UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
CURSO DE BIOMEDICINA

# CARACTERIZAÇÃO ESTRUTURAL E CONFORMACIONAL DA PROTROMBINA HUMANA

Carla Gottschald Chiodi

ORIENTADOR: Prof. Dr. Hugo Verli
CO-ORIENTADORA: MSc. Cláudia Lemelle Fernandes

Porto Alegre – Brasil
2010

# CARACTERIZAÇÃO ESTRUTURAL E CONFORMACIONAL DA PROTROMBINA HUMANA

*Carla Gottschald Chiodi*

Trabalho de conclusão de curso de Biomedicina elaborado no Grupo de Bioinformática Estrutural do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul sob orientação do professor doutor:

**Hugo Verli**

Porto Alegre – Brasil
2010

**AGRADECIMENTOS**

Ao Prof. Hugo Verli e à Cláudia Lemelle Fernandes, por terem me recebido junto ao Grupo de Bioinformática Estrutural, por terem me orientado e, principalmente, me ensinado não só a metodologia, mas também a questionar e interpretar os dados, auxiliando em meu crescimento intelectual e profissional. Além disso, agradeço também pela amizade e confiança durante esse período.

A UFRGS, aos responsáveis pelo curso de Biomedicina e ao Centro de Biotecnologia, por todo esforço e dedicação no sentido de oferecer para os alunos um ensino de qualidade, com infra-estrutura adequada, e de incentivar a Iniciação Científica.

Aos alunos do Grupo de Bioinformática Estrutural, pela amizade, convivência e auxílio, e nesse sentido não poderia deixar de agradecer ao Laércio Pol Fachin e ao Rodrigo Ligabue Braun.

Aos amigos e colegas de curso que, apesar de não gostarem de bioinformática, sempre me apoiaram e incentivaram, não só no âmbito profissional como no pessoal.

Aos membros da banca, por terem aceitado o convite.

Aos meus pais agradeço eternamente pelo imenso apoio, interesse, responsabilidade, ensinamentos e incentivo que me deram desde a minha infância e que se mostraram tão presentes e necessários durante esta etapa da minha formação.

***"Todas as verdades são fáceis de entender quando descobertas; o problema é descobri-las."***

Galileu Galilei

***"Pensar é o trabalho mais difícil que existe. Talvez por isso tão poucos se dediquem a ele."***

Henry Ford

***"Talvez a imaginação seja somente a inteligência se divertindo."***

George Scialabba

*Dedico este trabalho a todos amantes da ciência e a meus pais, Aires José Chiodi e Elisabeth Gottschald Chiodi,que me deram a educação, o apoio e o amor necessários para eu chegar até aqui e ir muito mais além.*

**ÍNDICE GERAL**

## LISTA DE ABREVIATURAS

CG – *Coarse-grained*

DM – Dinâmica Molecular

GLA – Gama-carboxiglutâmico

FFT – *Fast Fourier Transform*

FIX(a) – Fator IX (ativado)

FT – Fator Tecidual

FV (a) – Fator V (ativado)

FVII(a) – Fator VII (ativado)

FVIII(a) – Fator VIII (ativado)

FvW – Fator de Von Willebrand

FX(a) – Fator X (ativado)

FXI(a) – Fator XI (ativado)

FXIII(a) – Fator XIII (ativado)

LMWH – Heparina de Baixo Peso Molecular

RMN – Ressonância Magnética Nuclear

TFPI – *Tissue Factor Pathway Inhibitor* (Inibidor da Via do Fator Tecidual)

UFH – Heparina Não Fracionada

**RESUMO**

A protrombina é um importante zimogênio da cascata de coagulação. Sua forma ativa, a α-trombina, é a principal enzima na conversão de fibrinogênio solúvel em monômeros insolúveis de fibrina, os quais se organizam como uma rede protéica que estabiliza o tampão plaquetário e, assim, contribui para o controle de processos hemorrágicos. O complexo protrombinase, formado pelo fator Xa, fator Va, $Ca^{2+}$ e membrana de fosfolipídeos aniônicos, é essencial na ativação da protrombina, processo que culmina com a geração de α-trombina e fragmento F1.2. Por ser substrato de uma reação enzimática, a protrombina tende a ser uma molécula muito flexível, fato que pode ser correlacionado com a falta de sua estrutura cristalográfica. Nesse contexto, esse trabalho emprega técnicas de modelagem comparativa, cálculos de *docking* e simulações de dinâmica molecular na busca por um modelo capaz de auxiliar na interpretação estrutural de suas funções biológicas em nível atomístico. Os modelos obtidos foram devidamente validados, indicando a presença de movimentos de dobradiça entre os domínios da protrombina. Adicionalmente, regiões de flexibilidade destacadas foram observadas em regiões adicionais da proteína, como seu *N*-terminal (do resíduo 1 ao 327), característica esta que pode ser associadas a sua suscetibilidade à proteólise. A região *C*-terminal (domínio serino protease), em contrapartida, apresenta-se mais rígida, o que pode ser relacionado ao fato do domínio serino protease corresponder à trombina em sua forma ativa. Devido a essa alta flexibilidade da molécula, foram encontrados 3 estados conformacionais prevalentes em solução. A partir dos dados obtidos, espera-se que o modelo obtido possa ser empregado em futuros experimentos de desenvolvimento de novos agentes antitrombóticos.

*Palavras Chave:* Protrombina; Cascata de Coagulação; Modelagem Comparativa; *Docking*; *Coarse-Grained*.

## ÍNDICE DE FIGURAS

# 1. Introdução

## *1.1. A cascata de coagulação*

O modelo de cascata da coagulação foi proposto no início dos anos de 1960 quase simultaneamente por McFarlane, Davie e Ratnoff e, posteriormente, expandido por Morawitz (Becker, 2005). Nesse modelo, há 2 vias que iniciam a cascata, a intrínseca e a extrínseca. A primeira inicia-se com a exposição de cargas negativas que ativam, sucessivamente, uma série de fatores, sendo o último o fator X (FX). A via extrínseca começa com a exposição do fator tecidual (FT) e sua ligação com o fator VII (FVII), terminando com a ativação do FX também. A partir daí, as 2 vias culminam na via comum, onde a trombina é gerada para formar o coágulo de fibrina.

Embora empregados deste então, os termos "vias intrínseca e extrínseca" vêm recentemente entrando em desuso por limitar a aplicação deste modelo no entendimento da hemostasia e da trombose *in vivo* (Becker, 2005). Um exemplo disso seria que a cascata não consegue explicar porque os hemofílicos têm problemas de coagulação, ou seja, porque a via extrínseca não compensa o déficit da via intrínseca (Hoffman, 2003). Devido a esses problemas, foi proposto um modelo baseado em células que consiste numa visão fisiológica, integrada e funcional dos eventos bioquímicos que ocorrem na superfície celular durante a coagulação. Didaticamente, dividiu-se o evento inteiro em 6 subprocessos: iniciação, amplificação, propagação, terminação, eliminação e estabilização (Becker, 2005).

Em situações normais, perturbações no endotélio vascular por injúrias e estímulos inflamatórios dão início à coagulação. Nessa etapa de iniciação, algumas células da circulação sanguínea – *e.g.* monócitos, macrófagos, neutrófilos, células endoteliais ativadas, células musculares lisas – passam a expor o FT, ativando o fator VII, que forma o complexo FVIIa-FT. Esse complexo ativa o FX, o qual ativa o fator V (FV) que, por sua vez, gera uma pequena quantidade de trombina a partir da clivagem da protrombina (Becker, 2005).

Na fase de amplificação, a trombina formada anteriormente ativa plaquetas (agregadas previamente pelo fator de von Willebrand (FvW) solúvel e de superfície) e outros fatores de coagulação ancorados na membrana plaquetária, tais como FV, fator XI (FXI) e fator VIII (FVIII). Além disso, a superfície negativa exposta, decorrente do dano endotelial, também ativa os fatores XII, XI e IX (Guyton & Hall, 2006).

Na fase de propagação, há formação de complexos que são compostos pelos fatores ativados, *i.e.* tenase e protrombinase. O complexo tenase, composto por FIXa, FVIIIa, membranas e cálcio, ativa FX na superfície de plaquetas gerando FXa que, por sua vez, se complexa com FVa, cálcio e membranas para formar o complexo protrombinase. Este complexo cliva a protrombina em trombina, dessa vez produzida em grandes quantidades, promovendo uma significativa conversão de fibrinogênio solúvel em monômeros insolúveis de fibrina, além de promover a ativação do fator XIII (FXIII), que estabiliza o coágulo (Jenny & Mann, 1998) (Figura 1).

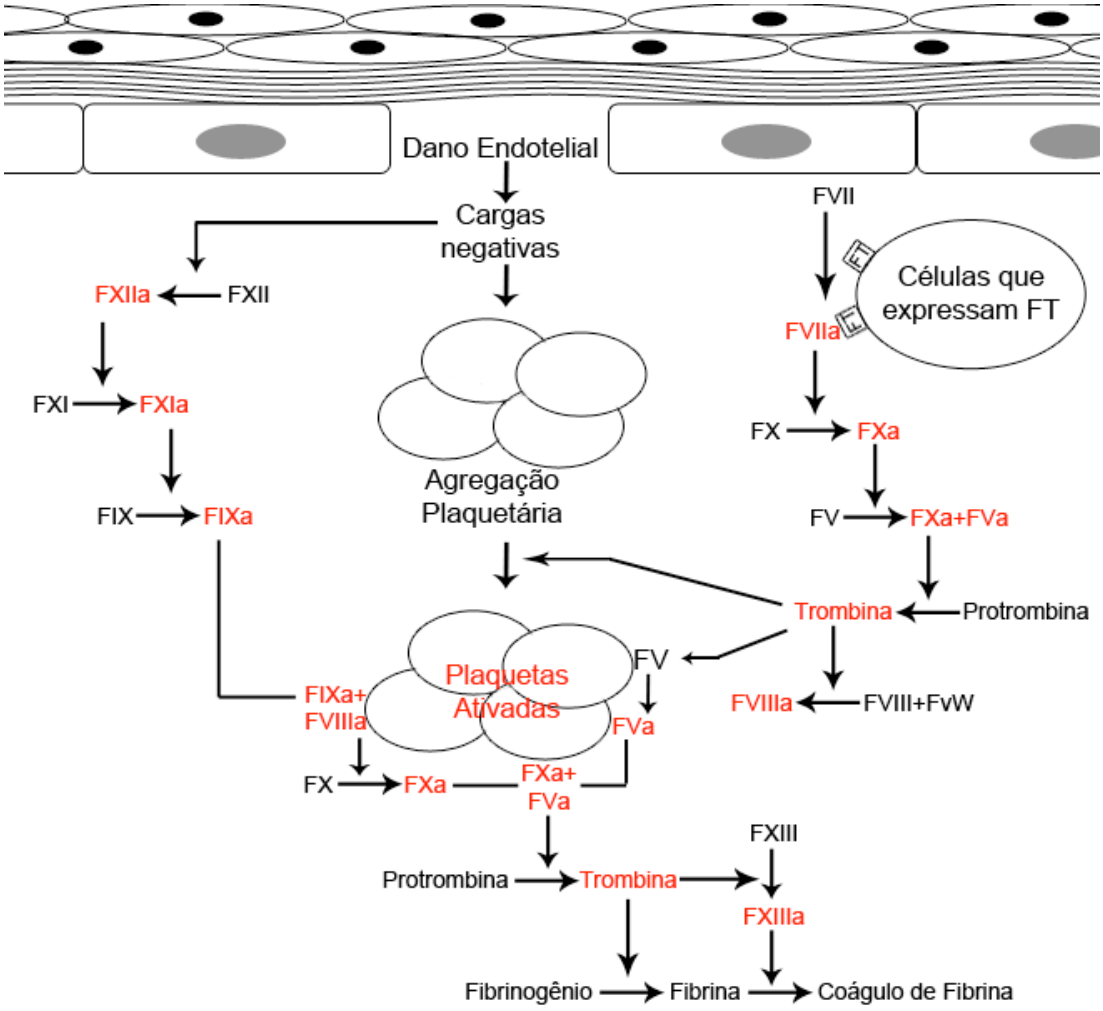



Figura 1: Representação esquemática da cascata da coagulação baseado no modelo de células, onde as principais ativações ocorrem na superfície plaquetária.

O controle e regulação da coagulação envolvem três principais moléculas: proteína C, TFPI (*tissue factor pathway inhibitor*, inibidor da via do fator tecidual) e antitrombina III. A proteína C é dependente de vitamina K, porque possui resíduos de glutamato modificados pela reação de carboxilação dependente de vitamina K,

formando o domínio gama-carboxiglutâmico (GLA). É ativada pelo complexo trombina-trombomodulina e pela ligação com seu cofator, a proteína S, resultando no complexo proteína C-proteína S capaz de inibir FVa e FVIIIa. O TFPI é liberado de células endoteliais e plaquetas e tem a função de inibir o FT, FVIIa e FXa. A antitrombina III, por sua vez, é uma serpina (*serine proteinase inhibitor*, inibidor de serino proteinases) cuja função é inibir as principais proteases da coagulação, *i.e.* trombina, FIXa, FXa, FXIa e complexo FVIIa-FT (Jenny & Mann, 1998).

Subsequente a esse processo de terminação, vem o de eliminação, que é a fibrinólise em si e cujo principal mecanismo consiste na ação do ativador do plasminogênio, que, como o próprio nome diz, ativa o plasminogênio em plasmina para degradação da fibrina. Concomitante a isso ocorre o processo de estabilização, que seria a contra-reação da eliminação, causada pelo aumento de FXIIIa, que estabiliza a fibrina. Portanto, tem de haver sempre um equilíbrio durante essas etapas a fim de não causar trombos nem hemorragia (Jenny & Mann, 1998).

### *1.2. Protrombina*

A protrombina é uma proteína sintetizada no fígado como uma glicoproteína de cadeia única com um peptídeo sinal correspondendo aos seus 43 resíduos de aminoácidos iniciais. A biossíntese dessa proteína envolve, ainda, a inclusão de carboidratos e a remoção do peptídeo sinal antes de sua secreção na circulação sanguínea. A inclusão de carboidratos refere-se à N-glicosilação das Asn78, Asn100 e Asn373 (Mizuochi, 1981) com ácido siálico, galactose, N-acetil-glicosamina e manose (Figura 2). A protrombina madura assim produzida possui 579 aminoácidos e 4 domínios (Figura 3): GLA, kringle 1, kringle 2 e serino protease (Degen & Davie, 1987). O domínio GLA é composto por cerca de 48 resíduos, sendo 10 glutamatos carboxilados pelo processo de gama-carboxilação dependente de vitamina K, domínio este necessário para a ligação da protrombina a íons cálcio e a membranas (Degen & Davie, 1987). Os domínios kringle são altamente conservados, contendo um padrão de alças estabilizadas por 3 pontes dissulfeto e, a partir deste arranjo, nomeado por sua semelhança a um *pretzel* escandinavo. Sua função está relacionada a interação com outras proteínas, principalmente o FV (Esmon & Jackson, 1974). Por fim, o domínio serino protease contém a região catalítica que irá corresponder a forma ativa da trombina (Elion, 1977).

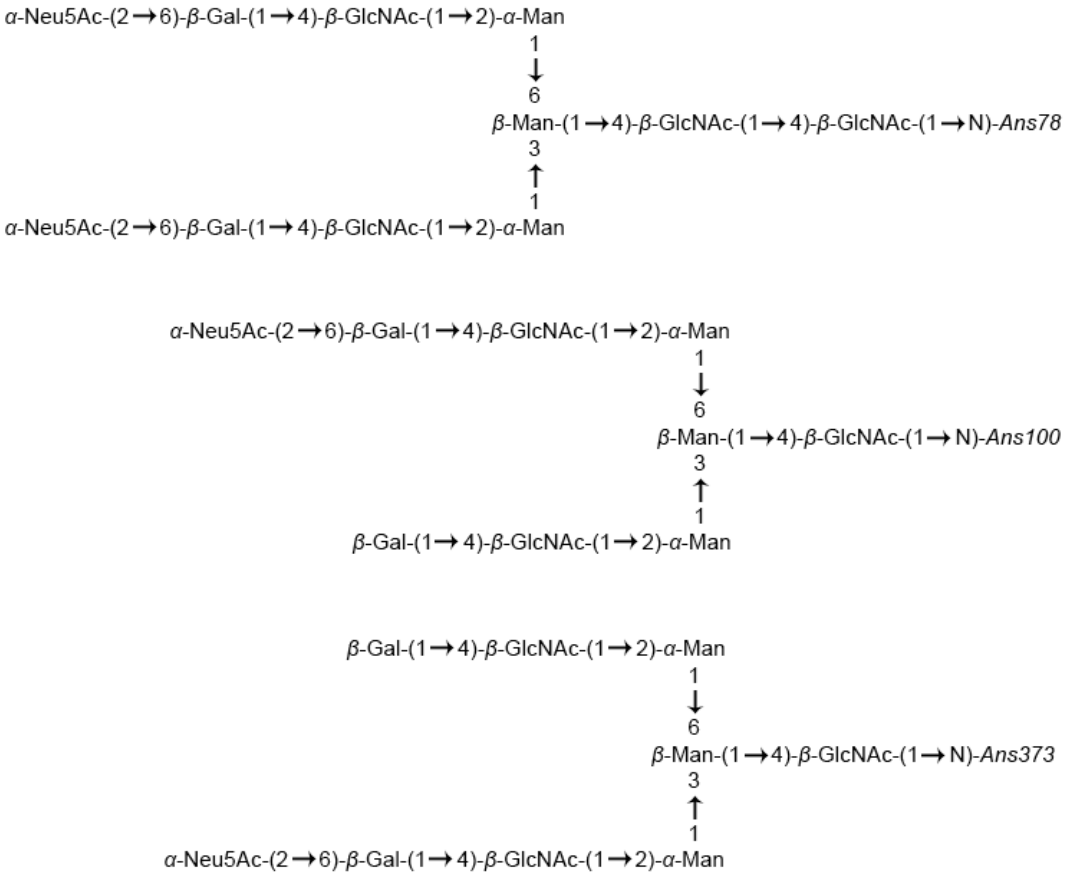



Figura 2: Árvores glicídicas da protrombina humana.

