Felipe Augusto Nunes Ferraz

Caracterização de Sítios Conformacionais de Fosforilação em Proteínas

Felipe Augusto Nunes Ferraz

Caracterização de Sítios Conformacionais de Fosforilação em Proteínas

Dissertação de mestrado. Requisito para a obtenção do título de Mestre em Ciências no programa de Bioinformática

Universidade de São Paulo - USP Instituto de Matemática e Estatística - IME Programa de Pós-Graduação Interunidades em Bioinformática

Orientador: Paulo Sergio Lopes de Oliveira

Brasil

2016

Felipe Augusto Nunes Ferraz

Caracterização de Sítios Conformacionais de Fosforilação em Proteínas/ Felipe Augusto Nunes Ferraz. – Brasil, 2016-

89 p.: il. (algumas color.); 30 cm.

Orientador: Paulo Sergio Lopes de Oliveira

Dissertação (mestrado) – Universidade de São Paulo - USP Instituto de Matemática e Estatística - IME Programa de Pós-Graduação Interunidades em Bioinformática, 2016.

1. Motivo de Fosforilação. 2. Biologia Estrutural I. Paulo Sergio Lopes de Oliveira II. Universidade de São Paulo III. Instituto de Matemática e Estatística IV. Caracterização de Sítios Conformacionais de Fosforilação em Proteínas

Resumo

A fosforilação de proteínas é o tipo de modificação pós-traducional mais recorrente nas vias de sinalização, desempenhando papel central numa vasta gama de eventos celulares. Um completo entendimento das circunstâncias que coordenam o evento de fosforilação permanece como um desafio para a ciência, a despeito do crescente número de abordagens e estudos realizados no assunto. Um mecanismo largamente descrito e aceito como essencial para coordenar a fosforilação de proteínas é a existência de sequências de aminoácidos que facilitam a fosforilação, conhecidos como consensos de fosforilação. Nesse modelo, cada proteína quinase reconhece sítios de fosforilação se os mesmos estiverem inseridos em uma sequência específica de resíduos na estrutura primária do substrato. Porém, com o crescente volume de dados sobre fosforilação, é possível notar a existência de sítios que são validados experimentalmente como fosforilados por uma determinada proteína quinase, que não apresentam o consenso de fosforilação. Neste trabalho, foi testada e comprovada a hipótese de que estes sítios de fosforilação sem consenso sequencial apresentam resíduos localizados em regiões da estrutura terciária adjacentes ao sítio de fosforilação, cuja as características estereoquímicas mimetizam um peptídeo substrato contendo o consenso de fosforilação. Para essa avaliação, utilizando substratos da PKA, foi constatado que mais de 90% dos sítios de fosforilação que não apresentam o consenso na estrutura primária, apresentam essa disposição na estrutura terciária. Resíduos distantes na estrutura primária se apresentam próximos espacialmente na estrutura tridimensional, em uma conformação semelhante a de um sítio com o consenso de fosforilação. Com isso nós propomos a existência de sítios conformacionais de fosforilação. Para confirmar que esses sítios conformacionais poderiam ser cruciais no reconhecimento do substrato, foram construídos modelos da interação da proteína quinase com os substratos, visando demonstrar a viabilidade da interação dos resíduos formadores do consenso conformacional com a proteína quinase de maneira análoga a de um substrato com o consenso de fosforilação. Para a comprovação experimental do fenômeno, foi utilizado o modelo de fosforilação da α -Tubulina, no qual foi constatada uma fosforilação no resíduo T253 que depende da atuação dos resíduos K163 e K164 para a interação com a proteína quinase, confirmando a coerência do modelo proposto. Diante da novidade da proposta, dos estudos computacionais feitos e da validação conseguida, torna-se clara a relevância de se estudar a estrutura tridimensional dos substratos de fosforilação, não só como uma forma de aprofundar os conhecimentos gerais na área de fosforilação, mas também como uma alternativa com potencial de ser explorada no desenvolvimento de novas tecnologias.

Palavras-chaves: Fosforilação. Consenso de Fosforilação. Sítios Conformacionais de Fosforilação. Proteínas-quinase. Estrutura tridimensional. Modelagem Molecular. Bioinformática. Biologia computacional.

Abstract

Protein phosphorylation is the most frequent type of post-translational modification in signaling pathway, developing a key role in a wide range of cell events. The full understanding of the circumstances that coordinate the phosphorylation event remains a challenge for science, despite the growing number of approaches and studies on the subject. A broadly described and accepted mechanism as essential for the coordination of protein phosphorylation is the existence of amino acids sequences that contribute to phosphorylation occurrence, known as phosphorylation consensus. In this model, each protein kinase is able to recognize phosphorylation sites inserted in a specific sequence on the primary structure. However, as the data about phosphorylation sites increases, it is possible to notice that there are sites that are validated experimentally as phosphorylated by a particular protein kinase, which do not have the consensus phosphorylation. In this work, it was tested and proved that phosphorylation sites without the sequence consensus presents anchors residues, that are close to the phosphorylation site on the tertiary structure, creating a structural conformation that mimics the stereochemical features of a substrate peptide containing the phosphorylation consensus. For this evaluation, using substrates of PKA, it was found that more than 90% of phosphorylation sites that have no consensus on the primary structure, presented this kind of disposition on the tertiary structure. Distant residues in the primary structure are spatially close on the three-dimensional structure, in a conformation similar to a phosphorylation site containing the consensus. Thus we proposed the existence of conformational phosphorylation sites. To confirm that these conformational sites could be crucial in substrate recognition, it was built kinase-substrate models, aiming to demonstrate the feasibility of residues forming the conformational consensus on the substrate to interact with the kinase analogously to a substrate with consensus phosphorylation. For experimental verification of this phenomenon, we used the phosphorylation model of α -Tubulin, in which we observed a phosphorylation at residue T253 that depends of residues K163 and K164 to interact with the protein kinase, confirming the consistency of proposed model. Faced with the novelty of the proposal, the computational data and the experimental validation, it becomes clear the importance of studying the three dimensional structure of phosphorylation sites, not only as a way of achieving deeper knowledge on phosphorylation field, but also as a potential prospect to be explored on the development of new technologies.

