displacement of PnTx2-6 upon washing (A) and upon application of strong (180 mV) depolarizing pulses (B), as shown in the inset, where t is the test pulse to 0 mV for 20 ms. The effect was measured with the same procedure described for Figure 3B.

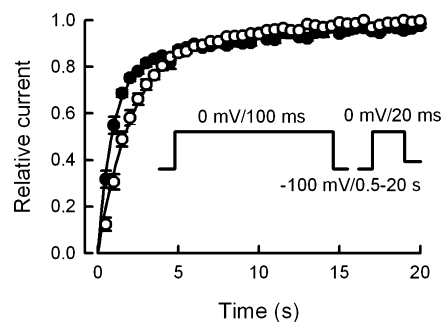


FIGURE 5: Recovery from inactivation in the presence of PnTx2-6. The pulse protocol is shown in the inset. The data were fitted with the sum of two-exponential functions: $f(T) = 1 - [k_1 \exp(-T/\tau_1) + k_2 \exp(-T/\tau_2)]$, where T is the recovery time and k_1 and k_2 are the proportions of the current that recover with time constants τ_1 and τ_2 , respectively. The best fits were obtained with the following values: $k_1 = 0.84 \pm 0.04$ and 0.79 ± 0.02 , $k_2 = 0.16 \pm 0.04$ and 0.21 ± 0.02 , $\tau_1 = 1.99 \pm 0.24$ and 0.91 ± 0.07 s, and $\tau_2 = 4.4 \pm 1.1$ and 10.7 ± 0.8 s in the absence (○) and presence (●) of PnTx2-6, respectively. τ_1 is significantly smaller in the presence of PnTx2-6.

align command, a structural alignment of the four template structures was generated, and then target sequences of each toxin were aligned together with the resulting structure-based template sequence alignment, using the Modeler *align2d* command. As predicted from target template sequence alignment, the pairs of cysteine residues (Cys3-Cys17, Cys10-Cys23, Cys14-Cys46, Cys16-Cys31, and Cys25-Cys29) from both toxin sequences were virtually constrained to form disulfide bridges. For each toxin, 100 models were calculated and evaluated using Modeler Objective Function (34), WHAT IF Check Report results (35), Procheck Ramachandran plot analysis (36), Verify3D score (37), and ProSA z-score (38).

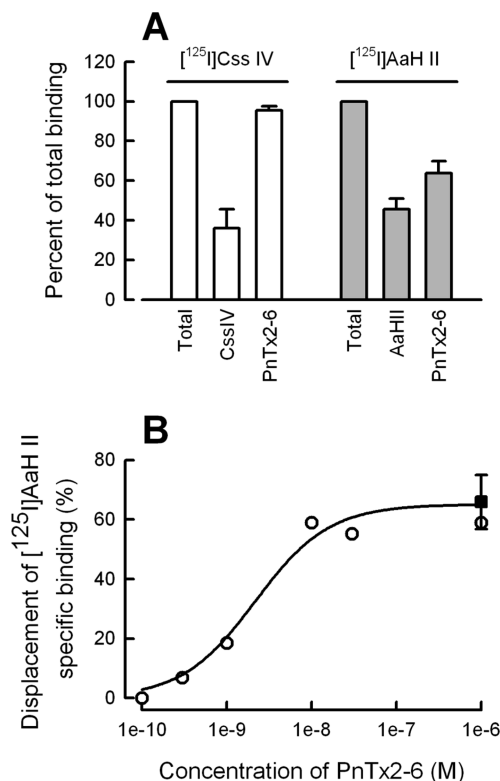


FIGURE 6: Competition binding assays of PnTx2-6. Panel A shows that PnTx2-6 does not compete with scorpion β -toxin [¹²⁵I]C_{ss}IV and partially competes with scorpion α -toxin [¹²⁵I]AaHIII. In panel B, one experiment carried out in triplicate shows the concentration dependence of the competition for the specific binding. The maximal displacement of specific binding of [¹²⁵I]AaHIII is 58%. The filled square represents the data obtained in the seven independent experiments shown in panel A, replotted as a percentage of displacement of specific binding.

The model which best satisfied the multiple evaluation criteria was selected for each toxin. Electrostatic potentials were calculated using GRASP2 (39), and structure representations were drawn using PyMol (40).

Statistics. Tests of significance were performed using an unpaired Student's *t* test. The level of significance is *p* < 0.05. Data are shown as averages \pm SEM.

RESULTS

Primary Structures of PnTx2-5 and PnTx2-6. Preliminary assays were needed to decide the correct primary structure of PnTx2-5, whose sequence was first determined by chemical sequencing (9). Via examination of a cDNA library constructed from stimulated venom glands, Kalapothakis and collaborators (41) were unable to find a sequence that fully matched the reported toxin. A similar sequence was found, in which residue 40 was a tryptophan (instead of an alanine) and cysteine 48 was not present. This sequence was named PnTx2-5a (41), because it was uncertain whether it was an isoform of PnTx2-5 or its true sequence. Mass spectroscopy of both toxins used in this work confirmed the expected mass calculated from the PnTx2-6 sequence (5289.31 ± 0.03 Da, calcd 5288.25 Da) and from PnTx2-5a (5112.31 ± 0.41 Da, calcd 5113.1 Da). Since chemical sequencing is prone to errors due to hangover after many cycles of reaction, especially when tiny amounts of material are available, and considering that the purification of our

sample was carried out according to the protocol originally used to purify PnTx2-5, we conclude that its correct sequence is the one shown in Figure 7A, which is identical to the one obtained from cDNA sequencing, while confirming the sequence of PnTx2-6 reported by chemical and cDNA sequencing (see Figure 7A). Furthermore, our data confirm the assumption that all cysteines are oxidized, forming disulfide bridges. Each of these toxins has 48 amino acids, with a 90% level of sequence identity. Using the confirmed sequences, extinction coefficients of 7570 and 14540 $M^{-1} \text{ cm}^{-1}$ were obtained for PnTx2-5 and PnTx2-6, respectively.