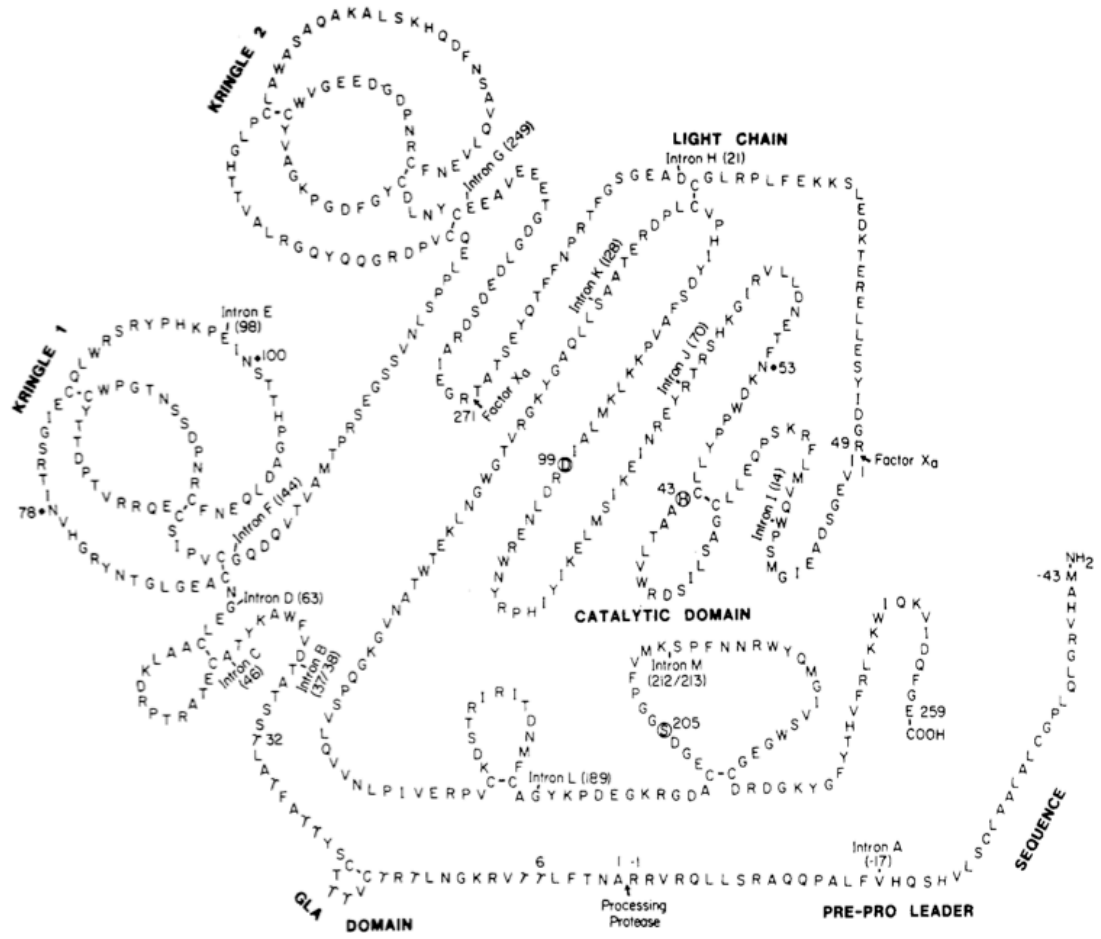



Figura 3: Representação bidimensional da sequência de aminoácidos da protrombina, mostrando os domínios, os sítios de clivagem e as pontes dissulfeto (Degen & Davie, 1987).

Quando há injúria no endotélio, a série de processos bioquímicos descritos acima leva à formação do complexo protrombinase. Neste complexo, FVa se liga ao domínio kringle 2 da protrombina (Kotkow, 1995), favorecendo a ação do FXa em 2 sítios deste zimogênio: Arg320-Ile321 e Arg271-Thr272. A ordem das clivagens só altera os intermediários – o fragmento 1.2 e pré-trombina ou a meizotrombina –, porém o produto final é o mesmo – o fragmento 1.2 e a α-trombina (Nesheim Mann, 1983; Krishnaswamy, 1987) (Figura 4).

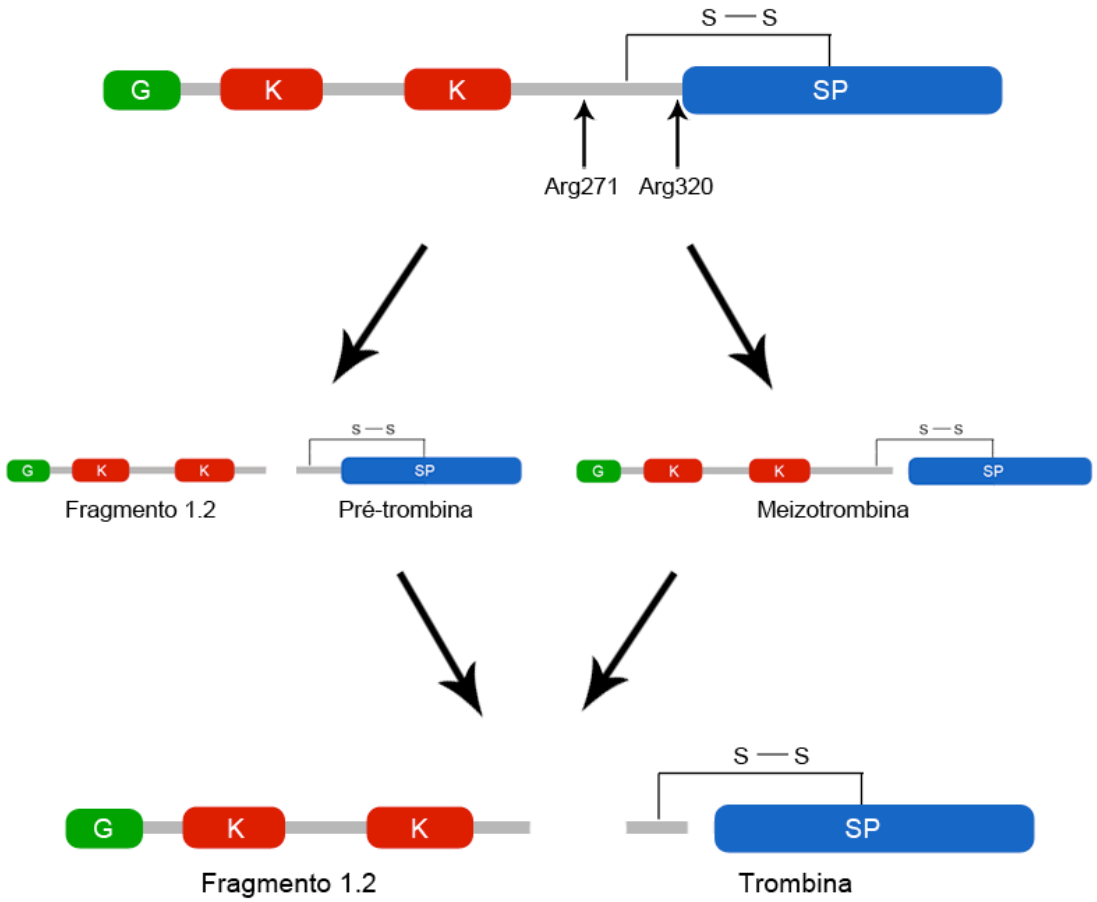



Figura 4: Representação das clivagens da protrombina nas 2 possíveis ordens, seus respectivos produtos e domínios. Em verde é o domínio GLA (G), em vermelho são os kringle (K) e em azul é o serino protease (SP).

Desde a descoberta da ação anticoagulante da heparina em 1916 por McLean, no Canadá, a partir de uma amostra de fígado de cão (Nader, 2001), o uso e desenvolvimento de novos fármacos desse tipo avançou muito no âmbito de prevenção e tratamento de eventos tromboembólicos. Atualmente, os anticoagulantes disponíveis são heparina de baixo peso molecular (LMWH), heparina não-fracionada (UFH), inibidores diretos de trombina, ativadores da fibrinólise, antagonistas de vitamina K (Katzung, 2006) e, recentemente, o *fondaparinux* (*Arixtra*), que é um análogo da heparina aprovado para uso clínico (Hirsh, 2003).

Apesar desses fármacos serem efetivos no tratamento e na redução de risco de doenças tromboembólicas, eles apresentam uma série de desvantagens que limitam seu uso e aceitação clínica, tais como administração parenteral de grande parte desses fármacos e interações com certos alimentos e outros fármacos, necessitando

monitoramento e ajuste de dose para obter o efeito desejado (Turpie, 2007). Por isso a indústria farmacêutica passou a buscar novos tipos de anticoagulantes, dentre os quais merecem destaque os inibidores diretos do FXa (Figura 5). O primeiro anti-Xa – DX 9065a – estava em fase clínica desde 1995, mas as pesquisas foram deslocadas em favor de outro agente da mesma classe – o Edoxaban. Atualmente existem cerca de 10 novos anticoagulantes, inibidores de FXa, em fase clínica, uns mais e outros menos avançados (Samama & Gerotziafas, 2010), tais como o Rivaroxaban, LY517717, Apixaban, DU176b, PRT054021, 813893 (Persborn, 2009).

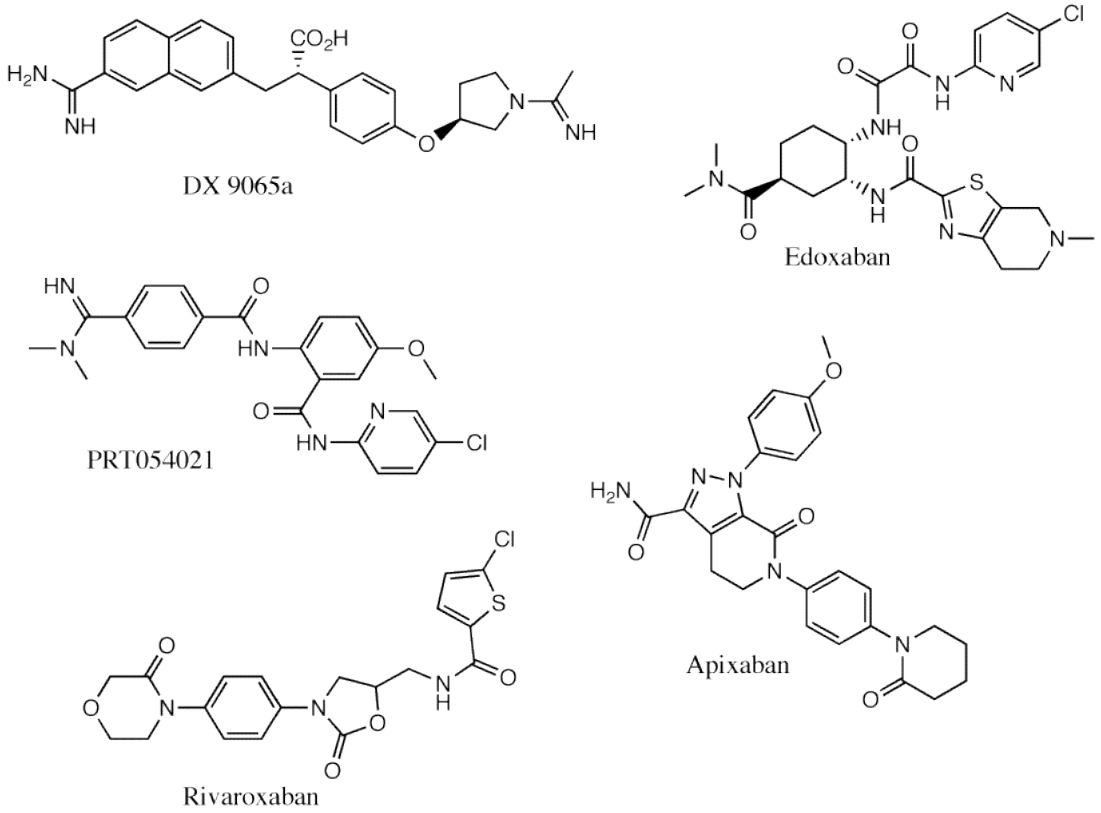



Figura 5: Estrutura bidimensional dos fármacos inibidores do FXa.

### *1.3. Importância da modelagem comparativa para descrição da estrutura tridimensional de proteínas*

O número de genes e genomas sequenciados vem crescendo progressivamente, alcançando um número 95 vezes superior ao número de estruturas determinadas experimentalmente (Ginalski, 2006). Um dos motivos que dificulta a obtenção de estruturas por cristalografia de raio-X é a flexibilidade de algumas proteínas e seu caráter transmembrânico, que dificultam a obtenção de cristais de boa qualidade para posteriores experimentos de difração. (Smyth & Martin, 2000). Tais

limitações vêm favorecendo o emprego de técnicas computacionais voltadas à predição de estruturas tridimensionais a partir de sequências de aminoácidos.

Apesar do progresso na predição de estrutura *de novo* (onde se obtém a estrutura terciária a partir da primária somente), a modelagem comparativa fornece modelos mais precisos e confiáveis, em menor tempo (Baker, 2001). Essa técnica consiste em 5 passos principais (Figura 6) que buscam comparar a sequência de aminoácidos de uma proteína de estrutura desconhecida – o alvo – com outras proteínas de estrutura conhecida, cuja sequência seja pelo menos 30% similar ao alvo – os moldes. Isso é possível porque a estrutura tridimensional é em torno de 3 vezes mais conservada do que a sequência, o que dá suporte à escolha de moldes de baixa identidade, quando não há nenhum de alta identidade (Martí-Renom, 2000).

Uma das principais potencialidades desta metodologia pode ser vista na sua capacidade de contribuir na análise das diferenças entre famílias de proteínas, identificação de resíduos catalíticos, interações proteína-proteína e proteína-ligante, design de protótipos a fármacos, correlacionar com dados de mutação, entre outros (Petrey & Honig, 2005), em situações onde não se conhece a estrutura tridimensional das proteínas em estudo a partir de métodos experimentais. Contudo, é necessário interpretar e validar os modelos obtidos primeiro, em termos de estrutura, estereoquímica, função, comportamento no ambiente. A estereoquímica das moléculas pode ser avaliada por programas como PROCHECK (Laskowski, 1998), que analisa o comprimento e os ângulos das ligações, a ligação peptídica, quiralidade, ângulos de torção da cadeia principal e lateral (Martí-Renom, 2000). Características como empacotamento, formação de um centro hidrofóbico, acessibilidade dos resíduos ao solvente, distribuição espacial de grupos carregados e volumes atômicos são critérios analisados por outros programas como VERIFY3D (Lüthy, 1992), ProsaII (Sippl, 1993) e ANOLEA (Melo, 1998), que comparam as propriedades de cada resíduo em um dado ambiente com dados prévios experimentais nesse ambiente, ou seja, avalia-se os perfis de energia a fim de detectar erros regionais (Martí-Renom, 2000). Além desses *softwares* específicos para validação, é possível utilizar outras abordagens, como *docking* de ligantes ou enzimas relacionadas com a função da molécula e, para refinar o modelo, pode-se fazer ainda simulações de dinâmica molecular (DM), mimetizando o ambiente no qual a proteína se encontra (Petrey & Honig, 2005).