Key-words: Phosphorylation. Phosphorylation Consensus. Conformational Phosphorylation Site. Protein Kinase. Three Dimensional. Molecular Modeling. Bioinformatics. Computational Biology.

Lista de ilustrações

Figura 1 –	Visão geral da proteína PKA e seu sítio catalítico	22
Figura 2 –	Exemplo de dados extraídos do banco de fosforilação	28
Figura 3 –	Esquema da busca estrutural de âncoras ao redor do sítio	32
Figura 4 –	Representação do algorítimo para definir o $docking$ guiado pelo PKI	39
Figura 5 –	Algorítimo utilizado para definir uma conformação termodinamicamente	
	viável do complexo PK-substrato	40
Figura 6 –	Interação do PKI com os subsítios da PKA	42
Figura 7 –	Possível modelo de sítio estrutural com âncoras em região sem cobertura:	
	Myosin regulatory light polypeptide 9	43
Figura 8 –	Sítio modelado sem consenso estrutural: Vitamin D3 receptor	44
Figura 9 –	Distribuição de RMSD da avaliação das estruturas dos complexos $$	46
Figura 10 –	Exemplo de complexo com consenso estrutural: Microtubule-associated	
	proteins 1A/1B light chain 3B	47
Figura 11 –	Exemplo de complexo com consenso estrutural: Regulator of G-protein	
	signaling 13	48
Figura 12 –	Exemplo de complexo com consenso estrutural: ribosomal protein S29,	
	mitochondrial	49
Figura 13 –	Exemplo de complexo com consenso linear: Glia maturation factor beta	50
Figura 14 –	Sítio e âncoras presente em α -hélice	51
Figura 15 –	Exemplo de complexo com consenso estrutural mais favorável que o	
	linear: $G1/S$ -specific cyclin-D1	52
Figura 16 –	Padrão de conservação de resíduos pertencentes ou não a sítios confor-	
	macionais	54
0		55
Figura 18 –	•	57
Figura 19 –	Painel de resultados experimentais da fosforilação da α -Tubulina	58

Lista de tabelas

Tabela 1 –	Correlação do Z-escore e o julgamento do YASARA	30
Tabela 2 –	Análise de RMSD para o substrato Microtubule-associated proteins	
	1A/1B light chain $3B$	47
Tabela 3 –	Análise de RMSD para o substrato Regulator of G-protein signaling 13	48
Tabela 4 –	Análise de RMSD para o substrato 28S ribosomal protein S29, mito-	
	chondrial	49
Tabela 5 –	Análise de RMSD para o substrato α -Tubulina	56
Tabela 6 –	Modelos de substratos contendo sítios sem consenso linear gerados por	
	homologia	74
Tabela 7 –	Avaliação dos complexos gerados para PKA	75

Lista de abreviaturas e siglas

ATP Adenosine Triphosphate

cAMP Cyclic adenosine monophosphate

DSC Discrimination of Secondary Structure Class

GFP Green Fluorescent Protein

HEK 293 Human Embryonic Kidney 293

SBS Substrate Binding Site (Sítio de Ligação do Substrato)

PDB Protein Data Bank

PK Protein Kinase

PKA Protein Kinase A

PKC Protein Kinase C

pN Piconewton

PSSP Profiles from Sequence and Structuraly related Proteins

RMSD Root Mean Square Distance

PP1 Protein Phosphatase 1

TGF β Transforming growth factor beta

Sumário

1	INTRODUÇÃO	15
1.1	Fosforilação e sinalização celular	15
1.2	Reconhecimento de Substratos – Visão Histórica	16
1.3	Reconhecimento de Substratos – Visão Estrutural	18
1.4	A família das proteína-quinase e seus grupos	20
1.5	Proteína quinase A e proteína quinase C	21
1.6	Dados Experimentais de Fosforilação	22
2	OBJETIVO	23
2.1	Objetivos específicos	23
3	JUSTIFICATIVA	25
4	MÉTODO	27
4.1	Prospecção do banco de dados	27
4.2	Geração da estrutura tridimensional de substratos por modelagem	
	por homologia	28
4.3	Prospecção estrutural dos sítios sem consenso linear na estrutura	
	dos substratos	31
4.4	Gerando modelos de interação com a PK	33
4.4.1	Visão geral	33
4.4.2	Docking geométrico de PKA-substrato guiado pelo modelo experimental PKA-PKI	33
4.4.3	Remoção de impedimentos estéricos do complexo PK-substrato	
4.4.4	Geração de conformações termodinamicamente viáveis para o complexo	55
	modelado PK-substrato	36
4.5	Avaliação dos modelos estruturais dos complexos gerados	37
4.6	Conservação dos resíduos âncoras em sítios conformacionais	38
5	RESULTADOS E DISCUSSÃO	41
5.1	Abundância de sítios de fosforilação conformacionais permite estu-	
	dos estruturais em larga escala	41
5.2	O conceito de sítio conformacional	41
5.3	Analise estrutural dos complexos PK-substratos modelados com sí-	ΛE
	tios conformacionais	43