Comparison of PnTx2-5 and PnTx2-6 with Respect to the Na⁺ Current of GH3 Cells. GH3 cells endogenously express neuronal types of the sodium channel, Nav1.1, Nav1.2, Nav1.3, and Nav1.6 (42), whose currents are sensitive to 300 nM tetrodotoxin (not shown). PnTx2-5 and PnTx2-6 modify the Na⁺ currents in a similar manner, slowing the inactivation kinetics and generating a persistent current at the end of 20 ms pulses (Figure 1A,B). The decrease in the inactivation rate leads to a hyperexcitability and can account for the symptoms of *Phoneutria* envenomation. Preliminary experiments showed no effect on K⁺ currents. These effects are similar to those produced by scorpion α -toxins, which bind to site 3 of sodium channels. Interestingly, no reversion of the PnTx2-6 effect was observed for up to 10 min after washing out (Figure 4A), in contrast with PnTx2-5 (Figure 1A). Panels D and E of Figure 1 show the normalized current-voltage (*I*-*V*) relationships in the presence of PnTx2-5 and PnTx2-6, respectively. Na⁺ currents were activated between -35 and -40 mV, while a maximal value near -5 mV was attained under control conditions. The negative shift of the voltage dependence of activation observed in the presence of the toxins is discussed below. In six paired experiments, we observed that addition of PnTx2-5 produced a small (5.2%), but significant, increase in the current amplitude at 0 mV. In seven paired experiments, the effect of 100 nM PnTx2-6 on the current amplitude was not statistically significant.

Effect of *P. nigriventer* Toxins on the Voltage Dependence of the Na⁺ Current. To further investigate the negative shift of the Na⁺ current activation, we calculated the conductance in each potential, normalized for the maximal conductance and plotted as function of the test potential (Figure 2A,B). The points were fitted with a Boltzmann function, and the best fit is shown as a solid line. The potential that activates half of the maximal conductance (*V*_g) is significantly shifted to negative values in the presence of both toxins. PnTx2-5 at 400 nM shifted *V*_g by -11.2 ± 0.9 mV (*n* = 4) and PnTx2-6 at 100 nM by -7.8 ± 0.8 mV (*n* = 5).

In addition, we investigated the effect of both toxins on the voltage dependence of steady-state inactivation (Figure 2C,D). Conditioning prepulses (-120 to -25 mV) with a duration of 100 ms were applied and followed by a test pulse to 0 mV, where the currents were measured. Each peak current is normalized to the peak current after the -120 mV conditioning prepulse, and the data were fitted with a Boltzmann function. The best fit is shown as a solid line. As shown in Figure 2C,D, the toxins shifted the *V*_h to hyperpolarized potentials by -14.0 ± 2.4 mV (*n* = 4) and -20.3 ± 2.9 mV (*n* = 5), in the presence of PnTx2-5 (400 nM) and PnTx2-6 (100 nM), respectively. The slope factor (*K*_h) was significantly modified by the toxins, and a consider-