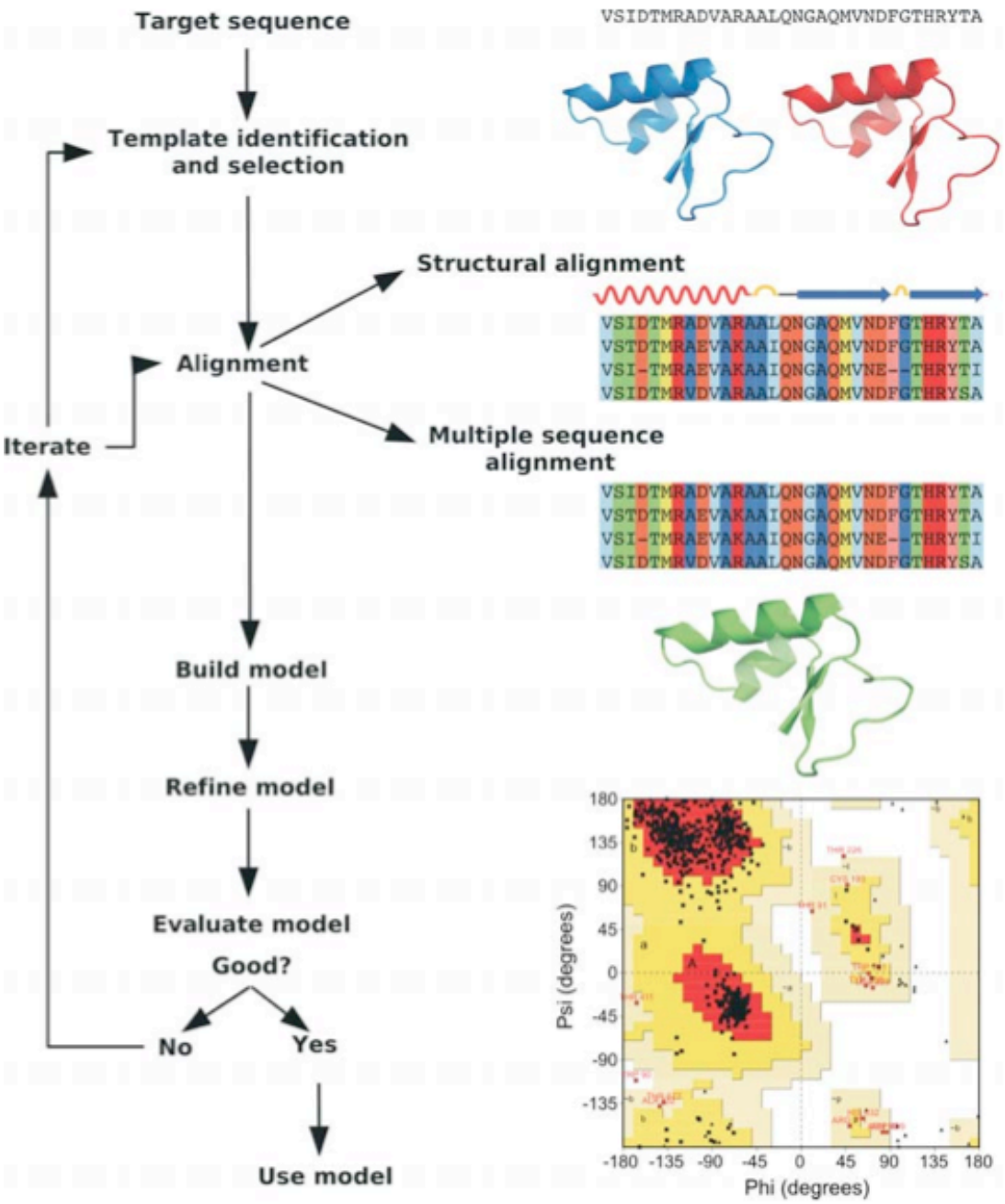



Figura 6: Representação esquemática da metodologia de modelagem comparativa (Bishop, 2008).

### *1.4. Estudos de docking na determinação de interações inter- e intra-moleculares*

Estudos sobre a interação de ligantes com diferentes tipos de proteínas, tanto para inibir quanto ativar, podem ser de extrema importância científico-tecnológica, tanto para a indústria farmacêutica quanto para a alimentícia, como no desenvolvimento de alimentos funcionais (Vaqué, 2008). A principal metodologia computacional empregada para estudar essas interações é o *docking*, também chamado de ancoramento molecular, que consiste em usar as coordenadas espaciais

do receptor e do ligante para predizer as coordenadas do complexo final. Apesar dos programas de *docking* utilizarem diferentes tipos de abordagens, há algumas características em comum entre eles, incluindo uma busca exaustiva das coordenadas possíveis das duas moléculas envolvidas para formação do complexo e o uso de uma função de hierarquização (*score*) para classificar os resultados obtidos, isto é, os conjuntos de orientações relativas entre receptor e ligante, de acordo com a melhor energia de interação intermolecular (Vaqué, 2008).

Em linhas gerais, os algoritmos de métodos de ancoramento molecular podem ser organizados segundo suas estratégias de busca conformacional, a saber:

1) Diretos ou sistemáticos:
    a) Busca conformacional;
    b) Fragmentação;
    c) Banco de dados.
2) Randômicos ou estocásticos:
    a) Monte Carlo;
    b) Algortimos genéticos;
    c) Busca *tabu*;
    d) Fast Fourier Transform

Em linhas gerais, há 2 tipos de algoritmos que permitem a busca do espaço conformacional do ligante: (1) diretos ou sistemáticos e (2) randômicos ou estocásticos. Os diretos possuem 3 subtipos de algoritmos, enquanto que os métodos estocásticos, 4. A Busca Conformacional tenta obter todas as possíveis conformações do ligante ao tratar suas ligações como flexíveis, podendo ser rotadas até 360º. Na Fragmentação pode-se utilizar 2 abordagens: a de colocar-e-juntar e a de incremento. Na primeira, o ligante é dividido em vários fragmentos rígidos que serão ancorados e após isso se tenta remontar o ligante, unindo esses fragmentos. Na outra abordagem, o ligante é dividido em diversos segmentos e um centro rígido, o qual é ancorado primeiro e após isso se une os outros segmentos. O Banco de Dados utiliza conformações semelhantes as encontradas em bibliotecas pré-geradas e que depois são submetidas a um *docking* rígido. (Vaqué, 2008).

No Monte Carlo, o ligante é randomicamente colocado no sítio de ligação do receptor e é gerado um *score*, então uma mudança aleatória na posição do ligante e nas ligações flexíveis forma uma nova conformação, gerando outro *score*. Se esse valor for melhor que o primeiro, essa conformação é aceita para entrar nos

resultados. Se não, uma função de probabilidade baseada em Boltzmann é aplicada a fim de avaliar se essa conformação será aceita ou não (Vaqué, 2008).

O Algoritmo Genético é baseado no modelo de evolução do Darwin, onde a população do ligante (genótipo) é criada aleatoriamente, contendo diferentes conformações (genes), e as coordenadas atômicas dos resultados (fenótipos) são analisados pela adaptação (*fitness*), ou seja, quanto mais adaptado, maior a chance de se ligar ao receptor. As orientações com melhor *fitness* serão utilizadas para gerar, randomicamente através operadores genéticos (mutações, permutas e migrações), outra população a ser testada e assim sucessivamente a fim de aprimorar os resultados (Wellock, 2001). A Busca *tabu* impõe restrições em áreas já exploradas do espaço conformacional do ligante, prevenindo que seja testada mais uma vez (Vaqué, 2008). Já o Fast Fourier Transform (FFT) foi introduzido em 1992 por Katchalski-Katzir e baseia-se na complementaridade de formas, isto é, o ligante e o receptor são considerados rígidos e as mudanças conformacionais são descritas por certos níveis de penetração inter-proteína (Wenfan, 2005).

### *1.5. Uso da DM como ferramenta para o estudo da conformação de biomoléculas*

A DM baseia-se na integração da equação de movimento de Newton, $F_i=m_i.a_i.$, onde a força "$F_i$" é aplicada a um átomo *i*, cuja massa é "$m_i$", acarretando em sua aceleração, "$a_i$", durante um intervalo de tempo, $\Delta t$ (Leach, 2001). Como isso ocorre em todos os átomos do sistema, é possível avaliar esses intervalos de tempo em sucessão e obter um conjunto de coordenadas tridimensionais que, dependendo do sistema, podem representar a conformação de biomoléculas em solução aquosa e a difusão de moléculas de água, dentre outros (Leach, 2001).

Além da equação de Newton, há um conjunto de funções de potencial, chamado de Campo de Força, que definem a energia de estiramento da ligação, a energia de distorção do ângulo de ligação, a posição e a distância dos átomos, termos de átomos ligados (ângulos, número de ligações, ângulos de diedros) e de átomos não ligados, entre outros (Scott, 1999). Alguns campos de força podem conter termos de interação entre átomos não ligados, de efeitos eletrostáticos, de ligação de hidrogênio e de outros efeitos estruturais (de Sant'Anna, 2002).

A adequada descrição do comportamento molecular por estas equações consiste em um dos principais fatores determinantes da confiabilidade das predições de dinâmica molecular (van Gunsteren & Berendsen, 1990). Contudo, é ainda

necessário avaliar o modelo produzido pela DM, comparando-o com propriedades experimentais, ou seja, se a DM reproduziu os dados de maneira adequada (Karplus & Petsko, 1990).

### *1.5.1. Simulação de Coarse-Grained (CG)*

Como visto acima, simulações de DM são ferramentas úteis para análise estrutural e interpretação de dados experimentais. Contudo, apesar do avanço no poder computacional disponível, as simulações de dinâmica ainda apresentam limitações importantes quanto ao número de átomos necessários ao estudo de sistemas biológicos (Monticelli, 2008). Processos celulares de duração acima de nanosegundos e que envolvem moléculas grandes ou a interação de várias moléculas necessitam de simplificações no sistema de forma a torná-los possíveis de serem tratados computacionalmente (Tozzini, 2010). Nesse contexto, foi desenvolvido o modelo denominado *coarse-grained* (CG), que apresenta simplificações capazes de aumentar sua velocidade e capacidade de lidar com problemas moleculares complexos, tais quais a formação e fusão de vesículas (Marrink, 2003), transformação de fase lamelar, estrutura e dinâmica de complexos membrana-proteína (Rzepiela, 2010), dinâmica de membranas (Risselada & Marrink, 2008), abertura e fechamento de canais de membrana (Treptow, 2008) e movimentos de domínios (Marrink, 2007), dentre outros. O principal campo de força para esse tipo de simulação é o MARTINI, tendo sido empregado com sucesso no estudo de diversos sistemas moleculares (Marrink, 2008).

O campo de força MARTINI é baseado em um *mapping* 4-para-1 (Figura 7), onde uma média de 4 átomos pesados são representados por um único centro de interação (*bead* ou esfera), a exceção de estruturas em anel, que são mapeadas em maior resolução, de 2-para-1 (Monticelli, 2008). Adicionalmente, só são considerados 4 tipos de sítios de interação: polar, não polar, apolar e carregado. Além disso, cada um desses tipos possui subtipos, permitindo uma representação mais precisa da natureza química, incluindo: capacidade de realizar ligação de hidrogênio (doador, aceptor, ambos ou nenhum) e graus de polaridade (de 1, com menor polaridade, a 5, com a maior) (Marrink, 2007). Por isso, como se pode ver na Figura 7, a cadeia lateral dos aminoácidos apolares, *i.e.* leucina, prolina, isoleucina, valina, cisteína e metionina, são representados por esferas apolares (azul ou ciano), assim como os aminoácidos polares sem carga, *i.e.* treonina, serina, asparagina e

glutamina, são representados por *beads* polares (amarelos ou laranjas). Resíduos carregados negativamente na cadeia lateral (glutamato e aspartato) são descritos por esferas carregadas (vermelhos), enquanto os carregados positivamente (arginina e lisina) são modelados pela combinação de partículas carregadas (vermelhas) e não carregadas (verde ou azul). Já os aminoácidos que possuem anéis, *i.e.* histidina, fenilalanina, tirosina e triptofano, são modelados com 3 ou 4 *beads* de uma partícula em anel de classe especial. A glicina e a alanina são os únicos resíduos que só são representados pela partícula da cadeia principal. Embora todas as esferas da cadeia principal estejam representada como polar na Figura 7, esta representação depende da estrutura secundária da sequência polipeptídica: aminoácidos livres em solução ou em alças possuirão partículas polares, porém nas hélices ou folhas-beta por esferas de polaridade intermediária, porque as ligações de hidrogênio entre as cadeias, que ocorrem nessas estruturas, reduzem a polaridade (Monticelli, 2008).

Esse tipo de simulação já foi utilizado em vários trabalhos para diversos sistemas, porém o CG apresenta algumas limitações na parametrização (Marrink, 2008). Um deles é que aplicação de CG em peptídeos e proteínas tem de ser feito com cuidado porque transformações de estrutura secundária não são parametrizadas, isto é, elas são fixas por uma função de potencial de energia do diedro, portanto processos em que o enovelamento ou o desenovelamento são essenciais não são recomendados para serem analisados por CG (Marrink, 2008).

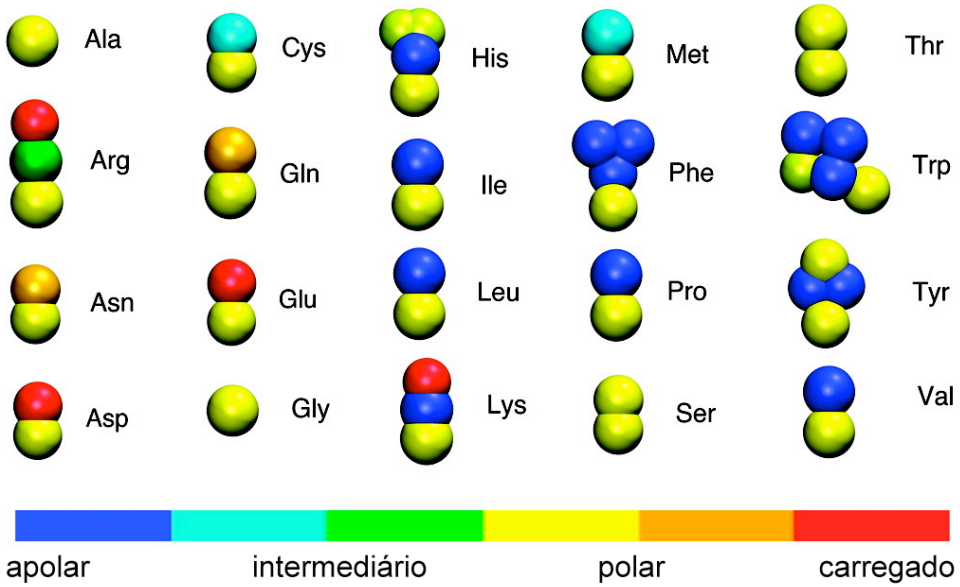



Figura 7: *Mapping* 4-para-1 dos 20 aminoácidos na simulação de CG, onde aproximadamente 4 átomos pesados são representados como um centro de interação (Ligabue-Braun, 2010).

## 2. Objetivos

A partir do exposto, o presente trabalho tem como objetivo a caracterização estrutural e conformacional da protrombina humana em solução aquosa, produzindo por fim um modelo teórico capaz de explicar e reproduzir dados experimentais prévios neste sistema. Para tal, o trabalho envolve alguns objetivos específicos, incluindo:

- Modelagem comparativa da protrombina por meio da similaridade de sua sequência com a de potenciais moldes;
- Ancoramento de regiões da proteína que não poderiam ser unidas devido a lacunas (*gaps*) sem estrutura tridimensional;
- Validação dos modelos obtidos;
- Uso de DM CG a fim de refinar os modelos obtidos;
- Análise da flexibilidade da proteína inteira e de suas regiões, bem como de possíveis movimentos interdomínios.

Uma proposta deste trabalho é que, com o modelo da protrombina humana completa, seja possível desenvolver análogos estruturais que interajam com o FXa, inibindo-o, e dessa maneira contornar as desvantagens dos outros anticoagulantes a fim de melhorar o tratamento e a qualidade de vida de pacientes com doenças tromboembólicas.

## 3. Trabalho experimental em formato de artigo científico

A revista escolhida para submissão do presente trabalho foi a *Proteins: Structure, Function and Bioinformatics*. Sob abreviatura *Proteins*, a revista teve seu primeiro ano de publicação em 1986 pela editora *Wiley*, em língua inglesa, sobre temas relacionados a pesquisa com proteínas, incluindo estrutura, função, interação com receptores, com ácidos nucléicos e com outros ligantes, computação, genética e design.

Esse periódico teve índice de impacto, em 2009, de 3,09 (Journal Citation Reports®, 2009). Suas publicações incluem artigos de pesquisa, notas, comunicações rápidas e revisões quando convidado.

De forma a ficar mais didático, as tabelas e as figuras 6, 7, 10 e 11 foram inseridas no corpo do texto, contudo, no manuscrito original, elas estão no material suplementar.

# Structural and conformational characterization of human prothrombin

Chiodi, C. G.[a], Fernandes, C. L.[a] and Verli, H.[a,b,*]

[a]*Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970, RS, Brazil*

[b]*Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil*

*Corresponding author. Tel.: +55-51-3316-7770; fax: +55-51-3316-7309; e-mail address: hverli@cbiot.ufrgs.br.