5.4	Analise estrutural dos complexos quinase-substratos modelados com
	sítios com o consenso linear
5.5	Possíveis aspectos da relevância biológica de sítios conformacionais 52
5.6	Os sítios conformacionais em ação: Uma visão in vivo da fosforilação
	de alpha-tubulina
5.7	Considerações sobre o método de docking PK-substrato desenvol-
	vido neste trabalho
6	CONCLUSÃO
	REFERÊNCIAS 63
	ANEXOS 71
	ANEXO A – TABELA DOS SUBSTRATOS DE PKA GERADOS POR HOMOLOGIA
	ANEXO B – TABELA COM A AVALIAÇÃO POR RMSD DOS COMPLEXOS GERADOS PARA OS SUBSTRATOS DE PKA
	ANEXO C – ARTIGO: PROTEIN FOLDING CREATES STRUCTURE- BASED, NONCONTIGUOUS CONSENSUS PHOSPHORY LATION MOTIFS RECOGNIZED BY KINASES 81

1 Introdução

1.1 Fosforilação e sinalização celular

A fosforilação de proteínas é o tipo de modificação pós-traducional mais recorrente nas vias de sinalização celular. Virtualmente todos os processos celulares são afetados por fosforilação, com o espectro de atuação abrangendo uma grande diversidade de eventos, incluindo metabolismo, divisão, diferenciação, mobilidade, movimentação de organelas, transporte de membranas, apoptose, contração muscular, imunidade, aprendizado e memória (BRINKWORTH; BREINL; KOBE, 2003; JR; UBERSAX, 2007).

Do ponto de vista químico a fosforilação pode ser descrita como a transferência do fosfato-gama de uma molécula de ATP para um substrato. No caso em que o substrato é uma proteína, essa transferência é realizada para aminoácidos específicos. Em eucariotos, os aminoácidos mais comuns de servirem como sítios de fosforilação são serina, treonina e tirosina.

Por ter um papel central em um cenário abrangente dentro das vias de sinalização celular, as quinases que fosforilam proteínas são extensivamente estudadas. As PKs formam uma das maiores famílias de genes em eucariotos, representando aproximadamente 2% do genoma (MONTENEGRO et al., 2007). Estima-se que 30% das proteínas celulares possuam ao menos um resíduo fosforilado. Para ter uma noção do espaço de possibilidades que uma PK encontra para fosforilar, podemos estimar que em uma célula eucariótica existem cerca de 10.000 proteínas. Se o tamanho médio dessas proteínas for de 400 aminoácidos e sabendo que na distribuição dos aminoácidos temos 8,5% de serina, 5,7% de treonina e 3% de tirosina (ECHOLS, 2002) totalizando 17% de resíduos fosforiláveis, podemos estimar que existem aproximadamente 700.000 sítios de fosforilação em potencial. Mesmo considerando que parte desses potenciais sítios se encontrem enterrados na estrutura do substrato e portanto inacessíveis a PK, ainda assim temos um universo grande de possíveis pontos de interação entre substratos e PKs. Obviamente esses sítios não são fosforilados de maneira errática, o que geraria um caos na rede de sinalização celular. Não obstante, muitos tipos de câncer são causados ou modulados por alterações em PK (ZHANG; YANG; GRAY, 2009). Existem mecanismos que coordenam a interação das PKs com os substratos, como a alternância entre os estados ativo e inativo da PK, e os mecanismos de especificidade na seleção de substrato.

1.2 Reconhecimento de Substratos - Visão Histórica

Entender a seletividade exercida pelas quinases é um tema de pesquisa que apresenta uma longa história e consequentemente uma extensiva literatura. Trabalhos elucidando a especificidade de diferentes tipos de quinases, como a hexoquinase (SOLS; CRANE, 1954) e creatina quinase (ENNOR; ROSENBERG; ARMSTRONG, 1955), remontam aos anos 1950. Já para proteínas-quinase, encontramos os primeiros registros de estudos versando a especificidade das mesmas na década de 1960 (RABINOWITZ; LIPMANN, 1960).

Estudos para a determinação dos critérios de especificidade de substratos de PKs apresentam um elevado grau de complexidade. Um passo crucial para a correta caracterização de uma PK é a determinação de substratos fisiológicos. Em 1979, Edwin G. Krebs e Joseph A. Beavo realizaram uma revisão bibliográfica e delinearam quatro critérios a serem considerados para a determinação de um substrato fisiológico (KREBS; BEAVO, 1979):

- 1. O substrato tem que ser fosforilado estequiometricamente e desfosforilado por uma fosfatase *in vitro* a uma taxa cinética significativa.
- Propriedades funcionais do substrato devem se alterar em correlação com o grau de fosforilação.
- 3. A fosforilação do substrato deve ser verificada *in vivo* ou em células intactas acompanhada de alteração funcional.
- 4. Correlação entre o nível de PK e a extensão de seus efeitos.

Apesar de não serem definitivos, a observância desses critérios são diretrizes para a caracterização de sítios como fisiológicos, que são os verdadeiramente interessantes de serem estudados. Porém mesmo estudando apenas sítios comprovadamente fisiológicos, o entendimento do mecanismo molecular da interação entre quinase e substrato ainda não era completamente entendido. Nessa mesma revisão, Krebs e Beavo ressaltam a importância da estrutura primária para o reconhecimento de substratos. Essa conclusão deriva de uma série interessante de observações da literatura disponível da época.