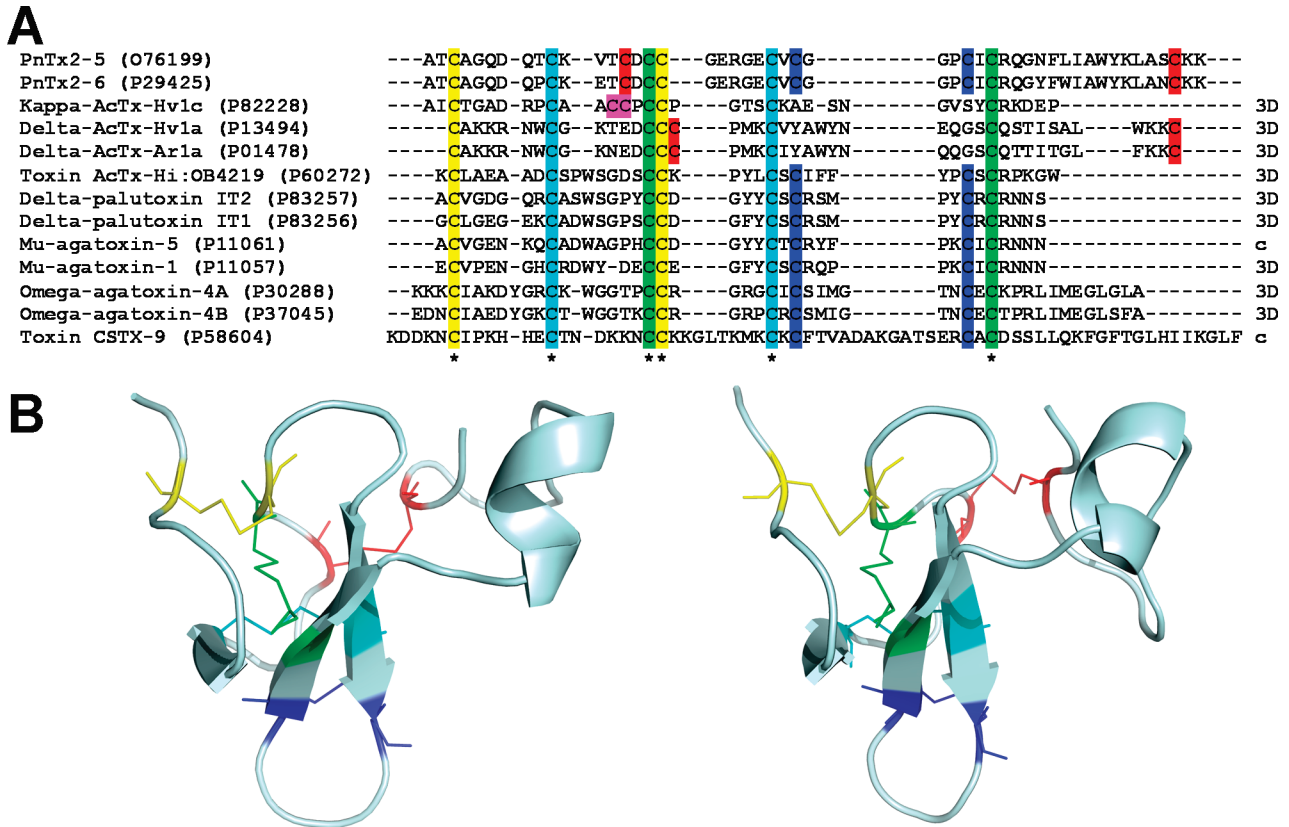


FIGURE 7: PnTx2-5 and PnTx2-6 molecular modeling. (A) Sequence and cysteine connectivity pattern of cysteine-rich spider toxins. PnTx2-5 and PnTx2-6 sequences are aligned with all cysteine-rich peptide toxins containing at least eight cysteine residues whose disulfide bridges were experimentally identified, by three-dimensional structure resolution [κ -atracotoxin-Hv1c (55), δ -atracotoxin-Hv1a (versutoxin) (29), δ -atracotoxin-Arla (robustoxin) (56), AcTx-Hi:OB4219 (32), δ -palutoxins IT2 and IT1 (30), μ -agatoxin-1 (31), ω -agatoxin-4A (57), and ω -agatoxin-4B (58)] or by chemical assignment [μ -agatoxin-5 (59) and CSTX-9 (60)]. Cysteine residues that form disulfide bridges are highlighted with the same color. (B) Cartoon representation of the three-dimensional theoretical models of PnTx2-5 (left) and PnTx2-6 (right). Disulfide bridges Cys3-Cys17, Cys10-Cys23, Cys14-Cys46, Cys16-Cys31, and Cys25-Cys29 are colored yellow, turquoise, red, green, and dark blue, respectively. The amino and carboxy termini are located on the left and right of the molecule, respectively.

Table 1: Ramachandran Plot Statistics of the Template Structures and Models^a

	no. of amino acid residues (% of non-Gly non-Pro nonterminal residues)					
	1VTX	1V91	1EIT	1KQH	PnTx2-5	PnTx2-6
most favored regions	24 (64.9)	15 (51.7)	18 (62.1)	24 (80.0)	30 (74.4)	32 (84.2)
additional allowed regions	9 (24.3)	13 (44.8)	10 (34.5)	6 (20.0)	7 (20.5)	4 (10.5)
generously allowed regions	3 (8.1)	0 (0.0)	1 (3.4)	0 (0.0)	2 (5.1)	2 (5.3)
disallowed regions	1 (2.7)	1 (3.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
all regions	37 (100.0)	29 (100.0)	29 (100.0)	30 (100.0)	39 (100.0)	38 (100.0)

^a The table shows, for each structure, the number and percentage of amino acid residues in each dihedral region of the Ramachandran plot (excluding glycine, proline, and the end terminal residues). Data were obtained using Procheck (36).

able percentage of channels (15–18%) did not inactivate even with potentials of -25 mV.

Although the effect of PnTx2-5 on the inactivation rate, K_h , and noninactivated fraction were reversible, its effect on V_g was not, while V_h reverted partially upon PnTx2-5 washout (Figure 2A,C).

Sodium Channel Inactivation Kinetics. To characterize the effect of the toxins on the inactivation process, the inactivation rates of the Na⁺ currents were analyzed. During control inactivation, Na⁺ currents decay as a single exponential with a fast time constant ($390 \pm 23 \mu\text{s}$). As shown in Figure 1, both toxins inhibit fast inactivation. We fitted the decay of individual currents at 0 mV with a single-exponential function (eq 1) or the sum of two exponentials (eq 2), in the presence of toxins PnTx2-5 and PnTx2-6. The Na⁺ currents in the presence of either toxin are best fitted with

two exponentials. PnTx2-5 modifies the Na⁺ currents by adding a slower exponential component (τ_s) to the fast component (τ_f) whose rate remained similar to the control. In contrast, PnTx2-6 slows the fast component (τ_f) in addition to generating a slow component (τ_s) (Figure 1). The same pattern is observed in all potentials where the current decay is dominated by inactivation. We used in our analysis records at 0 mV, because at that potential the sodium conductance has reached its maximum and remains maximal regardless of the voltage shift. In five experiments, the fast time constant of inactivation changed from 0.390 ± 0.023 ms in the absence of PnTx2-6 to 1.21 ± 0.15 ms after addition of PnTx2-6.

Concentration-Response Relationship. Since PnTx2-5 adds a slower time constant (τ_s) to the exponential decay, while preserving a fraction of the current with a fast time

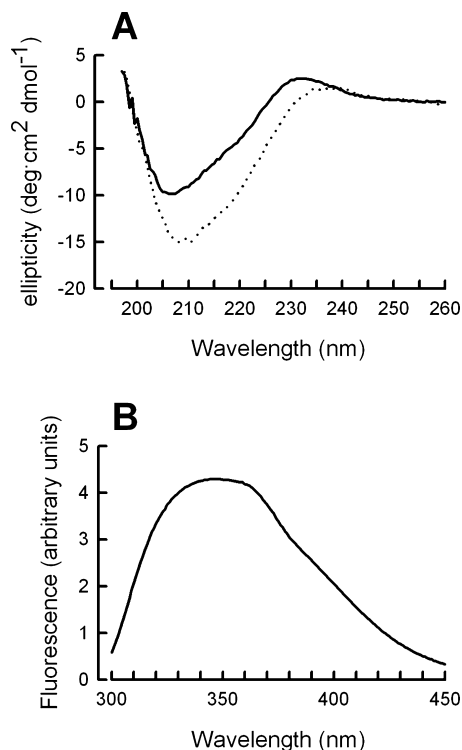


FIGURE 8: Experimental tests of the proposed model of PnTx2-6. (A) Circular dichroism spectrum of PnTx2-6 (—) compared with the spectrum of ω -atracotoxin Hv1a (48) (···). (B) Fluorescence spectrum of PnTx2-6 in water with a maximal fluorescence at 346 nm, consistent with the tryptophan residues being exposed to the hydrophilic medium.