**Abstract**

Prothrombin is an important zymogen of the coagulation cascade since its active form, α-thrombin, is the key enzyme for converting soluble fibrinogen into insoluble fibrin monomers, which in turn organize themselves in a proteic clot that establish the platelet plug and, consequently, contribute to control the hemorrhagic processes. The prothrombinase complex (formed by factor Xa, factor Va, $Ca^{2+}$ and anionic phospholipid-containing membranes) is essential in prothrombin activation, in a process that culminates in the generation of α-thrombin and fragment 1.2. As substrate of an enzymatic reaction, prothrombin appears to be a very flexible molecule, a property that may be correlated with difficulties in obtaining its crystallographic structure. In this context, the current work employs comparative modeling techniques, docking calculations and molecular dynamics simulations in order to construct a theoretical model of the complete human prothrombin, capable to support the structural interpretation of its biological roles at the atomic level and so be potentially employed in further experiments to develop new antithrombotic agents. The obtained model points to the presence of hinge movements between prothrombin domains and, consequently, to the co-existence of multiple conformational states in biological solutions. Such flexible behavior of protrombin was also observed in intra-domain regions of the protein, as in its N-terminal region and in some exposed loops, which may be associated with its proteolysis susceptibility and specificity.

*Keywords:* Comparative Modeling; Molecular Dynamics; Docking; Coarse-Grained; Coagulation Cascade; Prothrombin.

## 1. Introduction

Prothrombin is a zymogen from coagulation cascade, also known as factor II, which converts soluble fibrinogen into insoluble fibrin monomers when activated. This occurs in the coagulation cascade due to the action of the prothrombinase complex, formed by factor Xa (FXa), factor Va (FVa), $Ca^{2+}$ and anionic phospholipid-containing membranes[1]. The prothrombinase complex cleaves prothrombin at Arg320-Ile321, generating meizothrombin, and Arg271-Thr272, producing α-thrombin and fragment 1.2[2].

The biosynthesis of prothrombin, before its secretion into the blood, includes glycosylation, removal of a prepro leader sequence in the liver and a vitamin K-dependent γ-carboxylation[3]. This prepro peptide contains the first 43 residues of prothrombin and is proposed to contain a recognition site for γ-carboxylation[4,5]. The final zymogen found in bloodstream contains 579 amino acid residues and 4 domains: γ-carboxyglutamic acid (GLA), kringle-1, kringle-2 and serine protease domains. The GLA domain contains 10 Glu residues modified by γ-carboxylation and plays a role in membrane binding[6]. Kringle domains are triple-looped structures, with highly conserved folding, constrained by three disulfide bridges[7] and involved in binding to FVa[8]. Finally, the serine protease domain comprises the catalytic machinery of the active thrombin. It is, therefore, the local where the prothrombinase complex, specifically FXa, acts, cleaving the connection between the domains. These four domains are didactically combined to form 2 fragments: fragment 1, composed by GLA and kringle-1 domains, and fragment 2, composed by kringle-2 and serine protease domains.

In terms of crystallographic studies, the first 3D structure of thrombin was described in 1992 and, until April 2010, approximately 300 structures were available

in the Protein Data Bank for the active form of prothrombin. However, only 7 structures were available for prothrombin fragments. Additionally, previous studies had modeled the tridimensional structure of the serine protease domain of porcine prothrombin based on human thrombin crystal[9] and the 4 domains of rhesus prothrombin separately, without considering the loops that connect them[10]. Accordingly, the absence of the complete tridimensional structure of prothrombin may be correlated with the fact that this protein is a substrate of an enzymatic reaction, thus it is supposed to be very flexible[11] and so increasing the difficulties in its crystallization.

Moreover, considering the role of thromboembolic diseases as a leading cause of morbidity and mortality[12], the anticoagulant market is projected to grow from around US$6 billion in 2008 to over US$9 billion in 2014[13]. In this process, one of the main targets in coagulation cascade for new antithrombotic agents appears to be FXa, which is expected to present an almost 9 fold increase in sales per year in this period[13]. As the substrate of FXa, the 3D structure of prothrombin may offer original scaffolds for developing new antithrombotic agents. In this context, the current work intends to obtain a tridimensional model of the human prothrombin in aqueous solution through strategies as comparative modeling, docking calculations and molecular dynamics simulations. The potential implications derived from such model are accordingly discussed.

## 2. Computational Methods

**2.1. Comparative modeling.** The amino acid sequence of human prothrombin (access code AAC63054 GI 339641[3]) was retrieved from the National Center of Biotechnology Information (NCBI). Homologue proteins were searched using

BLASTp tool[14] and the ones chosen were: PDB ID 2PF2[15], named $template_{44-187}$ and composed by $Ca^{2+}$ ions and the GLA domain of prothrombin fragment 1; PDB ID 1A0H[16], named $template_{207-363}$ and composed by the crystallographic complex between PPACK and meizothrombin; and PDB ID 1HAG[17], named $template_{328-622}$ and composed by isomorphous structures of prethrombin2, hirugen and PPACK-thrombin. There were regions without homologue crystallographic structures, as the prepro leader sequence, from amino acid residue 1 to 43, and a loop between $template_{44-187}$ and $template_{207-363}$ ($gap_{188-206}$), being the latter modeled (see below at item 2.3).

Based on a pair-wise alignment made with CLUSTAL W[18], the human prothrombin was modeled according to the tridimensional structure of the templates using the Swiss PDB Viewer (SPDBV) 3.7 program[19], and the obtained models were submitted to SWISS-MODEL server[20] for refinement, which generated the $model_{44-187}$, $model_{207-363}$ and $model_{328-622}$. The amino acid residues shared by $model_{207-363}$ and $model_{328-622}$ (Figure 1) were employed for alignment of the structures, resulting in a hybrid protein, named $model_{207-622}$.

**2.2. Docking.** As there is no structural information on the residues linking $model_{44-187}$ and $model_{207-363}$, 3 different docking strategies were employed, including AutoDock 3.0[21], HEX 5.0[22] and ZDock 3.0.1[23], in order to obtain a suitable orientation of both segments of prothrombin and thus support the modeling of $gap_{188-206}$,. Each program has a particular calculation approach; therefore orientations of the obtained models could be compared and used for cross validation. In AutoDock $model_{44-187}$ and $model_{207-622}$ were prepared as described previously[24]. The AutoGrid was used to generate grid maps of 0.527Å with 126 grid points for all coordinates in the N-terminal aminoacids of $model_{207-622}$. The Lamarckian Genetic Algorithm was

applied, with default parameters, except for the maximum number of generations, which was of 270,000, followed by 100 runs. Orientations of $model_{44\text{-}187}$ were selected based on the lower energy and most abundant orientations on a cluster of 0.5Å cutoff.

In HEX, dockings were performed in 3D Fast Lite Fast-Fourier Transform (FFT) mode with default parameters except for grid dimension of 1Å. The best 1000 orientations from N=16 (N is the order of Fourier expansions) scan phase were further refined by combining higher shape and electrostatic correlations at N=25. The $model_{44\text{-}187}$ orientations nearest to the N-terminal of $model_{207\text{-}622}$ and with the lowest energies were chosen.

In ZDock, $model_{44\text{-}187}$ and $model_{207\text{-}622}$ were prepared using the program default parameters, as described previously[25]. A rigid docking of 1000 runs was performed in the N-terminal site of $model_{207\text{-}622}$ using fast-fourier transform (FFT). Orientations were chosen based on higher scores and distance from $model_{207\text{-}622}$.

**2.3. Modeling gaps.** The $gap_{188\text{-}206}$ was predicted as a loop in PSIPRED[26] and PHYRE[27] servers. Thus, this region was built in the docking-derived models using SPDBV 4.0 and further submitted to ModLoop[28] for construction and refinement.

**2.4 Validation of the models.** In order to validate the obtained models three techniques were employed: 1) PROCHECK[29] stereochemical validation, 2) analyses of maintenance of kringle domains folding, and 3) docking with FXa proteolytic site (PDB ID 1EZQ[30]). Kringle domains have a highly conserved structure, hence they were superimposed using SPDBV 4.0 with other kringles, as PDB ID 1PMK[31] (kringle 4 of plasminogen), PDB ID 2KNF[4] (kringle 5 of plasminogen), PDB ID 1HPJ[32] (kringle 1 of plasminogen), PDB ID 1I71[33] (kringle 7 type IV of apolipoprotein) and PDB ID 2K4R[34] (neurotrypsin kringle). These superimpositions

were compared by means of root mean square deviation (RMSD) analysis. Additionally, as FXa is the main responsible for the cleavage of prothrombin, at Arg271-Thr272 and Arg320-Ile321 regions, docking studies between enzyme and zymogen were performed with ZDock, following the above described procedure, except for this time the aminoacids that were distant from the catalytic (residues 1-49, 71-89, 116-183 and 206- 244) and cleavage (residues 1-229 and 351-579) sites were blocked.

**2.5. Molecular dynamics simulations.** Each of the so obtained models for human prothrombin was submitted to Coarse-Grained (CG) molecular dynamics (MD) simulation under GROMACS 3.3 simulation suite[35] with Martini forcefield and water model of Marrink and co-workers[36,37]. An integration step of 30 fs after minimization using Steepest Decents algorithm was used in all CG-MD simulations, with coordinates saved every 150 ns for subsequent analysis up to 1.05 μs. Periodic boundary conditions were employed, and the Berendsen temperature and pressure coupling thermostats were used with time constants of 1 ps. Anisotropic pressure coupling was used, with the same compressibility (4.5 × 10-5) in all directions. Lennard-Jones interactions were shifted to zero between 0.9 and 1.2 nm, and electrostatics were shifted to zero between 0 to 1.2 nm, with a relative dielectric constant of 15.

## 3. Results and Discussion

### 3.1. Comparative Modeling

In the search for crystallographic homologues, $template_{44-187}$, $template_{207-363}$ and $template_{328-622}$ were chosen. As can be seen in the pair-wise alignment (Figure 1), each one has a high identity - 75%, 72% and 100%, respectively - with different parts

of prothrombin, and the three templates together have 81.5% of identity with the entire zymogen. Such high similarities may be expected to contribute for the model overall quality. The first 43 aminoacids without crystallographic structure correspond to the prepro leader sequence, which was not modeled, as it is not found in plasma prothrombin. Additionally, while the regions modeled based on $model_{207\text{-}363}$ and $model_{328\text{-}622}$ share a common region composed by 36 residues, supporting the building of $model_{207\text{-}622}$, the region corresponding to the $model_{44\text{-}187}$ has no structural reference to support a connection to the other templates to compose the final prothrombin model. As a consequence, these three templates supported the building of two initial models, that is, $model_{44\text{-}187}$ (based on $template_{44\text{-}187}$) and $model_{207\text{-}622}$ (based on $template_{207\text{-}363}$ and $template_{328\text{-}622}$). Accordingly, the stereochemical parameters of these two initial models were shown as adequate by PROCHECK analysis, in which the Ramachandran had more than 90% of the residues in allowed regions and the G-factor were less than -0.25, that is, the particular properties of the amino acid residues are not unusual. (Table I).

**Table I.** Results of stereochemical validation from PROCHECK.

| **Raw Models** | **Ramachandran** | | | | **G-factor** |
|---|---|---|---|---|---|
| | *Most Favored* | *Allowed* | *General* | *Disallowed* | *Total* |
| $Model_{44\text{-}187}$ | 76.4% | 19.7% | 3.1% | 0.8% | -0.16 |
| | *Total:* 96.1% | | | | |
| $Model_{207\text{-}622}$ | 67.1% | 29.7% | 2.3% | 0.8% | -0.26 |
| | *Total:* 96.8% | | | | |

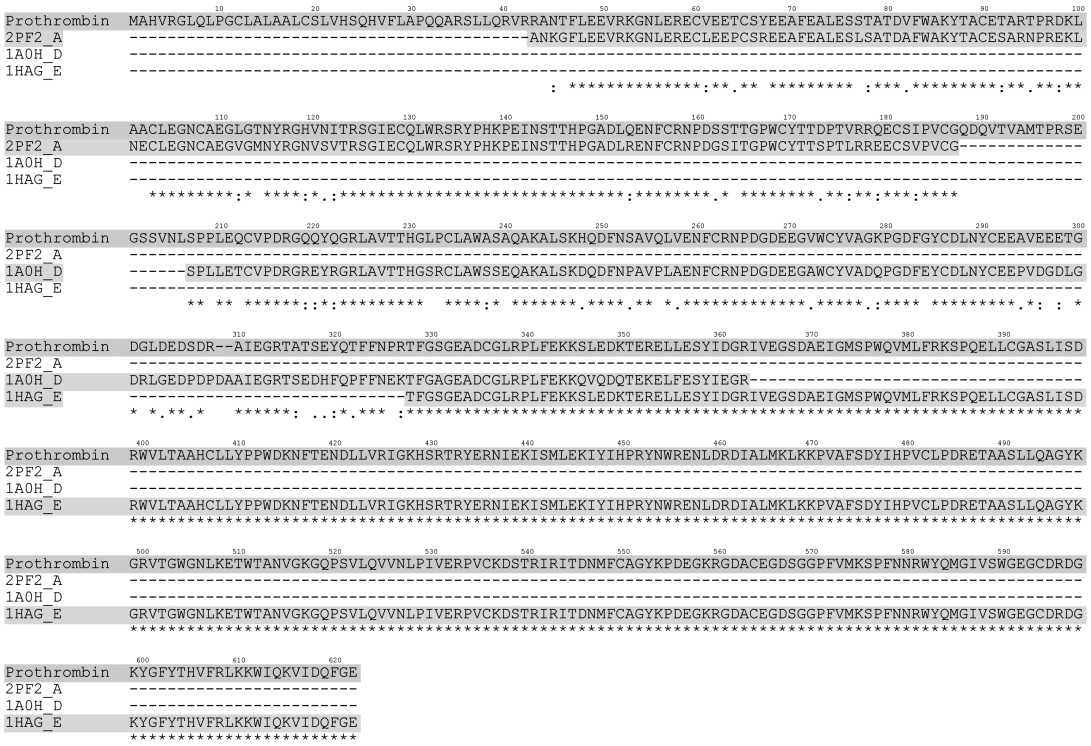



**Figure 1.** CLUSTALW pairwise alignment of $Template_{44\text{-}187}$, $Template_{207\text{-}363}$ and $Template_{328\text{-}622}$ with the target sequence. Templates sequences are shaded where they align for the purpose of highlighting the similarities among them.

Also, as kringle domains have a high conserved structure, such domains in the obtained models (kringle 1 is located between residues 107 and 186, and kringle 2 between residues 212 and 286) were superimposed all to all with previously determined kringle structures. The RMSD for kringle 1 was of 1.012 Å for 2KNF, 0.812 Å for 1PMK, 2.48 Å for 1HPJ, 4.137 Å for 2K4R, and 0.728 Å for 1I71. And for kringle 2, the obtained RMSD values were 1.940 Å, 1.708 Å, 3.513 Å 4.150 Å and 1.709 Å, respectively. Although kringle 1 and 2 had a high RMS with 1HPJ and 2K4R, the secondary structure was very similar, including the pattern of 3 dissulfide bridges and the β-sheets. Additionally, these data not only support the validation of prothrombin structure, but also partially contribute to its functional validation, since

the conformation and relative orientation of kringle domains are essential to factor Va binding.

### 3.2. Docking

Given the absence of further information concerning the relative orientation between these two segments of the whole protein, that is, between residues 44-187 and 207-622, we employed docking calculations in order to obtain a reliable proposal for the relative orientation between these two sets of residues. Accordingly, three docking programs with different approaches were employed, that is, Autodock, Hex and ZDock. In each algorithm the docked complexes were chosen based on the following criteria: 1) distance between the C-terminal of the $model_{44-187}$ and the N-terminal of the $model_{207-622}$ no longer than the length of the 19 missing aminoacid residues, that is, approximately 30 Å; 2) lowest energy or highest score, depending on the program; and 3) relative abundances. Therefore, among all results, four complexes were chosen: Model A from Autodock, Models B and C from Hex, and Model D from ZDock. Additionally, the $gap_{188-206}$ between $Model_{44-187}$ and $Model_{207-622}$ was constructed and modeled employing SPDBV and ModLoop server. This supported the obtainment of four models for human prothrombin, potentially representing different conformational states of the protein.

This finding appears to point to a high flexibility for this zymogen. In fact, as a zymogen, prothrombin must be activated, and the main physiological molecule responsible for that is FXa, which cleavages the protein in order to generate α-thrombin. Taking this in consideration, we attempted to validate the prothrombin models A-D through docking with FXa (PDBID: 1EZQ[30]) and so support a functional validation of the derived data. The obtained complexes between FXa and

prothrombin models$_{44\text{-}622}$ indicate that the catalytic site of the enzyme is facing both cleavage sites of the four zymogen models (Figure 2).