Em 1975, Philip Cohen e colaboradores, ao estudarem a atividade da fosforilase quinase na glicogenólise, constataram que a ativação dessa enzima dependia de fosforilações realizadas por PKA (COHEN; WATSON; DIXON, 1975). O conceito corrente na época era de que a PKA apresentava uma baixa seletividade, uma vez que era conhecida a capacidade dessa PK de reconhecer diversos substratos. Nesse trabalho eles refutam essa ideia baseando-se em uma série de evidências:

- 1. A PKA fosforilou in vitro apenas duas serinas em toda a estrutura da fosforilase quinase, com cada fosforilação apresentando uma função bem definida. A fosforilação do sítio presente na subunidade β aumentou a atividade da fosforilase quinase, enquanto que na subunidade α tornava a fosforilase quinase um substrato efetivo da fosforilase quinase fosfatase.
- Na literatura, eles encontraram um estudo funcional bem caracterizado de uma única fosforilação promovida por PKA em histona F1 (LANGAN, 1971).
- 3. A inativação da glicogênio sintase se correlaciona com a incorporação de fosfato, em uma proporção de um mol de fosfato para cada mol da enzima (SODERLING et al., 1970).
- 4. A taxa de fosforilação de serinas em qualquer outra enzima na via da glicogenólise ou em um conjunto testado de proteínas do músculo esquelético, apresentaram uma taxa de fosforilação negligenciável se comparada com o que havia sido constatado nas proteínas fosforilase quinase, glicogênio sintase e histona F1.

Pelo que foi observado nos estudos, a PKA fosforila substratos seletos exercendo funções altamente específicas, tornando claro que essa PK não fosforila a esmo qualquer serina exposta. A primeira análise feita pelos autores desse estudo foi a comparação da sequência de resíduos em torno das serinas fosforiladas nos substratos estudados, na qual eles concluem que não há uma similaridade sequencial maior do que a que ocorreria ao acaso. Com isso, é descartada a possibilidade de PKs reconhecerem sequências lineares em substratos. Diante desses dados, a conclusão do trabalho é que a PKA deve reconhecer alguma forma de estrutura tridimensional comum a todos os substratos fosforilados. Conclusão semelhante havia sido obtida em um outro trabalho (LANGAN, 1973).

Porém essa hipótese de reconhecimento de sítios tridimensionais foi desacreditada por uma série de trabalhos. Na revisão de Krebs e Beavo, 17 artigos são citados para refutar essa possível explicação para a seleção de substratos pela PK. Um breve sumários dos argumentos e citações utilizados será feito a seguir.

Um trabalho chave para atestar a impossibilidade de o reconhecimento de sítios por PK ser feito pela estrutura tridimensional dos substratos foi feito por Peter Daile e Patrick R. Carnegie. Esse estudo, realizado concomitantemente ao do Phillipe Cohen e colaboradores, deliberadamente refuta os resultados obtidos em (LANGAN, 1973), definindo a importância da estrutura primária no reconhecimento. Para tal, os autores realizaram uma digestão peptídica de um substrato conhecido de PKA e demonstraram que era possível selecionar um peptídio derivado desse processo para servir de substrato da PKA, ou seja, era possível fosforilar o peptídeo isolado, independe do restante da estrutura do substrato (DAILE; CARNEGIE, 1974).

Outro argumento contrário ao reconhecimento estrutural é encontrado em (BY-LUND; KREBS, 1975) e (HUMBLE et al., 1975), nos quais foram constatados que o desenvelamento da proteína aumentava a suscetibilidade a fosforilação do substrato.

Observações qualitativas sobre a importância da composição de aminoácidos ao redor de sítios conhecidos corroboraram para a aceitação do conceito de reconhecimento sequencial. Estudos destacando a importância da presença de um resíduo básico na direção N-terminal para o reconhecimento de substratos de PKA foram feitos utilizando variações genéticas do substrato β -caseína (KEMP et al., 1975), pela observação de sequências ricas em lisina de peptídeos derivados de histonas (SHLYAPNIKOV et al., 1975) e pelo desenvolvimento de peptídeos sintéticos como potenciais substratos de PKA (DAILE; CARNEGIE; YOUNG, 1975; KEMP; BENJAMINI; KREBS, 1976). Também surgiram trabalhos destacando a importância de dois resíduos básicos adjacentes ao sítio de fosforilação (HJELMQUIST et al., 1974; KEMP et al., 1977). O conceito de reconhecimento de sequência linear foi expandido por avaliação de substratos naturais de PKA (YEAMAN et al., 1977; COHEN; RYLATT; NIMMO, 1977) além de mais trabalhos envolvendo peptídeos sintéticos (POMERANTZ et al., 1977; KUROCHKIN et al., 1978) .

Uma vez estabelecido que a estrutura primária era o elemento chave para o reconhecimento de sítios de fosforilação, estudos sobre o tema convergiram para a avaliação dos substratos nessa esfera de complexidade. Uma consequência direta desses estudos foi o surgimento do conceito de consenso de fosforilação, isso é, baseando-se na inspeção sequencial de sítios conhecidos, definem-se características mínimas de uma sequência plausível de ser um sítio de fosforilação. Para a PKA, o consenso foi inicialmente determinado como KRXXS ou RRXS (KREBS; BEAVO, 1979). Esse conceito de consenso de fosforilação sequencial perdura até os dias de hoje. Em 2001, John B. Schabb escreveu uma extensa revisão sobre sítios fisiológicos de PKA, no qual o consenso de PKA relatado já apresenta um padrão mais abrangente, podendo ser RRX[S/T], [R/K]XX[S/T] ou [R/K]X[S/T] (SHABB, 2001). Nesse mesmo trabalho existe a menção a sítios fisiológicos de PKA que não estão contidos no consenso tradicional de fosforilação.