constant (τ_f) whose value is unchanged, it is conceivable that the slow component reflects the toxin-modified channels and the fast component the unbound channels. This interpretation is supported by the experiments shown in Figure 3A. The inactivation kinetics at 0 mV at different concentrations of PnTx2-5 was analyzed, and the relative contribution of the slow component was calculated. The data were initially fitted with a Hill equation [$f(c) = P \times c^n / (K_{0.5}^n + c^n)$]. Since the value of n was found to be close to 1, a single-site saturation kinetics was considered and the value of n was fixed as 1. The best fit to the equation is shown as a solid line in Figure 3. The maximal percentage of modified current was 59%, and the concentration of PnTx2-5 that modifies half of the maximal proportion ($K_{0.5}$) was 190 nM. The same reasoning could not be applied to the PnTx2-6 effect, because it modifies the fast time constant (τ_f) besides generating the slow component. To construct a dose-response curve for PnTx2-6, we reanalyzed all curves and measured the fraction of the maximal current that remained after it decayed for a period equal to three time constants of the control. The choice of this time is critical, because it will have a significant effect on the apparent concentration dependence. Our choice was based on the following reasoning. (i) After a time equivalent to three time constants, the contribution of the unmodified channels will be less than 5%. (ii) At that time, the current generated by modified channels will still be high. For the sake of comparison, the effect of PnTx2-5 was also measured with this method and plotted with the results obtained with PnTx2-6 (Figure 3B). The results obtained with PnTx2-5 are similar to those depicted in Figure 3A, supporting the adequacy of the procedure. The graph in Figure 3B was plotted with the best fit obtained with the

same equation, having $K_{0.5}$ equal to 140 and 23 nM, and maximal effects of 58 and 71%, for PnTx2-5 and PnTx2-6, respectively. This means that, in addition to showing reversibility, PnTx2-5 has a $K_{0.5}$ 6 times larger than that of PnTx2-6. The fact that the maximal effect of PnTx2-6 is greater than that of PnTx2-5 may also be meaningful.

Dissociation of the Toxin from the Sodium Channel. We have already shown that PnTx2-5 can be washed off the sodium channel (Figure 1A). In contrast, the effect of PnTx2-6 could not be removed upon washing (Figure 4A). One important characteristic of scorpion α -toxins is that their dissociation from the sodium channel is dependent on the membrane potential and they can be displaced by strong depolarizing pulses (43–46). Since the effects of the *Phoneutria* toxins were similar to those of scorpion α -toxins, we used a strong depolarizing pulse after each test pulse to 0 mV to try to dissociate PnTx2-6 from its active site (see the inset of Figure 4B). This protocol was repeated 24 or 40 times (with a 2 s interval) in the absence of free toxin in the external solution. Figure 4B shows that the inhibition of inactivation produced by PnTx2-6 (1 μ M) remained even if 40 strong depolarizing pulses (180 mV for 20 ms) were applied alternated with the test pulses, in a medium devoid of the toxin. The rate of dissociation of PnTx2-5 was not significantly altered by the same protocol (not shown).

Although PnTx2-5 and PnTx2-6 differ by only five amino acid residues, they have significant differences in their affinities and binding properties of the sodium channel. We decided to further investigate PnTx2-6, the more potent of these toxins.

Recovery from Inactivation. To characterize the effect of PnTx2-6 on the recovery of sodium channel inactivation, we applied two depolarizing pulses to 0 mV (with a 20 ms duration) separated by a variable interval (from 0.5 to 20 s, with 0.5 s increments), when the membrane was hyperpolarized to -100 mV (see the pulse protocol in the legend of Figure 5). Figure 5 shows the peak value of the Na^+ current during the second pulse normalized by the value of the peak of the current generated by the first pulse and plotted as function of the hyperpolarization time. The curves were fitted with a double-exponential equation, and the best fit is shown as a solid line. In the presence of 500 nM PnTx2-6 (Figure 5), the channels recovered faster from inactivation. Statistical analysis shows that the proportion of each exponential component, k_1 and k_2 , was not different from the control, but the time constants of recovery, τ_1 and τ_2 , were different.

PnTx2-6 Competition Assay. The effects of PnTx2-6 resemble those of scorpion α -toxins and, less clearly, those of scorpion β -toxins. To determine if PnTx2-6 acts by binding to sites 3 and 4 of the sodium channel, we investigated if it can compete for the binding site of ^{125}I -radiolabeled AaHIII and CsxIV, typical scorpion α - and β -toxins, respectively. The results depicted in Figure 6A show that PnTx2-6 at a concentration up to 1 μ M did not compete with radiolabeled β -toxin CsxIV (100 pM) in cortical brain synaptosomes. In a control, excess unlabeled CsxIV (100 nM) displaced 64% ($n = 2$) of the radiolabeled toxin, and this percentage is considered the specific binding.

In contrast, PnTx2-6 displaced up to 36% of 100 pM radiolabeled AaHIII (Figure 6B) on the fresh P2 fraction of rat brain, whereas the excess of unlabeled AaHIII displaced 54%. This means that PnTx2-6 competes partially (67 \pm