However, while the four models adequately represent the role of prothrombin as a zymogen of the coagulation cascade as FXa substrate, each docking program has produced slightly different orientations between Model$_{44\text{-}187}$ and Model$_{207\text{-}622}$. Instead of representing different results, these models appear to represent different paths in a hinge movement between prothrombin kringle domains (Figure 3). Such flexibility may correlate to difficulties in obtaining its crystallographic structure, until nowadays absent. So we considered as a further strategy for the refinement of these four docked complexes coarse-grained (CG) molecular dynamics (MD) simulations as capable of sampling large time scales and so describe complex hinge movements[38,39].

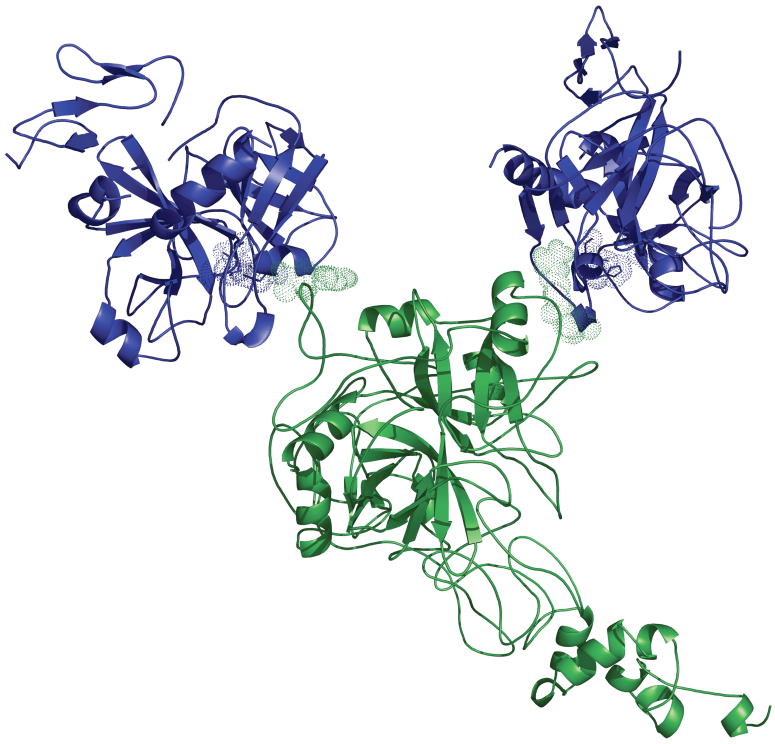


**Figure 2.** Docking of FXa (blue, PDBID: 1EZQ) to the whole prothrombin (green). The cleavage sites (green dots) go toward the catalytic sites (blue dots), validating the models.

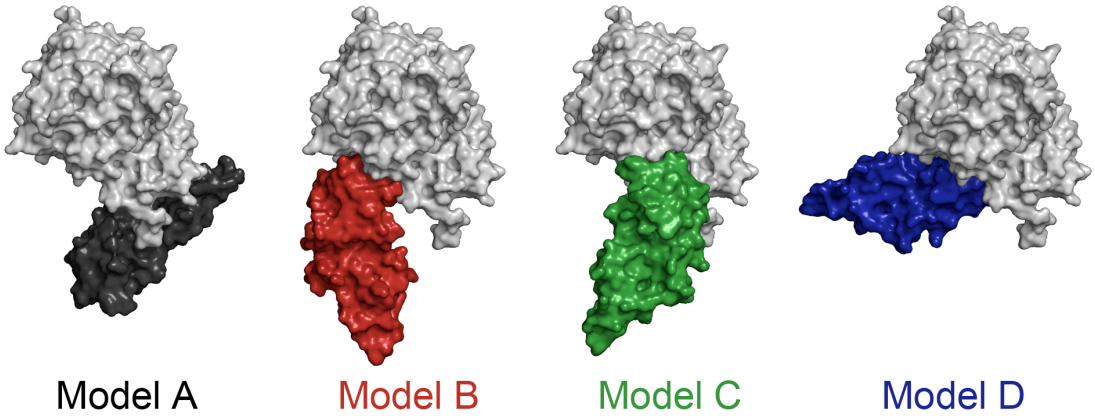



**Figure 3.** Models originated from docking. $Model_{207-622}$ (surface gray) is fixed to highlight the different orientations of the $model_{44-187}$ (blue, red, green and black) and the movement between them.

### 3.3. CG MD simulations

CG MD simulation is a methodology that simplifies the molecular system under study, withdrawing the effective friction caused by fine-grained degrees of freedom and so supporting the observation of large-scale motions during simulations with a larger time step[37]. Thus, CG is capable to labor with complex molecular systems, *e.g.* vesicle fusions[40], membrane dynamics[41], opening and closing of membrane channels[39], domain motions[36,38]. Accordingly, each of the prothrombin models obtained from docking calculations, that is, models A to D, was subsequently submitted to CG MD simulations.

During the simulations, the protein preserves its form, but becomes a little more globular with, approximately, 2.5 Å of radius of gyration (Figure 4). The 4 systems ended up equilibrated though they had high RMSD (Figure 4), which is normal in CG simulations.

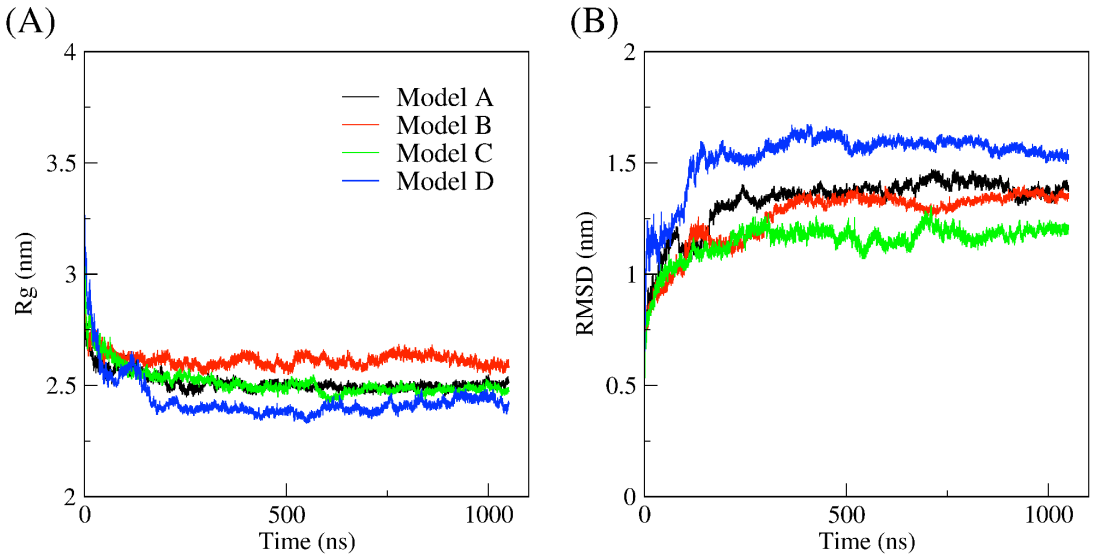



**Figure 4.** Prothrombin behavioral analyses during the simulations. (A) In the radius of gyration, it can be noticed that prothrombin became a little more globular. (B) In the RMSD graph, from 250 ns until the end of simulations, the systems became stable and do not undergo denaturation processes.

Model A becomes more globular in the middle of the structure, near kringles and serine protease, and the first domains described movements upwards and downwards, similar to a hinge. Model B appeared to be a little more rigid than model A, but presents some flexibility. Model C has a flexible N-terminal and a large number of hinge motions can be observed in its dynamics. Model D starts to become totally globular, however, after 500 ns, the first domains move a lot in the same hinge type of movement. Despite some differences in MD, all models presented high flexibility, which looked like hinge motions, primarily the N-terminal region that comprises the first 3 domains (see video in Supplementary Data). Since the C-terminal, which is the serine protease domain, correspond to the activate prothrombin form, thrombin, it was expected to be more rigid and stable. In order to support and quantify these findings, RMSD (Figure 5-A,B) and a root mean square fluctuation

(RMSF) (Figure 6-A) and RMSD analyses were made. The fluctuation of serine protease domain was very similar among the 4 models, however the N-terminal had diverse peaks and some were really different from the others. Hence the black peak in kringle 1, the region between kringles – corresponding to the loop modeled – and the blue peak between kringle 2 and serine protease were chosen in order to observe to which secondary structure they correspond, and as expected they were all loops (Figure 6-B). Interestingly, the residues that compound the loop that was modeled had different fluctuations in each simulation, fact that can be correlated with its role for the hinge motions.

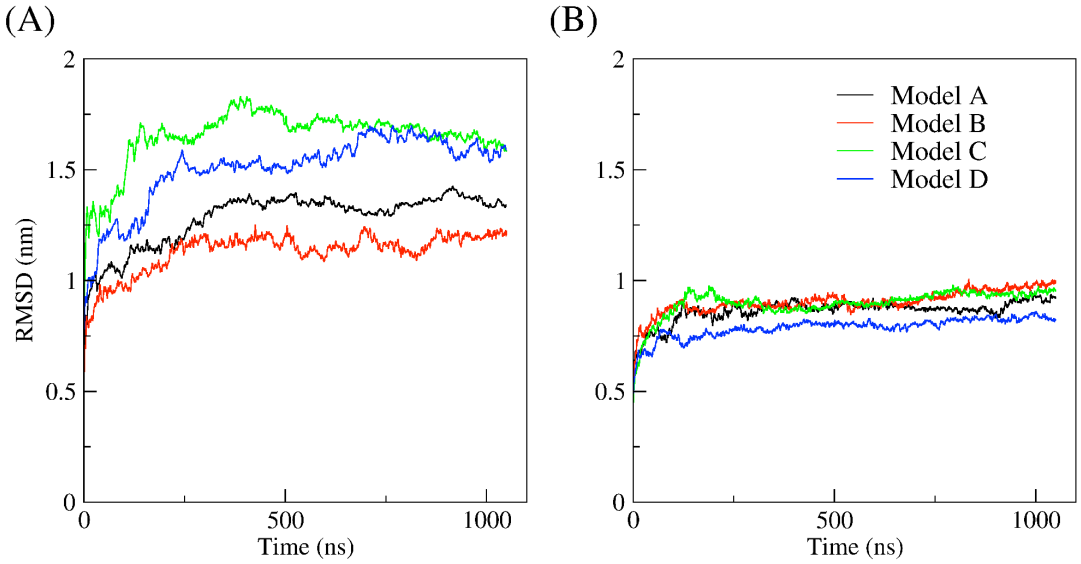
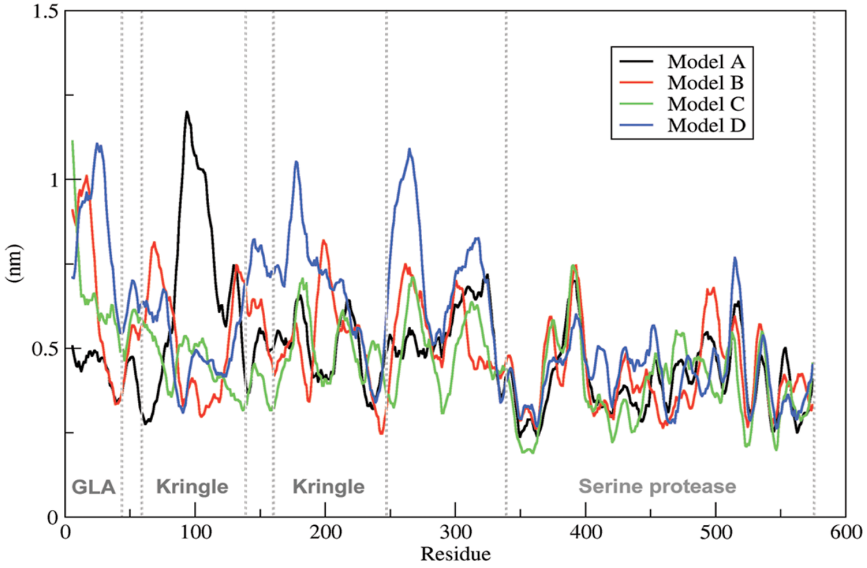



**Figure 5.** RMSD analysis of the N- (comprises GLA and kringles domains) and C-terminal (serine protease) in each model. (A) RMSD of the N-terminal, showing differences among each model in the simulation. The same does not occur in (B) the RMSD of the C-terminal, which is very similar in the 4 models.

(A)




(B)

**Figure 6.** Regions of prothrombin that have major flexibility. (A) RMSF performed for the 4 models in the whole simulations showed that the fluctuation of serine protease domain is much lower than the N-terminal region. (B) Some peaks of RMSF – black peak in kringle 1 (orange), residues between the kringles (blue) and between kringle 1 and serine protease (red) – were analyzed in prothrombin tridimensional structure and the 3 were loops, as expected due to flexibility.

In terms of hinge motions, the atomistic models A-D were submitted to HingeProt server[42] in order to evaluate this motion and determine which amino acids are implicated. For the 4 models, there were 2 kind of movements, however the residues involved were different, *e.g.* in model A, Ala294, Glu105, Gly187, Asp304, Ile447, Leu465, Glu532 and Phe549 determine the motions, whereas in model B they are Gly187, Asn115, Val185 and Glu296; model C has Tyr221, Leu104 and Leu303 involved; and model D has Gly187, Ala101 and Glu297. Interestingly, Gly187 was appointed for 3 models and this residue correspond to the first amino acid of the constructed loop. Indeed these motions can be seen in the trajectories and, guided by HingeProt results, we were able to analyze the angle distribution, as well as the average angle and the standard deviation of the residues involved in the hinge movements (Table II). Analyzing the standard deviations, Ala294, Glu296 and Gly187 are the amino acids that have larger motions, whereas Ala101, Leu104 and Glu105 have minor motions (Figure 7). Curiously, these residues represent interdomain loops. The latter, that have fewer deviations, connect GLA to kringle 1, Ala294 and Glu296 are between kringle 2 and serine protease, while Gly187 connect the kringles. These findings support the RMSF analysis, whose major fluctuation are in these interdomain loops, and can be correlated with the biological activity, since kringle 2 and serine protease have to be accommodated to interact with FVa and FXa, respectively, from prothrombinase complex.

**Table II.** Angle distribution and average of the residues presented in the HingeProt results for the 4 simulations.

| Residues | Model A | | Model B | | Model C | | Model D | |
|---|---|---|---|---|---|---|---|---|
| | Average | St. Dev. | Average | St. Dev. | Average | St. Dev. | Average | St. Dev. |
| 100-101-102 | 95.39º | 3.23º | 96.77º | 3.35º | 96.77º | 2.39º | 96.64º | 3.37º |
| 103-104-105 | 94.98º | 3.43º | 96.79º | 3.44º | 96.88º | 3.42º | 96.58º | 3.47º |
| 104-105-106 | 97.12º | 3.41º | 97.569º | 3.46º | 126.11º | 14.38º | 97.56º | 3.54º |
| 114-115-116 | 114.88º | 11.71º | 109.65º | 11.43º | 121.09º | 15.90º | 118.38º | 13.38º |
| 184-185-186 | 114.14º | 11.00º | 117.72º | 16.61º | 107.05º | 17.00º | 112.48º | 13.98º |
| 186-187-188 | 145.67º | 15.35º | 148.32º | 12.84º | 150.40º | 12.97º | 136.72º | 17.19º |
| 220-221-222 | 117.89º | 12.95º | 118.24º | 13.66º | 121.09º | 14.37º | 121.62º | 14.82º |
| 293-294-295 | 137.98º | 17.36º | 137.03º | 16.67º | 130.92º | 13.35º | 138.45º | 17.28º |
| 295-296-297 | 128.47º | 15.92º | 128.84º | 15.81º | 136.77º | 13.83º | 129.29º | 16.03º |
| 296-297-298 | 133.83º | 14.81º | 134.57º | 14.98º | 142.36º | 13.09º | 133.86º | 14.74º |
| 302-303-304 | 122.81º | 14.57º | 121.43º | 14.38º | 123.27º | 14.91º | 122.58º | 14.57º |
| 303-304-305 | 116.31º | 12.31º | 114.87º | 13.04º | 120.65º | 13.72º | 118.65º | 13.69º |
| 446-447-448 | 149.36º | 10.17º | 136.11º | 10.21º | 132.58º | 10.88º | 139.58º | 10.70º |
| 464-465-466 | 145.82º | 11.58º | 147.28º | 13.79º | 140.32º | 11.29º | 142.03º | 12.38º |
| 531-532-533 | 125.27º | 14.60º | 125.52º | 17.28º | 129.44º | 15.45º | 131.08º | 15.16º |
| 548-549-550 | 142.14º | 10.94º | 120.195º | 12.09º | 136.13º | 10.80º | 124.41º | 12.10º |

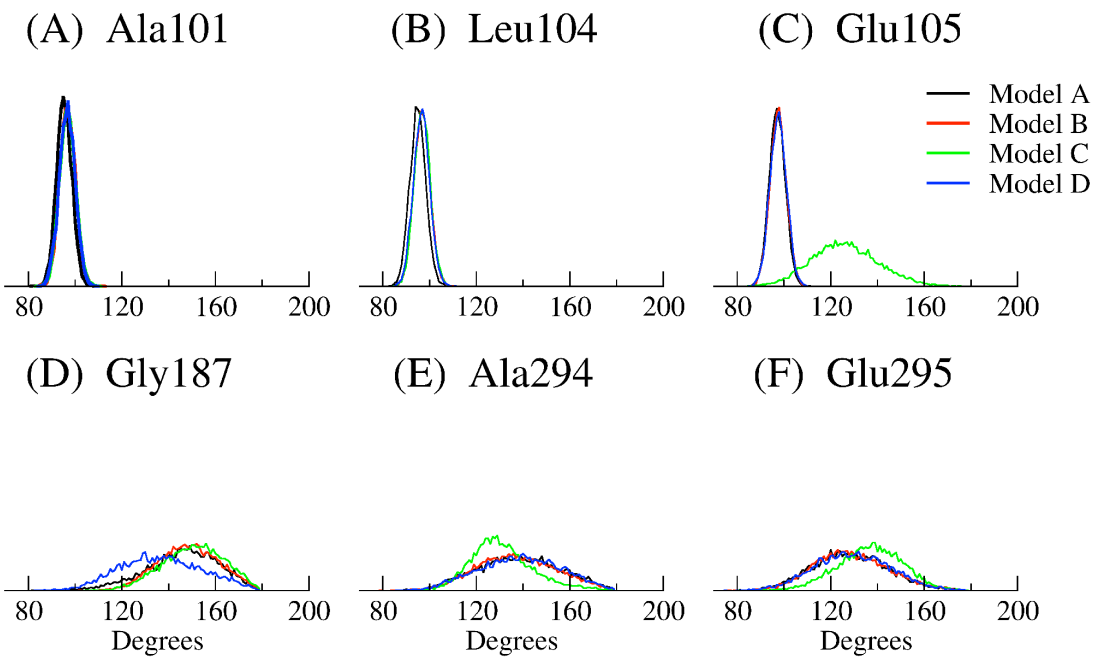



**Figure 7.** Angle distributions, during the 4 simulations, of the residues with major and minor standard deviations involved in hinge motions, which are (A) Ala294 (B) Glu105 (C) Gly187 (D) Ala101 (E) Leu104 (F) Glu296.