1.3 Reconhecimento de Substratos - Visão Estrutural

As PKs eucarióticas no geral são estruturalmente muito similares. O domínio catalítico canônico de uma PK apresenta cerca de 250 aminoácidos. A sua estrutura pode ser dividida em um lóbulo N-terminal pequeno com predominância de folhas β e um lóbulo C-terminal formado por α -hélices. O sítio de ligação do ATP está localizado numa fenda entre os dois lóbulos enquanto o substrato proteico se liga ao longo da fenda, região denominada de sítio de ligação do substrato (SBS). Mesmo que as PKs compartilhem uma estrutura global em comum, características específicas do SBS contribuem para a

capacidade de selecionar substratos.

A profundidade do SBS está relacionada com a preferência pelo resíduo a ser fosforilado. As PKs com maior afinidade por fosforilar tirosina (tirosina quinase), apresentam a fenda do SBS mais profunda do que PKs que fosforilam preferencialmente serina ou treonina (serina/treonina quinase). Isso pode ser explicado pelo tamanho da cadeia lateral desses resíduos, uma vez que tirosinas conseguem se adaptar a uma maior distância entre a entrada da fenda e o sítio catalítico (HUBBARD, 1997; JOHNSON et al., 1999). Evidentemente esse mecanismo de seleção não é absoluto, havendo casos de fosforilação cruzada, sendo mais comum a fosforilação de tirosinas por serina/treonina quinases do que o inverso.

Um fator determinante para a especifidade das PKs é a interação do SBS da PK com os resíduos que flanqueiam o sítio de fosforilação no substrato. A presença de resíduos no entorno do sítio de fosforilação que apresentem características físico-químicas complementares àquelas presentes na superfície do SBS favorece energeticamente a formação do complexo do substrato com a PK. A importância desses resíduos, que serão referenciados como resíduos âncoras ao longo do texto, já foi evidenciada experimentalmente por estudos de mutação de substratos conhecidos (KEMP et al., 1975; DAILE; CARNEGIE; YOUNG, 1975) e por estudos in silico de modelagem computacional (BRINKWORTH; BREINL; KOBE, 2003).

Essa observação da importância de resíduos âncoras é coerente com o conceito de consensos de fosforilação. Isso porque resíduos próximos ao sítio de fosforilação na sequência possuem alta probabilidade de interagirem com o SBS simultaneamente ao sítio, e portanto podem auxiliar a interação do substrato com a PK se apresentarem características físicoquímica favoráveis, sendo esse conjunto de características uma peculiaridade de cada PK. Dentro da sequência fosforilada, o sítio é comumente nomeado como P0, com os resíduos que o antecedem, isso é na direção N-terminal, nomeados de forma decrescente (P-1, P-2, P-3...), enquanto que os resíduos que o procedem, isso é, em direção ao C-terminal, de forma crescente (P1, P2, P3...). Em contrapartida, o SBS pode ser divido em sub-regiões relacionadas com a parte do substrato com a qual interage, seguindo uma numeração análoga a utilizada na denominação do substrato. Então, o sítio catalítico, que interage com o resíduo a ser fosforilado, é denominado S0, e as regiões que flanqueiam o sítio catalítico seguem numerações decrescentes em um dos sentidos (S-1, S-2, S-3...) e crescentes no outro (S1, S2, S3...). Portanto os resíduos no entorno dos sítios de fosforilação podem atuar como âncoras quando há uma complementariedade entre o resíduo em P(X) no substrato com os resíduos formadores da região S(X) da PK, criando interações favoráveis por carga, ligação de hidrogênio ou interações hidrofóbicas.

A determinação do consenso de fosforilação pode auxiliar o entendimento da atuação de uma PK. Baseado em sítios conhecidos, existem métodos que avaliam a conservação dos

resíduos na sequência visando determinar o consenso de fosforilação (ALTO et al., 2003; HJERRILD et al., 2004). Uma possível aplicação desse dado é buscar e restringir o rol de possíveis PKs candidatas a atuarem em um sítio de fosforilação recém descoberto. Uma das aplicações mais promissoras é a prospecção do genoma em busca de proteínas passíveis de serem fosforiladas por uma determinada PK (YAFFE et al., 2001). Entretanto, a presença única e exclusiva do consenso de fosforilação na proteína não garante que a mesma é um subtrato in vivo da PK. Métodos computacionais preditivos incorporam cada vez mais informação no modelo com a finalidade de melhorar a prospecção de possíveis alvos para uma determinada PK. Introdução de informações estruturais do substrato é uma tendencia (DUREK et al., 2009; ZANZONI et al., 2010). Entre as informações incorporadas por esses métodos para predizer a probabilidade de um resíduo ser fosforilado, podemos citar: utilização da desordem intrínseca da região (programa DIPHOS) (IAKOUCHEVA, 2004) e a utilização de dados estruturais da cadeia principal de sítios conhecidos cujas estruturas estejam disponíveis no banco de dados de proteínas (programa NetPhos) (BLOM et al., 2004).