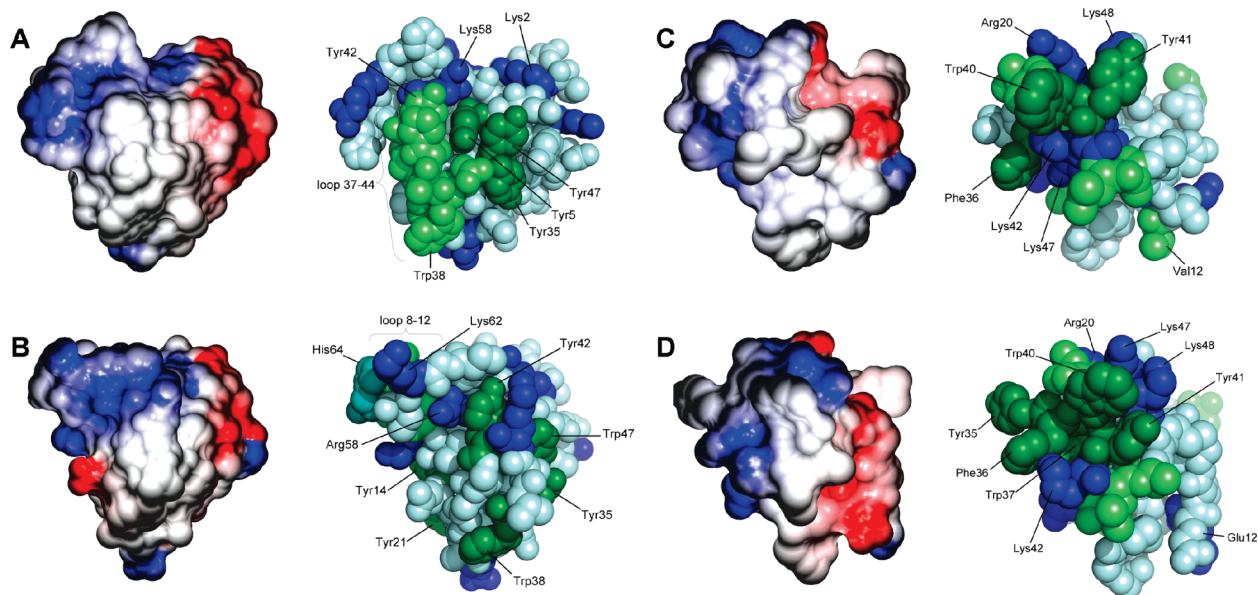


FIGURE 9: Bioactive surface comparison of site 3 mammalian sodium channel modifier neurotoxins AaHII (A), BmKM1 (B), PnTx2-5 (C), and PnTx2-6 (D). Electrostatic potential is mapped to the molecular surface of each molecule structure (left side). The potential is shown at -5.0 to $+5.0$ kT/e , with k standing for the Boltzmann constant, T the temperature, and e the electron charge. Red indicates regions of negative electrostatic potential; white indicates neutral regions, and blue indicates positive regions. These maps were drawn using GRASP2 (39). The sphere representation generated using PyMOL (40) is shown for each molecule (right side). Amino acid residues presumed or known to be involved in the interaction with the channel are colored and labeled. Blue denotes positively charged amino acid residues. Hydrophobic residues are colored light green, and aromatic residues are colored dark green.

9%) with AaHII for its specific binding sites. Because the scorpion α -toxin binding depends on the membrane potential, the P2 preparation was found to be more appropriate than synaptosomes, since it better preserves the resting potential of the membrane. When using different concentrations of PnTx2-6 (100 pM to 1 μ M), we found that the concentration that attained half of the maximal displacement was 2 nM. It is important to notice that the competition is partial, and the data suggest no tendency toward a complete competition.

Description of the Structures. Figure 7A shows the sequence alignment of PnTx2-5 and PnTx2-6 with 11 cysteine-rich toxins whose cysteine connectivities were experimentally determined. The alignment shows that we can assign to the *Phoneutria* toxins a conserved ICK motif (21). A fourth Cys bridge (dark blue) was deduced by similarity, which leads to the fifth pairing (red). Alternative pairings have no resemblance with any known Cys connectivity. Figure 7B shows a cartoon representation of PnTx2-5 and PnTx2-6 molecular models, produced as described in Materials and Methods, with a highlight on the disulfide bridges. The models predict that the three-dimensional structures of PnTx2-5 and PnTx2-6 each contain a short triple-stranded antiparallel β -sheet formed by the Gln8-Cys10, Gly21-Cys25, and Cys29-Gln33 strands and a short α -helix formed by residues 35 (Asn in Tx2-5 and Tyr in Tx2-6) to 39 (Ala in both toxins).

Model Evaluation. The theoretical structures of PnTx2-5 and PnTx2-6 were evaluated with ProSA. Energy plots of PnTx2-5 and PnTx2-6 were found to be similar to those of known structures (not shown). Verify3D evaluation scores of PnTx2-5 and PnTx2-6 structures are rated "relatively good" and similar to those obtained with the evaluation of the experimental structures of the toxins used as a template (not shown). Although some uncommon structural features can be encountered in all template and model structures, analysis by WHAT IF reports that the models show no error

and have properties of good structures (not shown). The Procheck assessment of geometric quality of the structures is shown in Table 1. The results of Ramachandran analysis show that no amino acid residue was found in a disallowed region, and most of the non-Gly and non-Pro nonterminal residues (>74%) lie in the most favored region (Table 1). We noted that the geometric quality of the model structures is even better than that of most of the template structures, except for PDB entry 1KQH. Secondary structure prediction software produced results compatible with both models. These results suggest that the proposed theoretical models can be considered probable structures of toxins PnTx2-5 and PnTx2-6 in solution. In contrast with the scorpion α -toxins, our model predicts a very low content of α -helix. This prediction was verified experimentally by CD analysis of PnTx2-6, as shown in Figure 8A. The observed spectrum supports the proposed model. In fact, the estimated proportion of α -helix using the ellipticity at 222 nm (47) is 7.7% (corresponding to 3.7 amino acid residues), which fits well with the proposed model. For the sake of comparison, we also show the experimental CD spectrum of a toxin with a similar experimentally determined structure (ω -atracotoxin Hv1a, PDB entry 1AXH) (48). Furthermore, our model predicts that the tryptophan residues are exposed to the toxin surface. This is confirmed by the fluorescence experiment shown in Figure 8B. Besides being consistent, these models predict characteristics that can be verified experimentally, as discussed.

DISCUSSION

Effects of PnTx2-5 and PnTx2-6 on Neuronal Sodium Channels. Our data clearly demonstrate that toxins PnTx2-5 and PnTx2-6 markedly delay the fast inactivation kinetics of neuronal-type sodium channels (Figure 1B). Similar results with PnTx2-6 were reported on skeletal muscle Na⁺ currents

using the loose patch clamp technique (4), but showing a lower affinity. This effect can account for the prevailing symptoms of *P. nigriventer* envenomation, and despite several similarities with the action of site 3 toxins, important differences exist that may imply differences in their mechanisms of action. In addition, our data show that PnTx2-6 has an affinity 6 times higher than that of PnTx2-5, and its effects are not reversible up to 10 min after washing.

As shown in Figure 2, there is a shift of the voltage dependence of the Na⁺ current in the presence of the toxins, which may suggest an effect similar to that of scorpion β -toxins. However, a similar spontaneous shift was described previously in GH3 cells, which tends to reach a maximum of -20 to -25 mV in 30 min (49). We found shifts of -11.2 ± 0.9 and -7.8 ± 0.8 mV for the voltage dependence of activation in the presence of PnTx2-5 and PnTx2-6, respectively. After PnTx2-5 had been washed, a shift of -13.4 ± 1.0 mV (relative to the control) was observed, while its effect on the inactivation rate was reverted. The spontaneous shifts measured in four cells maintained under control conditions were -9.3 ± 1.1 and -12.4 ± 1.2 mV, 10 and 15 min after the initial recording, respectively. This is the approximate time of a complete experiment after the whole cell configuration is obtained. We conclude that the observed shift on the activation can be explained by the spontaneous shift and cannot be attributed to a β -like effect.