The high flexibility and the hinge motions shown in MD provided data to hypothesize that some interconversions between the models during the simulations may occur. Therefore GROMACS cluster analysis was employed in order to attest this hypothesis. Cluster size graphic (Figure 8-A) shows 3 different population states (State I, II and III) for the 4 models in the 4 trajectories, proposing that the protein is always rearranging but have 3 major conformations represented as the average of each cluster (Figure 8-A, structures above the clusters). In order to analyze the differences between the states, they were superimposed with each other's C-terminal (Figure 8-B), providing more subsistent results about the domain motions. Besides that, the angle of the same residues described above (those from the HingeProt results) was calculated for the states (Table III) in order to measure the distinctions among them. It was found that five groups of residues had more than 10º of

difference between some of the states. In Glu105, the state I have this segment widely open comparing to the others, while in Ala294 it is the narrowest. In Ile447, state III has the greatest width. State II has the minor angle in Phe549 and the major one in Leu303.

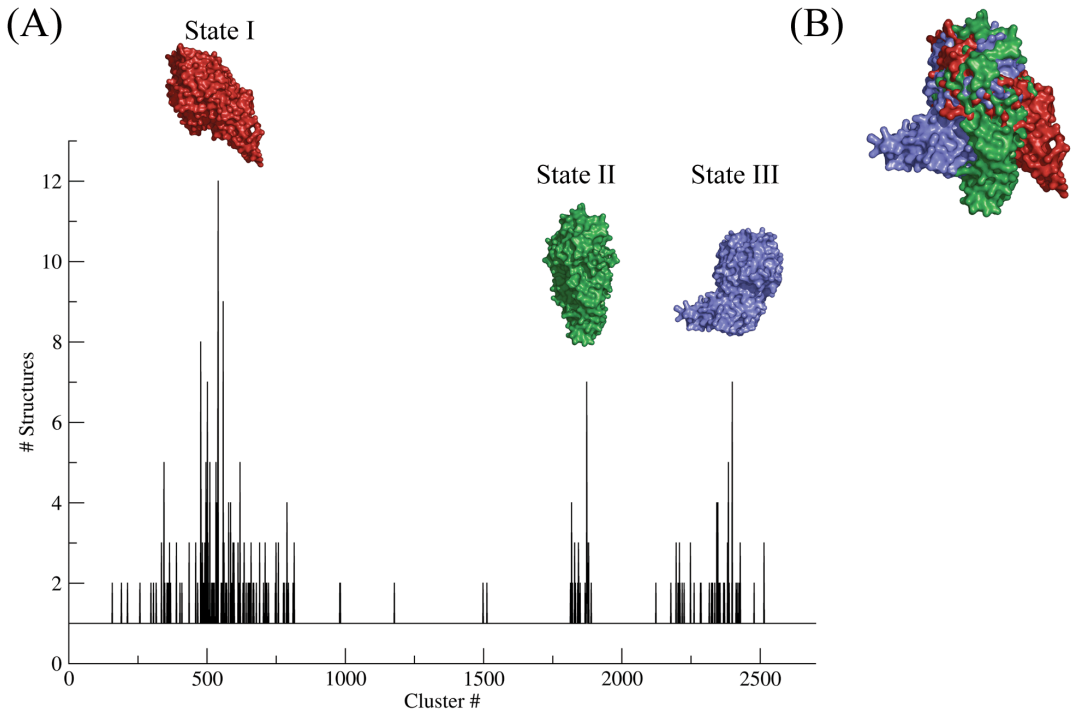



**Figure 8.** Structures of the 4 dynamics organized in clusters of 0.2nm. (A) There can be evidenced 3 different populational states with each representative conformation above the graphic. (B) Superimposition of the C-terminal portion of the 3 states, highlighting the differences between them and the domain motions of the N-terminal segment.

**Table III.** Angle distribution and average of the residues presented in the HingeProt results for the clusters.

| Residues | State I | State II | State III | Average |
|---|---|---|---|---|
| 100-101-102 | 98.14º | 98.23º | 95.37º | 97.25º |
| 103-104-105 | 97.32º | 98.03º | 94.51º | 96.62º |
| 104-105-106 | 143.39º | 94.81º | 97.82º | 112,01º |
| 114-115-116 | 111.26º | 108.63º | 102.80º | 107.56º |
| 184-185-186 | 117.21º | 106.76º | 115.71º | 113.23º |
| 186-187-188 | 167.24º | 168.67º | 158.40º | 164.77º |
| 220-221-222 | 114.27º | 123.90º | 113.72º | 117.30º |
| 293-294-295 | 130.39º | 157.70º | 163.77º | 150.62º |
| 295-296-297 | 154.91º | 146.04º | 143.85º | 148.27º |
| 296-297-298 | 151.51º | 130.84º | 141.60º | 141.32º |
| 302-303-304 | 125.61º | 142.16º | 125.66º | 131.14º |
| 303-304-305 | 114.28º | 128.37º | 115.04º | 119.23º |
| 446-447-448 | 126.22º | 128.08º | 146.24º | 133.51º |
| 464-465-466 | 142.01º | 143.49º | 145.10º | 143.53º |
| 531-532-533 | 138.11º | 151.93º | 123.15º | 137.73º |
| 548-549-550 | 151.48º | 130.64º | 152.08º | 144.73º |

In order to confirm and highlight punctual differences among the states, each of them were compared two by two in the contact maps generated by g_mdmat from GROMACS package (Figure 9). These maps show distinct interactions among the residues, for example the serine protease domain interacts with kringle 1 in State II, with kringle 2 in State I and with amino acids between GLA and kringle 1 and between kringles in State III. Besides these, State II is the only one that has GLA in contact with kringle 2. Some of these particular interactions were chosen and the distance between the respective residues were calculated during the 4 simulations

(Figure 10). As can be seen in the graphics, the distance between GLA and kringle 2 (Figure 10-A), domains, only exhibited in state III, has similar values in models B, C and D; therefore they may interconverge in dynamics. As model A has the lower distance, it is populating the state III. Additionally, the analysis of the distance kringle 1 and serine protease domains (Figure 10-B), whose interaction is only seen in state II, shows that state II comes from model B simulation and may come also from the first 200 ns of model C dynamics. Moreover, model A and D have similar distances, fact that strengthens the interconvergence hypothesis. Furthermore, the interaction of kringle 2 with serine protease, presented in state I, can be found in the models C and D, which have approximately 1.75 nm of distance, but not in models A and B, that have about 2.25 nm. Besides, the distances of some models decrease and increase over time, indicating the presence of hinge motions. These data support the hypothesis of hinge movement and interconversion of the models during simulation. Contact maps also showed that the interactions in serine protease domain were more constant than in N-terminal region – comprising GLA, kringle 1 and kringle 2 domains –; which were more variable, corroborating with the RMSF and RMSD described above.

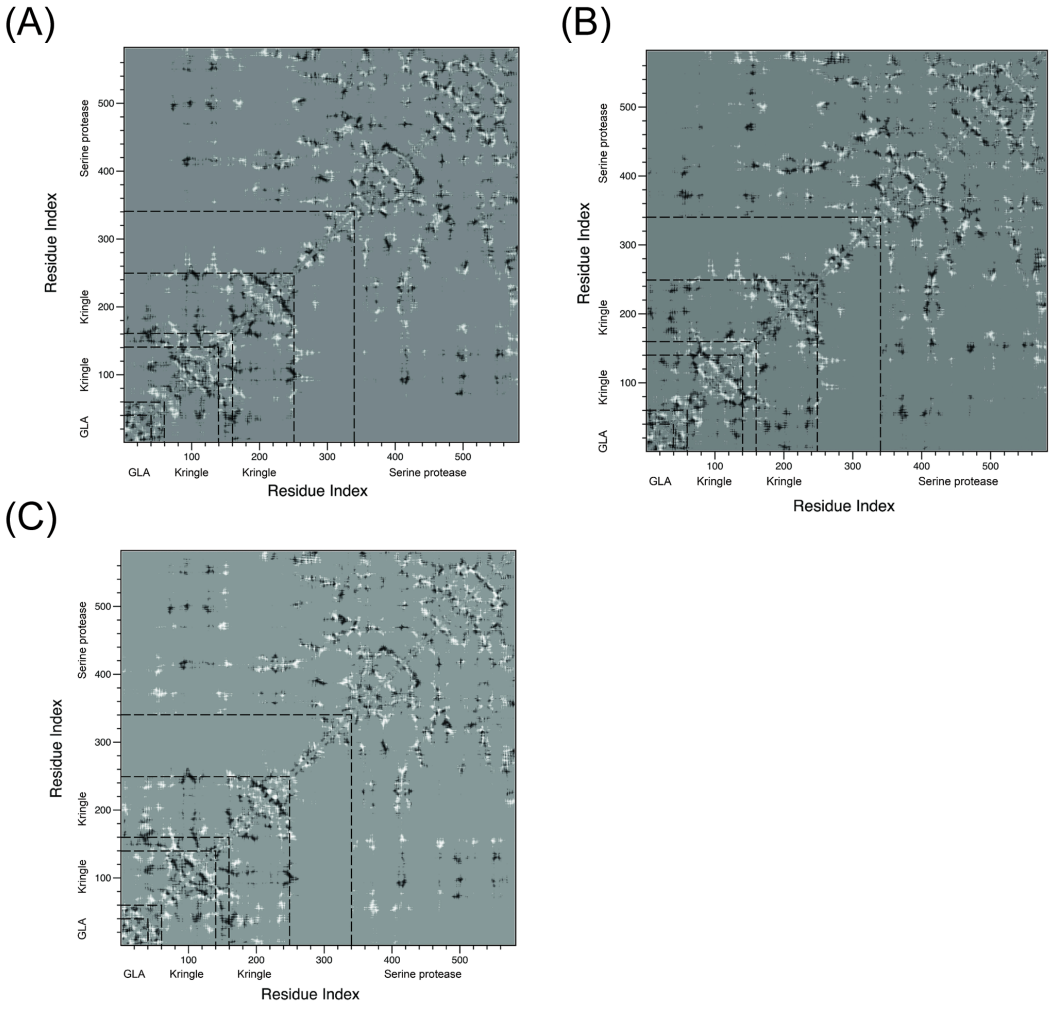



**Figure 9.** Inter-residue interactions analysis, demonstrating differences among the each state obtained from clusters. (A) Serine protease domain interacts differently in State I (white) and in State II (black), as can be seen the former interacts with kringle 2 and the latter with kringle-1. (B) Comparing State I (white) and State III (black), another different interactions with serine protease could be observed and this time the latter conformation has 2 regions – between the kringles and between GLA and kringle 1 – that are in contact with this domain. (C) In the comparison between State II (black) and State III (white), besides the differences observed previously, is important to highlight that only in State III kringle 2 interacts with GLA domain.

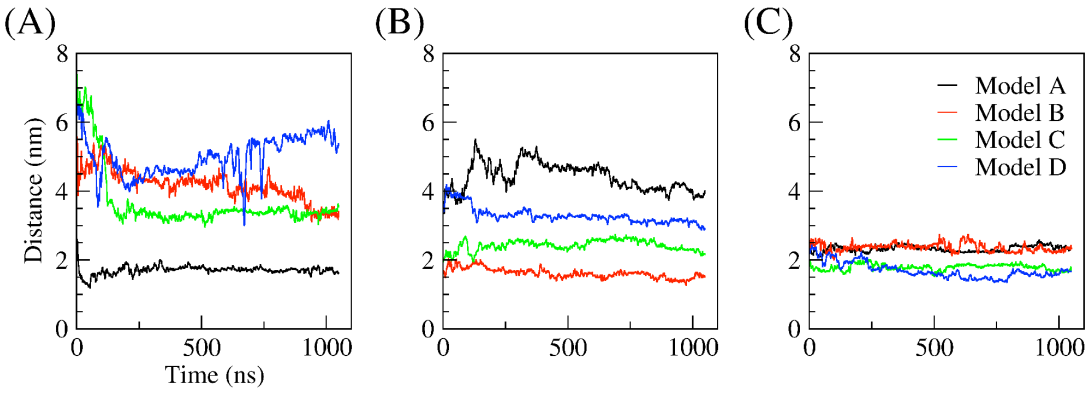



**Figure 10.** Distance between some residues chosen from Figure 9, reinforcing the interconversion of models in MD. (A) Distance between residues 8 to 15 and 195 to 205. The interaction of these residues – that comprises GLA and kringle 2, respectively – only was exhibited in state III. (B) Distance between residues 95 to 105 – kringle 1 – and 495 to 505 – serine protease – whose interaction is only seen in state II. (C) Distance between residues 190 to 200 and 415 to 425. The interaction of these residues – that comprises kringle 2 and serine protease, respectively – only was exhibited in state I.

These findings suggest that prothrombin is a transitional protein in solution, which has a major conformation in solution, characterized here as state I. The vast number of conformations is due to its hinge motions that are responsible for the major flexibility of N-terminal region comparing to C-terminal, fact that can be explained by its biological role, since the C-terminal is the local that comprises the catalytic machinery of thrombin. Therefore the CG MD simulations provide conformational data to characterize the entire human prothrombin and its motions, corroborating with the potential of this methodology to describe larger systems and time steps.

## 4. Conclusions

The lack of prothrombin structural information is mainly due to experimental difficulties with very flexible proteins, therefore this problem was overcame by CG MD simulations, a technique that enabled the structural and conformational characterization of this protein. The models presented here proved to be valid for prothrombin description at molecular level. State III and state II are structures equally prevalent in dynamics, for that reason they should be transient states of the protein in the human blood circulation, while state I is the most present conformation.

**Acknowledgments**

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq #420015/2005-1 and #472174/2007-0), MCT and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), MEC, Brasília, DF, Brazil.