1.4 A família das proteína-quinase e seus grupos

As proteínas-quinase eucarióticas constituem uma numerosa família de proteínas homologas (HANKS; HUNTER, 1995). Considerando apenas o genoma humano, é possível identificar mais de 500 membros dessa família (MANNING, 2002). Todos os membros dessa família de enzimas são fortemente relacionados pelos seus domínios catalítico, apresentando um considerável grau de conservação de elementos da estrutura primária, elevado grau de conservação da estrutura terciária e consequentemente, do mecanismo funcional básico. PKs que apresentam domínios catalíticos altamente conservados são produtos de genes que passaram por uma diferenciação evolutiva relativamente recente. Consequentemente, membros de uma mesma família tendem a compartilhar características chaves, como o modo de ativação e a especificidade de substratos (HANKS; HUNTER, 1995).

Um esquema de classificação padronizado das PKs foi proposto inicialmente por (HANKS; HUNTER, 1995) e posteriormente aperfeiçoado e estendido em (MANNING, 2002). O sistema utiliza informações funcionais, estruturais, sequenciais e evolutivas para classificar as PKs. A classificação realizada é hierárquica e se divide em grupos, que abrangem famílias e que em alguns casos se dividem em subfamílias. Uma visão geral do esquema de classificação é dado abaixo.

Grupos: De maneira geral, divide as quinase quanto a especificidade de substrato.
 A lista dos grupos atualmente considerados para quinases eucarióticas são AGC,
 CMGC, CAMK, CK1, STE, TK, TKL, RGC, PKL, atípicas e um grupo agregando todas as quinases que não se encaixam nos demais.

- Famílias: Conjunto de quinases relacionadas tanto pela similaridade sequencial quanto pela função biológica. Alguns exemplos de famílas seria a divisão da família AGC nas famílias de PKA, PKC, PKG, PKN PDK1, AKT entre outras, ou o grupo CMGC, dividido nas famílias CDK, CDKL, CK2, CLK, DYRK, GSK, MAPK, RCK e SRPK.
- Subfamílias: Mesmo princípio das famílias, mas com maior grau de similaridade. Por exemplo a família de PKA é dividia nas subfamílias PKA, PKX, PKAHP e outras PKAs basais, enquanto que a família MAPK apresenta as subfamílias ERK1, ERK3, ERK5, JNK, p38, nmo, ERK7.

Nesse trabalho, foram utilizadas apenas quinases pertencentes ao grupo AGC, um dos primeiros a ser caracterizado, possuindo muitas quinases do núcleo de sinalização intracelular, sendo portanto um dos grupos mais bem caracterizados e estudados. O grupo possui esse nome em referência as principais famílias que o constitui (PKA, PKC e PKG). No total, 16 famílias constituem o grupo AGC, dos quais 8 provavelmente já existiam nos primeiros eucariotos.

1.5 Proteína quinase A e proteína quinase C

A proteína quinase A, também conhecida como proteína quinase cAMP dependente, é uma das enzimas mais estudadas, sendo portanto um modelo de estudo interessante pelo vasto conhecimento a respeito dela disponível. A sua estrutura foi a primeira PK a ser resolvida (KNIGHTON et al., 1991a), sendo inclusive resolvida a estrutura da mesma em complexo com o seu inibidor natural (KNIGHTON et al., 1991b). Seu consenso de fosforilação é bem conhecido, apresentando dois resíduos básicos nas posições P-2 e P-3. Tal preferência é explicada pela formação de bolsões com ambientes carregados negativamente na região do sítio de ligação (Figura 1).

Assim como a PKA, a PKC também apresenta um consenso de fosforilação conhecido, com resíduos básicos nas posições P-2 e P-3 (WOODGETT; GOULD; HUNTER, 1986). Além de validação *in vitro*, podemos confirmar esse consenso de fosforilação pelo fato de existir um mecanismo regulatório presente nas PKs da família das PKC que consiste num pseudosubtrato de sequência Arg19-Phe-Ala-Arg-Lys-Gly-Ala25-Leu-Arg-Gln-Lys-Asn-Val-His-Glu-Val-Lys-Asn36. Esse pseudosubstrato, presente no domínio regulatório dessas PKs, é capaz de se ligar ao sítio do domínio catalítico, mantendo o sítio ocupado e portanto a PK inativa. A substituição da ALA25 por uma serina, transforma a região em um substrato de alta afinidade (HOUSE; KEMP, 1987).

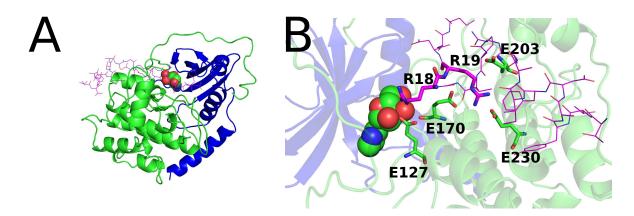


Figura 1 – A - Visão geral da PKA em complexo com o seu peptídeo inibidor natural [PDB: 1FMO]. Em cartoon azul o lóbulo N-teminal e em verde o C-terminal. Na representação de esferas uma adenosina ocupando o sítio de ligação do ATP e em lines rosa o peptídeo inibidor. B - Detalhes da região do sítio catalítico em que ocorre a interação das âncoras do peptídeo. Destaque em sticks no peptídeo R19 (posição P-2), que interage no bolsão formado pelos resíduos E170, E203 e E230, e R18 (P-3), que interage com o resíduo E127 da PKA, além de interagir com o ligante no sítio de ligação do ATP.

1.6 Dados Experimentais de Fosforilação

Avanços tecnológicos em métodos experimentais geram um volume cada vez maior de informações sobre sítios de fosforilação. O surgimento de métodos de detecção de fosforilação em larga escala, notadamente técnicas de espectrometria de massa, causaram um grande impacto nesse cenário. Estudos baseados em métodos de detecção mais tradicionais, como por exemplo os que utilizam anticorpos fosfo-específicos e experimentos de mutagênese, são tipicamente focados em poucos sítios de fosforilação, contrastando com os métodos de larga escala, capazes de detectar um amplo espectro de sítios.