In contrast, the spontaneous shifts observed on the voltage dependence of steady-state inactivation were -6.3 ± 1.2 and -7.4 ± 1.0 mV, at 10 and 15 min, respectively, which are significantly lower than those observed with each toxin. The meaningful difference, along with its partial reversion when PnTx2-5 was washed, leads to the conclusion that both toxins shifted the voltage dependence of inactivation toward more hyperpolarizing potentials.

Both toxins slowed the inactivation kinetics of the Na⁺ current. PnTx2-5 generates a slower component in addition to the normal fast one. On the other hand, PnTx2-6 had a more complex effect on neuronal sodium channels: besides adding a slower component, it also increased the time constant of the fast component (see below).

It is interesting to notice that the data shown in Figure 3 predict that, even at saturating concentrations, a significant proportion of the Na⁺ current (41%) remains unaffected by PnTx2-5. Although alternative explanations cannot be ruled out, this result suggests that part of the population of the sodium channels present in GH3 cells has its inactivation kinetics preserved, implying that PnTx2-5 may not act equally in all channel types expressed in these cells (Nav1.1, Nav1.2, Nav1.3, and Nav1.6). This result differs from what is observed with scorpion α -toxins (45) and may be the basis for some differential effects of these toxins, notably the priaptic effect of PnTx2-5. In this respect, the effect of PnTx2-6 is less clear, because of its effect on the fast component of inactivation that may overshadow the possibility that part of the population of sodium channels is resistant to PnTx2-6. A small increase in the time constant of inactivation of part of the population of sodium channels, not sufficiently high to be distinguishable from the unaffected channels, would produce this effect. The incomplete competition observed in Figure 6 reinforces the interpretation that there are neuronal types of sodium channels that bind AaHII, but not PnTx2-6.

It is important to notice that PnTx2-6 reduced the time for the sodium channel to recover from the inactivation state. This effect is similar to that reported for scorpion α -toxins (46) and may reflect its mechanism of action, as discussed below.

Mode of Action. The effects described in this paper resemble in many aspects those reported for scorpion α -toxins, namely, the decrease in the inactivation rate, the appearance of a noninactivating current, and the decrease in the time for recovery from inactivation. Furthermore, the experiments depicted in Figure 6 show that PnTx2-6 can displace 67% of the specific binding of AaHII, a typical scorpion α -toxin, while not competing with CssIV (a typical scorpion β -toxin) for its binding site. It is interesting to notice that the ability to displace AaHII is saturated at that level. This means that PnTx2-6 does not bind exactly to the same site (i.e., the same amino acid arrangement) as the scorpion α -toxins, while there may be an overlap. The simplest explanation is that PnTx2-6 cannot bind to all types of sodium channels present in the GH3 cells, while sharing overlapping sites where it can bind. This possibility requires further investigation using heterologously expressed cloned channels.

One striking difference between these toxins is shown in Figure 4B. It is well-known that the site 3 toxins, including sea anemone toxin ATX II, are removed by strong depolarizing pulses, and this property seems intrinsically related to their mechanism of action (45, 50). We report for the first time a toxin with effects similar to those of scorpion α -toxins, and possibly overlapping binding sites, that is not removed by strong depolarizing pulses. The affinity of PnTx2-6 is lower than that of the scorpion toxins and cannot account for this absence of displacement. This observation is currently under investigation.

A kinetic model that explains the effect of site 3 toxins was put forward by Campos and collaborators (45). This model explains the removal of site 3 toxins and predicts its voltage dependence. It also predicts the recovery from inactivation to be faster in the presence of these toxins, which was confirmed experimentally (46). In this work, we show a significant decrease in the time constant of recovery from inactivation in the presence of PnTx2-6 (Figure 5).

We propose that PnTx2-6 binds to a site that overlaps site 3 of the sodium channel, thus producing similar effects on the kinetic properties of the Na⁺ current.

Structural Comparison with Sodium Channel Scorpion α -Toxins. We compared the primary structures of PnTx2-5 and PnTx2-6 toxins with those of scorpion α -toxins and did not find any significant sequence similarity, even in the regions supposed to be participating in the interaction with the channel. However, the possible overlapping binding site of PnTx2-6 and similar electrophysiological properties of PnTx2-5 and PnTx2-6 with scorpion α -toxins prompted us to verify the hypothesis that, despite the lack of sequence similarity, they may have similar bioactive surfaces. It has been shown that scorpion α -toxins have a conserved hydrophobic (predominantly aromatic) core encompassed by basic residues (51, 52). Figure 9 shows the comparison of the predicted PnTx2-5 and PnTx2-6 surfaces with the potent α -toxin AaHII and with the well-studied scorpion α -like toxin BmKM1, from *Buthus martensii* Karsch. Similar surfaces can be identified, with a conserved hydrophobic

patch surrounded by positive charges. This observation is consistent with the hypothesis that the active surface of gating modifier toxins is an arrangement of hydrophobic and basic residues (51, 53). Moreover, of the residues that are nonidentical between the two toxins, two of them are located in the hydrophobic core and are exposed on the surface. One tyrosine (at position 35) and one tryptophan (at position 37), amino acids important for protein–protein interaction (54), are absent from the hydrophobic core of PnTx2–5.

The fluorescence data of PnTx2–6 show that the tryptophan residues are exposed to the solvent, as predicted in the models. Aromatic residues Phe36, Trp40, and Tyr41 are also exposed in both toxins. They also contain the same charged amino acid residues, at the same locations. In particular, positively charged residues Arg20, Arg32, Lys42, Lys47, and Lys48 surround the hydrophobic patch in the two structures (Figure 9C,D). This arrangement provides the expected basic requirement for the interaction with site 3 of the sodium channel and conforms to the “recognition patch” found in protein–protein interacting surfaces (54). We propose that this charge disposition together with the surrounding aromatic patch (residues Phe36, Trp40, and Tyr41) composes the bioactive surface of PnTx2–5 and PnTx2–6. The additional presence of Tyr35 and Trp37 in PnTx2–6 can account for its higher affinity for the channel.

ACKNOWLEDGMENT

We thank Dr. M.-F. Martin-Eauclaire for her generous gift of the pure samples of AaHIII and CsxIV scorpion toxins and Dr. Glenn King for kindly giving his CD data for the toxin ω -ACTX-Hv1a. We are indebted to Dr. Claude Granier for reviewing the manuscript.