**References**

1. Loscalzo J, Schafer AI. Thrombosis and Hemorrhage. Baltimore: Williams & Wilkins; 1998. p.
2. Krishnaswamy S *et al*. Activation of Human Prothrombin by Human Prothrombinase. J Biol Chem 1987;262:3291-3299.
3. Sanford DG *et al*. Structure of Prothrombin Containing the γ-Carboxylation Recognition Site Determined by Two-Dimensional NMR Spectroscopy. Biochem 1991;30:9835-9841.
4. Battistel MD *et al*. Solution Structure and Functional Characterization of Human Plasminogen Kringle 5. Biochemistry 2009;48:10208-10219

5. Esmon CT, Jackson CM. The Conversion of Prothrombin to Thrombin. J Biol Chem 1974;249:7791-7797.
6. Suttie *et al*. Vitamin K-dependent carboxylase: Possible role of the substrate "propeptide" as an intracellular recognition site. Proc Natl Acad Sci 1987;84:634-637.
7. Degen SJF, Davie EW. Nucleotide Sequence of the Gene for Human Prothrombin. Biochem 1987;26:6165-6177.
8. Huang M *et al*. Structural basis of membrane binding by Gla domains of vitamin K-dependents proteins. Nat Struct Biol 2003;10:751-756.
9. Chen Y *et al*. Full-lenght cDNA cloning and protein three-dimensional structure modeling of porcine prothrombin. Blood Cells, Molecules, and Diseases 2007;38:93-99.
10. Qin SF *et al*. cDNA Cloning, Protein Structure Modeling of Rhesus Monkey (*Macaca mulatta*) Prothrombin. Transplantation Proceedings 2008;40:603-606.
11. Smyth MS & Martin JHJ. X-Ray crystallography. Mol Path 2000;53:8-14.
12. Mackman, N. Triggers, targets and treatments for thrombosis. Nature 2008;451:914–918.
13. Melnikova, I. The anticoagulants market. Nature Reviews in Drug Discovery, 2009;8:353-354.
14. Altschul SF *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-3402.
15. Soriano-Garcia M *et al*. The $Ca^{2+}$ Ion and Membrane Binding Structure of the Gla Domain of Ca-Prothrombin Fragment 1. Biochem 1992;31:2554-2566.
16. Martin PD *et al*. New insights into the regulation of the blood clotting cascade derived from the X-ray crystal structure of bovine meizothrombin des F1 in

complex with PPACK. Structure 1997;5:1681-1693.

17. Vijayalakshmi J *et al*. The isomorphous structures of prethrombin2, hirugen-, and PPACK-thrombin: changes accompanying activation and exosite binding to thrombin. Protein Sci 1994;3:2254-2271.
18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673-4680.
19. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 1997;18:2714-2723.
20. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics 2006;22:195-201.
21. Morris GM *et al*. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J Computational Chemistry* 1998;**19**:1639-1662.
22. Mustard D, Ritchie DW. Docking Essential Dynamics Eigenstructures. PROTEINS: Struct. Funct. Bioinf. 2005;60:269-274.
23. Mintseris J *et al*. Integrating Statistical Pair Potentials into Protein Complex Prediction. PROTEINS: Struct. Funct. Bioinf. 2007;69:511-520.
24. Ruey R, Morris GM. Using AutoDock with AutoDockTools: A Tutorial. Scripps Research Institute, Molecular Graphics Laboratory, California, USA, 2006.
25. Chen R *et al*. ZDOCK Predictions for the CAPRI Challenge. PROTEINS: Struct., Funct. and Bioinf. 2003;52:68-73.

26. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000;16:404-405.

27. Kelley LA & Sternberg MJE. Proteins structure prediction on the web: a case study using Phyre server. Nature Protocols 2009;4:363-371.

28. Fiser A, Sali A. ModLoop: Automated modeling of loops in protein structures. Bioinformatics 2003;18:2500-2501.

29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26: 283-291.

30. Maignan S *et al*. Crystal structures of human factor Xa complexed with potent inhibitors. J Med Chem 2000;43:3226-3232.

31. Padmanabhan K *et al*. Kringle-kringle interactions in multimer kringle structures. Protein Sci. 1994;3:898-910.

32. Rejante RM & Llinas M. Solution structure of epsilon-aminohexanoic acid complex of human plasminogen kringle 1. Eur J Biochem 1994;221:939-949.

33. Ye Q *et al*. High-resolution crystal structure of apolipoprotein(a) kringle IV type 7: insights into ligand biding. Protein Sci 2001;10:1124-1129.

34. Ozhogina OA *et al*. NMR solution structure of the neurotrypsin Kringle domain. Biochemistry 2008;47:12290-12298.

35. van der Spoel D *et al*. J. Comput. Chem. 2005;*26*:1701–1718.

36. Marrink SJ *et al.* The MARTINI forcefield: coarse-grained model for biomolecular simulations. JPC-B 2007;111:7812-7824.

37. Monticelli *et al.* The MARTINI coarse-grained forcefield: extension to proteins. JCTC 2008;4:819-834.

38. Bahar I *et al.* Efficient Characterization of Colletive Motions and Interresidue

Correlations in Proteins by Low-Resolution Simulations. *Biochemistry*, **1997**;36:13512-13523.

39. Treptow W, Marrink SJ, Tarek M. Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations. *JPC-B*, **2008**;112:3277-3282.
40. Marrink, SJ, Mark, AE. The mechanism of vesicle fusion as revealed by molecular dynamics simulations. *JACS*, **2003**;125:11144-11145.
41. Risselada, HJ, Marrink, SJ. The molecular face of lipid rafts in model membranes. *PNAS*, **2008**;105:17367-17372.
42. Emekli U *et al.* HingeProt: Automated Prediction of Hinges in Protein Structures. *Proteins*, **2008**;70:1219-1227.

## 4. Discussão geral

Diversos tipos de metodologias estão disponíveis para determinação de estruturas tridimensionais de macro e micromoléculas, tais como cristalografia de raio X, RMN e modelagem molecular. Apesar dos avanços contínuos observado nas mesmas, diversos desafios se mantêm limitando o aumento do número de estruturas tridimensionais conhecidas. A técnica de RMN, por exemplo, é aplicada com sucesso em moléculas de até 12000 Da (Wüthrich, 1990), tendo dificuldades em compostos de massas maiores. Estas, em contrapartida, podem ser caracterizadas por cristalografia de raios-X, embora dificuldades na formação de cristais sejam frequentemente observadas em moléculas muito flexíveis (Faulon, 2002). Parte destes problemas pode ser contornado através do emprego de técnicas de modelagem molecular, permitindo a obtenção de modelos tridimensionais e posterior refinamento por dinâmica molecular na presença do solvente (ver item 1.3.). Nesse sentido, o presente trabalho identificou 3 moldes para a protrombina, suportando a obtenção de um modelo adequado para 3 regiões dessa proteína. A partir destes modelos iniciais, técnicas de *docking* foram empregadas na união destes 3 modelos. Como diferentes orientações foram encontradas para os domínios da protrombina dependendo do algoritmo de *docking* empregado, obteve-se 4 modelos iniciais da proteína de interesse. Cada um destes modelos foi devidamente validado de acordo com as características estereoquímicas pelo PROCHECK, com a estrutura dos domínios kringle altamente conservados e com a função dessa proteína na cascata de coagulação através da complexação da protrombina ao FXa por métodos de *docking*.

De forma complementar a estruturas tridimensionais rígidas oriundas de cristalografia de raios-X ou modelagem comparativa, o uso de simulações de dinâmica molecular vem contribuindo no entendimento do papel da flexibilidade molecular em processos fisiológicos (Karplus, 2002). Infelizmente, tais contribuições estão limitadas pela escala de tempo alcançada em simulações descrevendo moléculas de forma atomística, usualmente de até algumas centenas de nanosegundos. Com base nesses problemas, resolveu-se utilizar outro tipo de simulação capaz de demonstrar movimentos significativos em sistemas grandes ao alcançar escalas de tempo de micro e milisegundos, como é o caso da DM de CG.

O campo de força MARTINI de CG foi adequado para demonstrar a alta flexibilidade da protrombina, principalmente na sua região N-terminal, onde foram

observados movimentos interdomínios. Esta flexibilidade, aparentemente associada a movimentos do tipo "dobradiça" (*hinge movements*), levou-nos a identificar 3 estados conformacionais majoritários em solução, cujas principais diferenças residem na orientação dos resíduos da região N-terminal. No estado I, o mais prevalente das simulações, o domínio serino protease interage com o kringle 2 e possui maior ângulo no Glu105 e menor na Ala294, diferentemente do estado II, onde a interação do serino protease se dá com kringle 1 e o maior e menor ângulos são, respectivamente, Leu303 e Phe549. No estado III, a interação envolve os aminoácidos entre os kringle e entre o GLA e o kringle 1, e apresenta maior abertura na Ile447.

Além disso, utilizando a análise de distância dos estados durante as 4 trajetórias, pode-se notar que o estado I provém dos modelos$_{44\text{-}622}$ C e D, o II do B e o III do A. Portanto, os modelos obtidos da protrombina são válidos e descrevem seu comportamento flexível em solução aquosa, o que pode explicar as dificuldades na determinação de uma estrutura tridimensional completa para essa proteína.

## 5. Conclusões

A partir do exposto, o presente trabalho permitiu caracterizar conformacional- e estruturalmente a protrombina humana. Os resultados sugerem que essa proteína, devido à sua grande flexibilidade, destacadamente na porção N-terminal, tenha vários estados conformacionais em meio fisiológico, sendo o mais prevalente destes referente ao estado I. Adicionalmente, a alta flexibilidade encontrada é devido aos movimentos de dobradiça presentes na proteína.

Além disso, a rigidez da região C-terminal pode ser associada à sua função biológica, uma vez que ela é composta pelo domínio serino protease que irá compor a trombina.

Por fim, a metodologia apresentada nesse trabalho se mostrou adequada na construção de um modelo para a estrutura da protrombina. A partir do conjunto de validações empregado, espera-se que tais modelos possam ser utilizados em futuros trabalhos, principalmente no desenvolvimento racional de novos anticoagulantes inibidores de FXa por analogia estrutural a regiões da protrombina.

## 6. Anexo

- "Instruções para Autores" – Normas para publicação na revista *Proteins: Structure, Function and Bioinformatics*, em inglês.

### Types of Articles

*PROTEINS* accepts several different types of manuscripts:

RESEARCH ARTICLES
SHORT COMMUNICATION
REVIEW
RESEARCH COMMENTARY
STRUCTURE NOTE
PREDICTION REPORT (reserved for special issues on CASP and CAPRI meetings)

File Format Instructions for Online Submission Instructions for Manuscript Preparation Database Linking Criteria for Evaluating and Reporting Macromolecular Structures All PROTEINS manuscripts must be submitted online via ScholarOne Manuscripts at http://mc.manuscriptcentral.com/prot. **Please do not submit hard copies.** To find out if you have already created an account in ScholarOne Manuscripts, enter your e-mail address in the "Password Help" field. If you are submitting an article for the first time and/or do not have an existing account, carefully review the instructions posted under the heading "Resources" on the upper right hand side of log-in page; then click "Create Account" in the top right corner of the screen. At the end of a successful submission, a confirmation screen with a manuscript number will appear and you will receive an e-mail confirming the manuscript has been received by the journal. If you receive no confirmation, or if you experience technical difficulties during the submission process, click the "Get Help Now" link at the top-right corner of the log-in page to contact ScholarOne technical support. For general submission questions, contact Assistant Managing Editor Stefanie Alaimo at proteinsadmin@wiley.com or 201-748-6930

### *File Format Instructions for Online Submission*

Original submissions of Research Articles should be submitted as a **single PDF** document that includes text (i.e. title page, abstract, text, references), tables, figure legends, and figures. The manuscript should conform to the format described in the instructions for manuscript preparation, below.

All revised manuscripts and the first submission of all manuscripts that are not regular research articles (e.g. Structure Notes, Short Communication, Research Commentary, Reviews) must be submitted with individual files, one each, for text, tables and for each figure. Carefully follow the instructions below to ensure that, if your manuscript is accepted, it will be published as quickly as possible. The manuscripts must also adhere to the format described in the instructions for manuscript preparation, below.

Any article materials that are to appear in print must be uploaded under the category "Manuscript Files for Review" in ScholarOne Manuscripts. **Text** 1. Submit your

text file in .DOC or .RTF format. 2. Do not embed figures or tables in this document. 3. Designate this file as "Main Document" in the File Type field when you upload it to ScholarOne Manuscripts. 4. References and Figure Legends should be included in this document.

**Tables**

- 1. Create tables as text files and save in either .DOC or .RTF format. 2. Do not embed tables in your manuscript text file. 3. Designate this file as "Table" in the File Type field when you upload it to ScholarOne Manuscripts.

**Figures**

1. Submit your figure files in .TIFF or .EPS format. Authors are encouraged to use the file tags and figure legend options when uploading image files in order to take advantage of the linked HTML features of ManuscriptCentral. All figure panels (example: Figure 1A, 1B, 1C, etc.) must be submitted as a single file. 2. To ensure that your digital graphics are suitable for print purposes, please go to **RapidInspector** at http://rapidinspector.cadmus.com/zwi/index.jsp. This free, stand-alone software application will help you to inspect and verify illustrations right on your computer. 3. Vector-based figures (e.g., figures created in Adobe Illustrator) should be submitted in EPS format. 4. The following figure formats are unacceptable: JPG, GIF, PSD, CRD, PCT, PPT, PDF, XLS, DOC, BMP, 123 (or other Lotus Formats). 5. To ensure ease of legibility for reviewers as well as the highest print quality, TIFF and EPS files **must be submitted according to the following minimum resolutions:**

a. 1200 dpi (dots per inch) for black and white line art (simple bar graphs, charts, etc.) b. 300 dpi for halftones (black and white photographs) c. 600 dpi for combination halftones (photographs that also contain line art such as labeling or thin lines)

6. In addition to the above resolution guidelines, color figures must be submitted in CMYK colorspace.

a. Do not submit color figures as RGB. This color space does not reproduce well for print production.

7. Designate this file as “Image” in the File Type field when you upload it to ScholarOne Manuscripts.

**Instructions for Submitting LaTeX Files**

1. Submit a single, complete PDF file of your manuscript designated as “Main Document” in the File Type field on ScholarOne Manuscripts. 2. All source files must be uploaded as “Supplementary Files Not for Review.”

a. Submit your source file in LaTeX format as a single file. b. Submit figure and table files separately. Figures files must be named according to figure number. c. Follow resolution and formatting guidelines for figures as indicated above. d. Tables should saved in either .DOC or .RTF format. e. Do not embed tables in the text file. f. Bibliography Style: Use \cite{} to cite \bibitem{} listed in the reference list.

## Supplementary Online Material

Authors may submit supplementary material for their articles to be posted in the electronic version of the journal. These materials must be submitted online as "Supplementary Files for Review."

Supplementary text, tables and figures must be in PDF format. The only exceptions are audio and moving image files, which should be submitted in .mov or .mp3 format.

### *Instructions for Manuscript Preparation*

If your manuscript is a Structure Note, please see the special guidelines for structure notes.

**Manuscripts**

1. When submitting via ManuscriptCentral, authors **must** provide the names, e-mail addresses, and institutions (affiliation) of four possible reviewers and at least one preferred Editor. Options are also provided for designating names of persons that authors prefer not to review or edit their paper. The Editorial Office will refrain from assigning any person listed as "non-preferred". However, authors should note that preferred Editors and Reviewers may not be available nor will these persons necessarily be assigned in the peer-review process. 2. Authors may also include a cover letter to the Editor-in-Chief with the manuscript submission. This may be entered as plain text in the field provided, or uploaded as a .DOC, .RTF, or .PDF file. 3. All manuscripts must be be in 12-point Times New Roman font, in a single column and double spaced.

a. Authors in Japan may ask Wiley-Japan for a list of recommended services for checking and improving English. Please contact Yoko Kobayashi or A. Bocquet in the Wiley-Japan office by facsimile: 81 3-5689-7276 or by E-mail: editorial@wiley.co.jp for more information. Please indicate the name of the journal clearly.

4. Number all pages in sequence, beginning with the title page. 5. Follow the guidelines in CBE Style Manual Committee. CBE style manual: a guide for authors, editors, and publishers in the biological sciences. 5th ed. rev. and expanded. Bethesda, MD: Council of Biology Editors, Inc.; 1983.

Note that formats for Reviews and Research Commentary articles are flexible. Please consult with the Editorial Office before submission. The text for all other manuscript types should conform to the following format:

Manuscripts that do not conform to the following format will be returned. The parts of the manuscript must appear in the following order.

**Title Page**

The title page must contain the following information

1. The full title of the manuscript. 2. A short title of not more than 45 characters (including spaces). 3. Five to ten key words not used in the title that will adequately index the subject matter of the article. 4. The names and affiliations of all authors. 5.

The institution at which the work was performed. 6. Complete contact information for the author responsible for correspondence:

a. Name b. Address c. Phone/Fax numbers d. E-mail address

**Abstract**

This should summarize the purpose, methods, results, and major conclusions of the work (250 word limit).

**Introduction, Materials and Methods, Results, Discussion, and Conclusion**

1. The sections should be presented in the following order: Introduction, Materials and Methods, Results, Discussion, and Conclusion (if applicable). The Results and Discussion section can be combined into a single section, if convenient. 2. The manuscript should conform to standard scientific reporting style. 3. Sufficient data and information must be given so that the study can be replicated.

**References**

1. Cite references to published literature in the text numerically. 2. Provide full titles and complete page numbers for all works cited in the reference section. 3. Refer to the CBE style manual for the style of reference (see below for more information).

**Tables**

1.Indicate placement of all tables and illustrations in the text with a citation, i.e., (Table I). When uploading tables and figures, use the same tags for file names in order to utilize HTML linking in the manuscript to these elements. 2.Tables must be numbered in order of appearance with Roman numerals; illustrations with Arabic numerals.

a. A legend must accompany each illustration, and tables must have titles. b. All abbreviations used must be defined. c. All lettering must meet professional standards and be legible after reduction in size.