Nesse cenário, um desafio latente para a comunidade cientifica é a organização e a facilidade de acesso a esses volumosos dados. Nesse sentido, destaca-se uma iniciativa do grupo Cell Signaling Technology, o PhosphositePlus (HORNBECK et al., 2011), um banco público de modificações pós-traducionais curados da literatura. Para ilustrar o aumento no volume de informações sobre o assunto, foi feita uma comparação da quantidade de sítios de fosforilações registrados atualmente em relação a dados históricos do PhosphositePlus. Em consulta realizada no dia 10/11/2015, foram encontrados 369.798 sítios não redundantes, associados a 19.996 proteínas. Quando esse projeto foi inciado em 2003, contava com 2.389 sítios cadastrados (HORNBECK et al., 2004) e em 2011 eram 100.173 (HORNBECK et al., 2011). Isso representa um crescimento de 4000% entre 2003 e 2011, seguido por um crescimento de 269% entre 2011 e 2015.

2 Objetivo

Esse trabalho visou contribuir para o entendimento do mecanismo de reconhecimento de sítios de fosforilação em substratos de proteínas-quinase, do ponto de vista da biologia estrutural e computacional. Principalmente, dos mecanismos de reconhecimento de sítios de fosforilação que não apresentam o consenso de fosforilação.

2.1 Objetivos específicos

- 1. Quantificar a frequência de sítios que não apresentam o consenso de fosforilação.
- 2. Estudar a disposição desses sítios nas estruturas dos substratos.
- Explicar o mecanismo molecular da interação desses substratos com a PK por meio de modelagem computacional.
- 4. Definir um protocolo de baixo custo computacional capaz de fazer uma avaliação em larga escala dos itens 2 e 3.

3 Justificativa

A despeito da importância da fosforilação para os processos de sinalização celular, questões fundamentais como o mecanismo de reconhecimento dos sítios de fosforilação pelas PKs constituem um problema em aberto. Desde o conhecimento de padrões de fosforilação em substratos até aspectos estruturais e dinâmicos que podem estar envolvidos neste mecanismo de interação antes da ocorrência da catálise. Graças ao retumbante desenvolvimento de tecnologias para o estudo de proteômica, cada vez mais dados a respeito de modificações pós-traducionais em larga escala tem se tornado disponíveis à comunidade científica. Neste contexto, a prospecção de bancos de dados experimentais de fosforilação de proteínas tem permitido observar que uma elevada porcentagem de sítios conhecidos, fosforilados pelas PKs A e C, não apresentam a sequência consenso esperada. Considerando a recorrência desse tipo de sítio, não só fica evidente a existência de limitações na proposta vigente baseada em consensos sequenciais, mas também levanta o questionamento sobre o impacto que as limitações desse modelo impõem nas pesquisas atuais. Consenso sequencial de fosforilação, apesar de ser útil em muitos contextos, é um conceito que se apresenta defasado quando confrontado com o volume de dados existentes hoje. É preciso considerar que a ideia de consenso sequencial de fosforilação surgiu em uma época na qual se dispunha de um volume de dados consideravelmente menor acerca dos sítios de fosforilação e, quando comparado com a atualidade, poucos recursos tecnológicos para estudá-los. Diante de tudo isso, torna-se essencial uma revisão do tema utilizando dados recentes e novas abordagens. O uso de informações estruturais de proteínas combinado com técnicas de biologia computacional podem auxiliar o entendimento do processo de fosforilação desses sítios, tema pouco explorado na literatura científica atual.

4 Método

4.1 Prospecção do banco de dados

O estudo de modelos de interação entre substratos sem o consenso linear e a PK requer, inicialmente, a definição da PK a ser estudada e um conjunto de substratos dessa PK. A PK escolhida para o estudo foi a PKA. A seleção foi feita levando em consideração dois fatores. Primeiramente o fato dessa PK ter um consenso de fosforilação bem definido, consistindo de resíduos básicos nas posições P-2 e P-3 (KOBE et al., 2005). Outro fator determinante foi o fato dessa PK ter a sua estrutura cristalográfica resolvida em complexo com o seu peptídeo inibidor natural, o qual será usado como modelo de organização local do entorno do sítio de fosforilação na interação do substrato com a PK (Figura 1) (KNIGHTON et al., 1991a). Um dos aspectos importantes em projetos de biologia computacional é a capacidade de se comprovar in vitro ou in vivo as hipóteses levantadas. Neste sentido, este trabalho foi exitoso. Por conta de uma oportunidade de validação experimental em uma colaboração com o grupo da Dra. Deborah Schechtman do Instituto de Química da USP, que já trabalhava com a PKC β , estendeu-se o estudo para substratos da PKC, neste caso, com enfoque específico na proteína α -Tubulina. Além disso, é importante notar que a PKC β , assim como a PKA, é uma PK muito bem caracterizada.

Outro aspecto importante para esse trabalho foi a definição dos substratos da PK a serem estudados. Para este fim foi utilizado o sistema *PhosphositePlus* (HORNBECK et al., 2011), um banco de modificações pós-traducionais que coleta informações curadas manualmente da literatura sobre sítios de fosforilação em proteínas. O sistema oferece a possibilidade de baixar o banco completo ou a utilização do sistema web para realizar as consultas. As consultas no sistema podem ser feitas pelos sítios conhecidos em uma determinada proteína ou por substratos e sítios associados a uma determinada PK. Como essa segunda opção atende precisamente a proposta aqui exposta, optamos por realizar buscas utilizando o sistema do *PhosphositePlus* para encontrar substratos das PKs selecionadas.