SUPPORTING INFORMATION AVAILABLE

PnTx-5 and PnTx2–6 theoretical model three-dimensional coordinates. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

- Billen, B., Bosmans, F., and Tytgat, J. (2008) Animal peptides targeting voltage-activated sodium channels. *Curr. Pharm. Des.* 14, 2492–2502.
- Campos, F. V., Chanda, B., Beirão, P. S., and Bezanilla, F. (2007) β -Scorpion toxin modifies gating transitions in all four voltage sensors of the sodium channel. *J. Gen. Physiol.* 130, 257–268.
- Cestèle, S., and Catterall, W. A. (2000) Molecular mechanisms of neurotoxin action on voltage-gated sodium channels. *Biochimie* 82, 883–892.
- Matavel, A., Cruz, J. S., Penaforte, C. L., Araújo, D. A., Kalapothakis, E., Prado, V. F., Diniz, C. R., Cordeiro, M. N., and Beirão, P. S. (2002) Electrophysiological characterization and molecular identification of the *Phoneutria nigriventer* peptide toxin PnTx2–6. *FEBS Lett.* 523, 219–223.
- Araújo, D. A., Cordeiro, M. N., Diniz, C. R., and Beirão, P. S. (1993) Effects of a toxic fraction, PhTx2, from the spider *Phoneutria nigriventer* on the sodium current. *Naunyn-Schmiedeberg's Arch. Pharmacol.* 347, 205–208.
- Monjaraz, E., Navarrete, A., Lopez-Santiago, L. F., Vega, A. V., Arias-Montano, J. A., and Cota, G. (2000) L-type calcium channel activity regulates sodium channel levels in rat pituitary GH3 cells. *J. Physiol.* 523 (Part 1), 45–55.
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B., and Sigworth, F. J. (1981) Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pfluegers Arch.* 391, 85–100.
- Rezende Junior, L., Cordeiro, M. N., Oliveira, E. B., and Diniz, C. R. (1991) Isolation of neurotoxic peptides from the venom of the ‘armed’ spider *Phoneutria nigriventer*. *Toxicon* 29, 1225–1233.
- Cordeiro, M. N., Diniz, C. R., Valentim Ado, C., von Eickstedt, V. R., Gilroy, J., and Richardson, M. (1992) The purification and amino acid sequences of four Tx2 neurotoxins from the venom of the Brazilian ‘armed’ spider *Phoneutria nigriventer* (Keys). *FEBS Lett.* 310, 153–156.
- Edelhoc, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* 6, 1948–1954.
- Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326.
- Rochat, H., Tessier, M., Miranda, F., and Lissitzky, S. (1977) Radioiodination of scorpion and snake toxins. *Anal. Biochem.* 82, 532–548.
- Dunkley, P. R., Heath, J. W., Harrison, S. M., Jarvie, P. E., Glenfield, P. J., and Rostas, J. A. (1988) A rapid Percoll gradient procedure for isolation of synaptosomes directly from an S1 fraction: Homogeneity and morphology of subcellular fractions. *Brain Res.* 441, 59–71.
- Gray, E. G., and Whittaker, V. P. (1962) The isolation of nerve endings from brain: An electron-microscopic study of cell fragments derived by homogenization and centrifugation. *J. Anat.* 96, 79–88.
- Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951) Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* 193, 265–275.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404–405.
- Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.
- Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997) CATH: A hierarchic classification of protein domain structures. *Structure* 5, 1093–1108.
- Marti-Renom, M. A., Pieper, U., Madhusudhan, M. S., Rossi, A., Eswar, N., Davis, F. P., Al-Shahrouh, F., Dopazo, J., and Sali, A. (2007) DBALI tools: Mining the protein structure space. *Nucleic Acids Res.* 35, W393–W397.
- Craik, D. J., Daly, N. L., and Waite, C. (2001) The cystine knot motif in toxins and implications for drug design. *Toxicon* 39, 43–60.
- Gelly, J. C., Gracy, J., Kaas, Q., Le-Nguyen, D., Heitz, A., and Chiche, L. (2004) The KNOTTIN website and database: A new information system dedicated to the knottin scaffold. *Nucleic Acids Res.* 32, D156–D159.
- Conticello, S. G., Pilpel, Y., Glusman, G., and Fainzilber, M. (2000) Position-specific codon conservation in hypervariable gene families. *Trends Genet.* 16, 57–59.
- Rost, B., Sander, C., and Schneider, R. (1994) PHD: An automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* 10, 53–60.
- Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998) JPred: A consensus secondary structure prediction server. *Bioinformatics* 14, 892–893.
- Kneller, D. G., Cohen, F. E., and Langridge, R. (1990) Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* 214, 171–182.
- Rost, B., Sander, C., and Schneider, R. (1994) Redefining the goals of protein secondary structure prediction. *J. Mol. Biol.* 235, 13–26.
- Kozlov, S., and Grishin, E. (2005) Classification of spider neurotoxins using structural motifs by primary structure features. Single residue distribution analysis and pattern analysis techniques. *Toxicon* 46, 672–686.
- Fletcher, J. I., Chapman, B. E., Mackay, J. P., Howden, M. E., and King, G. F. (1997) The structure of versutoxin (δ -atracotoxin-Hv1) provides insights into the binding of site 3 neurotoxins to the voltage-gated sodium channel. *Structure* 5, 1525–1535.
- Ferrat, G., Bosmans, F., Tytgat, J., Pimentel, C., Chagot, B., Gilles, N., Nakajima, T., Darbon, H., and Corzo, G. (2005) Solution structure of two insect-specific spider toxins and their pharmacological interaction with the insect voltage-gated Na⁺ channel. *Proteins* 59, 368–379.