**Figure Legends**

1. Figure legends should be as clear and concise as possible.

**Figures**

1. Indicate placement of all illustrations in the text with a citation, i.e., (Fig.1). When uploading tables and figures, use the same tags for file names in order to utilize HTML linking in the manuscript to these elements. 2. Label the figures, i.e., (FIGURE 1). 3. For figures with multiple parts, the figures must be submitted in the assembled form in which they would appear in the published paper. 4. Figures must be submitted in the size (or close to it) in which they will appear in print. 5. Minimize the number of figures.

**Color Figure Preparation**

1. For best reproduction, bright, clear colors should be used. 2. Dark colors against a

dark background do not reproduce well; please place your color images against a white background wherever possible. 3. Unless otherwise arranged with the Editor, stereo pairs will be printed to a scale that yields a separation of 55-60 mm between corresponding points in the left and right images. Parallel-viewing pairs are preferred. Authors who choose to submit cross-eyed pairs must specify this feature in the legend. 4. Follow resolution guidelines listed above under "File Format Instruction for Online Submission."

**Color Reproduction Charges**

The color charge is $250 per figure.

**Cover Illustration**

Authors may nominate figures (or portions thereof) from their manuscript to be used as a cover illustration for the print issue of PROTEINS. Include a request in your cover letter and contact the Editorial Office.

***Database Linking***

For papers describing structures of biological macromolecules, the atomic coordinates and the related experimental data (structure factor amplitudes/intensities and/or NMR restraints) must be deposited at a member site of the Worldwide Protein Data Bank (www.wwpdb.org): RCSB PDB (www.pdb.org), PDBe (www.ebi.ac.uk/pdbe), PDBj (www.pdbj.org), or BMRB (www.bmrb.wisc.edu). The PDB ID should be included in the manuscript. Authors must agree to release the atomic coordinates and experimental data when the associated article is published. Questions relating to depositions should be sent to deposit@wwpdb.org.

Authors are encouraged to submit genetic and protein database information with their manuscript for the databases listed below and a hypertext link will appear in the online version of the article, via Wiley InterScience at www.interscience.wiley.com . The Genome Database (GDB) Protein Databank (PDB) Genbank Online Mendelian Inheritance in Man (OMIM) Molecular Modeling Database (MMDB) Entrez Genomes Entrez Proteins European Molecular Biology Laboratory (EMBL) SpecInfo ExPasy SWISS-PROT

1. To create hypertext links, authors must supply the gene name as it appears in the article, the database where the record appears, and the database specific identification number or name. 2. Please follow the instructions in the Database Linking Submittal Form, and submit a copy of that form with your manuscript. 3. It is the responsibility of the author(s) to ensure that the database information that is provided with the manuscript is correct and up to date. 4. The publisher will not submit new information to the databases. Incorrect information will result in the omission of hypertext links in the article. 5. For those articles containing gene and protein sequence information with a corresponding database record (see list of databases) hyperlinked database queries will be added to the online version for the full-text HTML version of the journal. 6. The hypertext links will appear in the Special Content Links section of the Abstract page, the text of the abstract and throughout the full text of the article.

***Wiley's Journal Styles Are Now in EndNote***

EndNote is a software product that we recommend to our journal authors to help simplify and streamline the research process. This feature also provides reviewers hyperlinked access to references in the HTML version of manuscripts. Using EndNote's bibliographic management tools, you can search bibliographic databases, build and organize your reference collection, and then instantly output your bibliography in any Wiley journal style.

Download Reference Style for this Journal: If you already use EndNote, you can download the reference style for this journal.

How to Order: To learn more about EndNote, or to purchase your own copy, visit www.endnote.com. Technical Support: If you need assistance using EndNote, contact endnote@isiresearchsoft.com, or visit www.endnote.com/support. Journal:

1. King VM, Armstrong DM, Apps R, Trott JR. Numerical aspects of pontine, lateral reticular, and inferior olivary projections to two paravermal cortical zones of the cat cerebellum. J Comp Neurol 1998;390:537-551. Book: 2. Voet D, Voet JG. Biochemistry. New York: John Wiley & Sons; 1990. 1223 p. Book Chapter: 3. Gilmor ML, Rouse ST, Heilman CJ, Nash NR, Levey AI. Receptor fusion proteins and analysis. In: Ariano MA, editor. Receptor localization. New York: Wiley-Liss; 1998. p 75-90. Electronic Media: 4. Bio-Xplor, Version 1.0. New York: Biostructure Inc.; 1991.

### *Electronic Proofing*

In order to expedite the publication and online posting of articles in Wiley InterScience, *PROTEINS: Structure, Function, and Bioinformatics* offers electronic proofing. Corresponding authors with e-mail addresses will be sent page proofs(and paperwork, such as reprint order forms)in .pdf format via e-mail. Please follow the instructions in the e-mail;contact names and numbers are given for questions and problems.

### *Accepted PrePrints*

Preprints of accepted manuscripts will be assigned a DOI number and posted to Wiley Interscience within a week of being sent out for production. Once the manuscript has been proofed and typeset, the preprint will be removed from Interscience and replaced with the most current version in EarlyView.

### *Criteria for Evaluating and Reporting Macromolecular Structures*

Below are guidelines for helping to assess the reliability and completeness of macromolecular structural reports.

A good summary of structure validation can be found in the article by Gerard Kleywegt, Acta Cryst. (2000) D 56, 249-265.

For all crystallographic studies, coordinates and structure factors should be deposited in the Protein Data Bank at the time of manuscript submission. The PDB id(s) should be included in the text. These data must be released upon publication of the accepted article (available on-line in EarlyView). Authors should note that the PROTEINS

Editorial Board encourages release of coordinates and structure factors at the time of PDB deposition.

For refined structures, the data submitted will depend somewhat upon the effective resolution of the analysis. In most cases, it would be helpful to include the following:

1. Effective resolution should be described clearly by: i) giving the percentage of data recorded in the various resolution ranges; ii) the internal agreement of the data ($R_{merge}$) in each of these ranges; iii) the percentage of data considered observable in each of the resolution ranges, along with the specific criteria that were used to select observable data; and iv) the average I/σ(I) value for each range. This information should be summarized in the manuscript, and the complete statistics provided as a table in the supplementary material.
2. The crystallographic R-index (Σ||Fobs|-|Fcalc||/Σ|Fobs|) should be tabulated as function of resolution. For each shell, the number and fraction of reflections observed should be given, as well as the number actually used in the calculation of the R-index. Any selection criteria should also be provided. The R-index should be computed with and without solvent included. The corresponding $R_{free}$ should be included along with the criteria used to select the test set. This information should be summarized in the manuscript, and the complete statistics provided as a table in the supplementary material.
3. A final Ramachandran plot or related conformational analysis should be provided for the reviewers with the favorable energy regions indicated. Distributions of χ$^{1}$, χ$^{2}$ values for sidechains might be presented, along with an analysis of deviations from expected minimum energy ranges. An indication of the range of the peptide angle ω should also be given. Unusual conformational features should be described in the manuscript.
4. The methods section should provide the following: i) adequate details regarding the steps followed in constructing the model and refining the structure; ii) the number of solvent atoms should be given, along with information regarding solvent B-values, and the approach that was used to identify solvent sites; iii) the history and salient details of the refinement methods employed, including the resolution ranges that were used at various stages of refinement; iv) the restraints used, including whether or not van der Waals' distances were restrained; v) a description of how the thermal parameters were treated; and vi) how solvent sites were selected and handled during refinement.
5. Hydrogen bonding patterns within the protein should be described. The number of hydrogen bond donors that are not involved in hydrogen bonding should be given, with particular attention to unsatisfied buried mainchain H-bonds. Deviations of bond lengths, bond angles, and planes from ideal geometries should be given. Close intramolecular and intermolecular van der Waals' contacts should be described.
6. Any structural features that are considered somewhat unusual should be described. Examples include: cis-peptide bonds, unoccupied volume inside the protein, buried charged groups that are not involved in salt bridges or reasonable hydrogen-bonding environments, unusual locations of glycine and proline residues, and unusual distributions of polar and hydrophobic groups within the molecule.

### *Disclaimers, Rights and Responsibilities*

All Manuscripts submitted to PROTEINS: Structure, Function, and Bioinformatics must be submitted solely to this journal, and may not have been published in another

publication of any type, professional or lay. Upon acceptance of a manuscript for publication, the author will be requested to sign an agreement transferring copyright to the publisher, who reserves copyright. No published material may be reproduced or published elsewhere without the written permission of the publisher and the author. All statements in, or omissions from, published manuscripts are the responsibility of the authors, who will assist the editor by reviewing proofs before publication. Reprint order forms will be sent with the proofs.

**Note to NIH Grantees** Pursuant to NIH mandate, Wiley-Blackwell will post the accepted version of contributions authored by NIH grant holders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

**Production Questions**

Ms. Deb Keener

Phone: 717-738-9334 Fax: 717-738-9478 or 717-738-9479

E-mail: keenerd@cadmus.com

## 7. Bibliografia


Baker, D.; Sali, A.: Protein structure prediction and structural genomics. *Science*, **2001**, 249, 93-96.

Becker, R. C.: Cell-Based Models of Coagulation: A Paradigm in Evolution. *J. Thomb. Thrombolysis*, **2005**, 20, 65-68.

Bishop, A. Ö. T.; de Beer, T. A. P.; Joubert, F.: Protein homology modelling and its use in South Africa. *South Africa J. Science*, **2008**, 104, 2-6.

de Sant'Anna, C. M. R.: Glossário de termos usados no planejamento de fármacos (recomendações IUPAC 1997). *Quím. Nova*, **2002**, *25*, 505-512.

Degen, S. J.; Davie, E. W.: Nucleotide Sequence of the Gene for Human Prothrombin. *Biochem.*, **1987**, 26, 6165-6177.

Elion J. *et al.*: Structure of human thrombin: comparison with other serine proteases. Em *Chemistry and Biology of Thrombin*; Ann Arbor Science, Michigan, **1977**, pp 97-112.

Esmon, C. T.; Jackson, C. M.: The Conversion of Prothrombin to Thrombin. *J. Biol. Chem.,* **1974**, 249, 7791-7797.

Faulon, J-L; Rintoul, M. D.; Young, M. M.: Constrained walks and self-avoiding walks: implications for protein structure determination. *J. Phys. A: Math. Gen.*, **2002**, 35, 1-19.

Ginalski, K.: Comparative modeling for protein structure prediction. *Current Opinion in Structural Biology*, **2006**, 16, 172-177.

Guyton, A. C.; Hall, J. E.: Hemostasia e Coagulação Sanguínea. Em *Tratado de Fisiologia Médica*, 11ª Ed.; Elsevier, Rio de Janeiro, 2006, pp 459-464.

Hirsh, J.: Current anticoagulant therapy-unmet clinical needs. *Thromb. Res.*, **2003**, *109*, S1-S8.

Hoffman, M.: A cell-based model of coagulation and the role of factor VIIa. *Blood Reviews*, **2003**, 17, 51-55.

Jenny, N. S.; Mann, K. G.: Coagulation Cascade: An Overview. Em *Thrombosis and Hemorrage,* 2ª Ed., Loscalzo, J.; Schafer, A. I. Eds.; Williams & Wilkins, Baltimore, **1998**, pp 3-22.

Journal Citation Reports® – Science Edition and Journal Citation Reports® - Social Sciences Edition Databases of the Institute for Scientific Information®, Inc., ISI®, Philadelphia, Pennsylvania, USA, Copyright© 2009.

Karplus, M.; McCammon, J. A.: Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.*, **2002**, 9, 646-652.

Karplus, M.; Petsko, G. A.: Molecular dynamics simulations in biology. *Nature*, **1990**, *347*, 2577-2637.

Katzung, B. G.: Fármacos usados em Desordens da Coagulação. Em *Farmacologia Básica e Clínica*, 9ª Ed; Guanabara Koogan, **2006,** pp 760-785.

Mizuochi, T. *et al.*: Studies on the Structures of the Carbohydrate Moiety of Human Prothrombin. *J. Biochem.*, **1981**, 90, 1023-1031.

Kotkow, K. J. *et al.*: The Second Kringle Domain of Prothrombin Promotes Factor Va-mediated Prothrombin Activation by Prothrombinase. *J. Biol. Chem.*, **1995**, 270, 4551-4557.

Krishnaswamy, S. *et al.:* Activation of Human Prothrombin by Human Prothrombinase: Influence of factor Va on the reaction mechanism. *J. Biol. Chem.*, **1987**, 262, 3291-3299.

Laskowski, R. A. *et al.*: PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystalogr.*, **1993**, 26, 283-291.

Leach, A. R.: Molecular Modelling Principles and Applications, 2nd Ed., **2001**, Longman, Cingapura.

Ligabue-Braun, R.; Verli, H.; Carlini, C. R.: Comportamento conformacional da urease de 'Canavalia ensiformis'. *Dissertação de Mestrado*, **2010**, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul.

Lüthy, R.; Bowie, J. U.; Eisenberg, D.: Assessment of protein models with three-dimensional profiles. *Nature*, **1992**, 356, 83-85.

Marrink, S. J. *et al.*: The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B*, **2007**, 111, 7812-7824.

Marrink, S. J. *et al.*: The MARTINI Forcefield. Em *Coarse graining of condensed phase and biomolecular systems*, CRC press, Salt Lake City, **2008**, pp 5-20.

Marrink, S. J.; Mark, A. E.: The mechanism of vesicle fusion as revealed by molecular dynamics simulations. *JACS*, **2003**, 125, 11144-11145.

Martí-Renom, M. A. *et al.*: Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.*, **2000**, 29, 291-325.

Melo, F.; Feytmans, E.: Assessing protein structures with a non-local atomic interaction energy. *J. Mol. Biol.*, **1998**, 277, 1141-1152.

Monticelli, L. *et al.*: The MARTINI Coarse-Grained Forcefield: Extension to Proteins. *Journal Chemical Theory and Computation*, **2008**, 4, 819-834.

Nader, H. B. *et al.*: Development of new heparin-like compounds and other antithrombotic drugs and their interaction with vacular endothelial cells. *Braz. J. Med. Biol. Res.*, **2001**, *34*, 699-709.

Nesheim, M. E.; Mann, K. G.: The Kinetics and Cofactor Dependence of the Two Cleavages Involved in Prothrombin Activation. *J. Biol. Chem.*, **1983**, 258, 5386-5391.

Persborn, E.: Factor Xa inhibitors: New anticoagulants for secondary haemostasis. *Hämostaseologie*, **2009**, 3, 260-267.

Petrey, D.; Honig, B.: Protein Structure Prediction: Inroads to Biology. *Molecular Cell*, **2005**, 20, 811-819.

Risselada, H. J.; Marrink, S. J.: The molecular face of lipid rafts in model membranes. *PNAS*, **2008**, 105, 17367-17372.

Rzepiela, A. *et al.*: Membrane poration by antimicrobial peptides combining atomistic and coarse grained descriptions. *Farad. Discuss.*, **2010**, 144, 431-443.

Samama, M. M.; Gerotziafas, G. T.: Newer anticoagulants in 2009. *J. Thromb. Thrombolysis,* **2010**, 29, 92-104.

Smyth, M. S.; Martin, J. H. J.: X-ray crystallography. *Mol. Path.*, **2000**, 53, 8-14.

Scott, W. R. P. *et al.*: The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A.*, **1999**, 103, 3596-3607.

Sippl, M. J.: Recognition of errors in three-dimensional structures of proteins. *Proteins*, **1993**, 17, 355-362.

Tozzini, V.: Multiscale Modeling of Proteins. *Acc. Chem. Res.*, **2010**, 43, 220-230.

Treptow, W.; Marrink, S. J.; Tarek, M.: Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations. *JPC-B*, **2008**, 112, 3277-3282.

Turpie, A. G. G.: Oral, Direct Factor Xa Inhibitors in Development for the Prevention and Treatment of Thromboembolic Diseases. *Arterioscler. Thromb. Vasc. Biol.*, **2007**, 27, 1238-1247.

van Gunsteren, W. F.; Berendsen, H. J. C.: Computer simulations of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.*, **1990**, *29*, 992-1023.

Vaqué, M. *et al.*: Protein-ligand Docking: A Review of Recent Advances and Future Perspectives. *Current Pharmaceutical Analysis*, **2008**, 4, 1-19.

Wellock, C.; Ross, B. J.: An Examination of Lamarckian Genetic Algorithms. *Proceedings of Late Breaking Papers, Genetic and Evolutionary Computation Conference (GECCO)*, **2001**, 474-481.

Wenfan, H.: Rigid Body Protein by Fast Fourier Transform. *Honor Year Project Report*, **2005**, National University of Singapore.

Wüthrich, K.: Protein Structure Determination in Solution by NMR Spectroscopy. *J. Biol. Chem.*, **1990**, 265, 22059-22062.