31. Omecinsky, D. O., Holub, K. E., Adams, M. E., and Reily, M. D. (1996) Three-dimensional structure analysis of μ -agatoxins: Further evidence for common motifs among neurotoxins with diverse ion channel specificities. *Biochemistry* 35, 2836–2844.
32. Rosengren, K. J., Wilson, D., Daly, N. L., Alewood, P. F., and Craik, D. J. (2002) Solution structures of the cis- and trans-Pro30 isomers of a novel 38-residue toxin from the venom of *Hadronyche infensa* sp. that contains a cystine-knot motif within its four disulfide bonds. *Biochemistry* 41, 3294–3301.
33. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
34. Fiser, A., and Sali, A. (2003) Modeller: Generation and refinement of homology-based protein structure models. *Methods Enzymol.* 374, 461–491.
35. Vriend, G. (1990) WHAT IF: A molecular modeling and drug design program. *J. Mol. Graphics* 8, 52–56, 29.
36. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
37. Eisenberg, D., Luthy, R., and Bowie, J. U. (1997) VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 277, 396–404.
38. Wiederstein, M., and Sippl, M. J. (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410.
39. Petrey, D., and Honig, B. (2003) GRASP2: Visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Methods Enzymol.* 374, 492–509.
40. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA.
41. Kalapothakis, E., Penaforte, C. L., Beirão, P. S., Romano-Silva, M. A., Cruz, J. S., Prado, M. A., Guimarães, P. E., Gomez, M. V., and Prado, V. F. (1998) Cloning of cDNAs encoding neurotoxic peptides from the spider *Phoneutria nigriventer*. *Toxicon* 36, 1843–1850.
42. Vega, A. V., Espinosa, J. L., Lopez-Dominguez, A. M., Lopez-Santiago, L. F., Navarrete, A., and Cota, G. (2003) L-type calcium channel activation up-regulates the mRNAs for two different sodium channel α subunits (Nav1.2 and Nav1.3) in rat pituitary GH3 cells. *Brain Res. Mol. Brain Res.* 116, 115–125.
43. Rogers, J. C., Qu, Y., Tanada, T. N., Scheuer, T., and Catterall, W. A. (1996) Molecular determinants of high affinity binding of α -scorpion toxin and sea anemone toxin in the S3-S4 extracellular loop in domain IV of the Na⁺ channel α subunit. *J. Biol. Chem.* 271, 15950–15962.
44. Chen, H., and Heinemann, S. H. (2001) Interaction of scorpion α -toxins with cardiac sodium channels: Binding properties and enhancement of slow inactivation. *J. Gen. Physiol.* 117, 505–518.
45. Campos, F. V., Coronas, F. I., and Beirão, P. S. (2004) Voltage-dependent displacement of the scorpion toxin Ts3 from sodium channels and its implication on the control of inactivation. *Br. J. Pharmacol.* 142, 1115–1122.
46. Campos, F. V., and Beirão, P. S. (2006) Effects of bound ts3 on voltage dependence of sodium channel transitions to and from inactivation and energetics of its unbinding. *Cell Biochem. Biophys.* 44, 424–430.
47. Chen, Y. H., Yang, J. T., and Martinez, H. M. (1972) Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry* 11, 4120–4131.
48. Tedford, H. W., Gilles, N., Menez, A., Doering, C. J., Zamponi, G. W., and King, G. F. (2004) Scanning mutagenesis of ω -atracotoxin-Hv1a reveals a spatially restricted epitope that confers selective activity against insect calcium channels. *J. Biol. Chem.* 279, 44133–44140.
49. Fernandez, J. M., Fox, A. P., and Krasne, S. (1984) Membrane patches and whole-cell membranes: A comparison of electrical properties in rat clonal pituitary (GH3) cells. *J. Physiol.* 356, 565–585.
50. Campos, F. V., Chanda, B., Beirão, P. S., and Bezanilla, F. (2008) α -Scorpion toxin impairs a conformational change that leads to fast inactivation of muscle sodium channel. *J. Gen. Physiol.* (in press).
51. Kharrat, R., Darbon, H., Rochat, H., and Granier, C. (1989) Structure/activity relationships of scorpion α -toxins. Multiple residues contribute to the interaction with receptors. *Eur. J. Biochem.* 181, 381–390.
52. Fontecilla-Camps, J. C., Habersetzer-Rochat, C., and Rochat, H. (1988) Orthorhombic crystals and three-dimensional structure of the potent toxin II from the scorpion *Androctonus australis* Hector. *Proc. Natl. Acad. Sci. U.S.A.* 85, 7443–7447.
53. Takahashi, H., Kim, J. I., Min, H. J., Sato, K., Swartz, K. J., and Shimada, I. (2000) Solution structure of hanatoxin1, a gating modifier of voltage-dependent K⁺ channels: Common surface features of gating modifier toxins. *J. Mol. Biol.* 297, 771–780.
54. Chakrabarti, P., and Janin, J. (2002) Dissecting protein-protein recognition sites. *Proteins* 47, 334–343.
55. Wang, X., Connor, M., Smith, R., Maciejewski, M. W., Howden, M. E., Nicholson, G. M., Christie, M. J., and King, G. F. (2000) Discovery and characterization of a family of insecticidal neurotoxins with a rare vicinal disulfide bridge. *Nat. Struct. Biol.* 7, 505–513.
56. Pallaghy, P. K., Alewood, D., Alewood, P. F., and Norton, R. S. (1997) Solution structure of robustoxin, the lethal neurotoxin from the funnel-web spider *Atrax robustus*. *FEBS Lett.* 419, 191–196.
57. Reily, M. D., Holub, K. E., Gray, W. R., Norris, T. M., and Adams, M. E. (1994) Structure-activity relationships for P-type calcium channel-selective ω -agatoxins. *Nat. Struct. Biol.* 1, 853–856.
58. Reily, M. D., Thanabal, V., and Adams, M. E. (1995) The solution structure of ω -Aga-IVB, a P-type calcium channel antagonist from venom of the funnel web spider, *Agelenopsis aperta*. *J. Biomol. NMR* 5, 122–132.
59. Skinner, W. S., Adams, M. E., Quistad, G. B., Kataoka, H., Cesarin, B. J., Enderlin, F. E., and Schooley, D. A. (1989) Purification and characterization of two classes of neurotoxins from the funnel web spider, *Agelenopsis aperta*. *J. Biol. Chem.* 264, 2150–2155.
60. Schaller, J., Kampf, U., Schurch, S., Kuhn-Nentwig, L., Haerberli, S., and Nentwig, W. (2001) CSTX-9, a toxic peptide from the spider *Cupiennius salei*: Amino acid sequence, disulphide bridge pattern and comparison with other spider toxins containing the cystine knot structure. *Cell. Mol. Life Sci.* 58, 1538–1545.

BI802158P