O conjunto de informações retornados pelo sistema consiste numa lista de sítios fosforilados validados experimentalmente. Essa lista incluí um conjunto de informações associadas ao substrato incluindo o nome da proteína, o símbolo do gene, o organismo, código de acesso no UNIPROT (APWEILER, 2004), peso molecular, o resíduo fosforilado, identificador do artigo no PubMed (ROBERTS, 2001) do qual a informação foi curada, a sequência do entorno do sítio com 7 resíduos no sentido N-terminal e mais 7 em direção ao C-terminal e informação sobre o método experimental utilizado para identificar o sítio, podendo ser *in vivo* ou *in vitro*. A figura 2 ilustra a saída obtida em uma consulta ao

28 Capítulo 4. Método

Protein	GeneSymbol	Organism	ACC#	MW(DA)	ResidueSite	GroupIdSite	Sequence	Method
ACC1	Acaca	rat	P11497	265,194	S77-p	448291	LAFHMRSsMSGLHLV	In Vitro
ACC1	Acaca	rat	P11497	265,194	S1200-p	448290	IPTLNRMsFASNLNH	In Vitro
ACC1 iso2	Acaca	rat	P11497-2	264,328	S1192-p	448290	TSHPNRMsFASNLNH	In Vitro
ACTA1	Acta1	rat	P68136	42,051	S35-p	2565387	APRAVFPsIVGRPRH	In Vitro
ACTA1	Acta1	rat	P68136	42,051	S201-p	2565391	ILTERGYSFVTTAER	In Vitro
ACTB	Actb	rat	P60711	41,737	S33-p	2565375	APRAVFPsIVGRPRH	In Vitro
ACTG1	Actgl	rat	P63259	41,793	S33-p	2565379	APRAVFPsIVGRPRH	In Vitro
ACTG2	Actg2	rat	P63269	41,877	S34-p	2565383	APRAVFPsIVGRPRH	In Vitro
ACTG2	Actg2	rat	P63269	41,877	S200-p	2565800	ILTERGYSFVTTAER	In Vitro
ACTN1	Actn1	rat	Q9Z1P2	102,960	S404-p	2565463	EKFRQKAsIHEAWTD	In Vitro
ACTN4	Actn4	rat	Q9QXQ0	104,915	S423-p	2565459	EKFRQKAsIHEAWTD	In Vitro
ADD1	ADD1	human	P35611	80,955	S59-p	451695	MEQKKRVsMILQSPA	In Vitro
ADD1	ADD1	human	P35611	80,955	5408-p	451594	REKSKKYsDVEVPAS	In Vitro
ADD1	ADD1	human	P35611	80,955	S436-p	451595	TCSPLRHsFQKQQRE	In Vitro
ADD1	ADD1	human	P35611	80,955	S481-p	451596	KEDGHRTsTSAVPNL	In Vitro
ADD1	ADD1	human	P35611	80,955	S726-p	449829	KKKFRTPsFLKKSKK	In Vitro
ADD2	ADD2	human	P35612	80,854	T55-p	451694	MEQKKRVtMILQSPS	In Vitro
ADD2	ADD2	human	P35612	80,854	S713-p	449802	KKKFRTPsFLKKSKK	In Vitro
ADRB1	ADRB1	human	P08588	51,323	S312-p	469625	RAGKRRPSRLVALRE	In Vivo
ADRB2	ADRB2	human	P07550	46,459	S345-p	447910	ELLCLRRsSLKAYGN	In Vitro
ADRB2	ADRB2	human	P07550	46,459	S346-p	447911	LLCLRRSsLKAYGNG	In Vitro
AHNAK	Ahnak	rat	NP 001178880	581,125	S136-p	2565487	KIKPRLRsEDGVEGD	In Vitro
AHNAK	Ahnak	rat	NP 001178880	581,125	T159-p	2565491	ITVTRRVtAYTVDVT	In Vitro
AHNAK	Ahnak	rat	NP 001178880	581,125	S5038-p	452760	EGPKVRGsLGATGEL	In Vitro
AID	AICDA	human	Q9GZX7	23,954	T27-p	471423	WAKGRRETYLCYVVK	In Vivo, In Vitro
AID	AICDA	human	Q9GZX7	23,954	S38-p	457604	YVVKRRDsATSFSLD	In Vivo, In Vitro
AID	Aicda	mouse	Q9WVE0	24,031	T27-p	471423	WAKGRHETYLCYVVK	In Vitro

Figura 2 – Exemplo do arquivo baixado como saída do banco *PhosphositePlus* para uma consulta de substratos de PKA. As colunas indicam o nome da proteína substrato, o gene, o organismo, o código UNIPROT, o peso molecular, o resíduo fosforilado, identificador do artigo PubMed de onde a informação foi extraída, a sequência do sítio e o tipo de experimento realizado na identificação do sítio.

Phosphosite Plus.

Tendo em vista que o foco do trabalho foi em sítios de fosforilação que não apresentam o consenso linear, o próximo passo foi classificar os substratos entre os que apresentam o consenso de fosforilação na estrutura primária dos que não apresentam. Para realizar essa tarefa de forma eficiente, foi desenvolvido um *script* em *Python* que dada uma saída de busca no banco de dados e um consenso de fosforilação, separa em arquivos distintos as entradas do banco que atendem ou não a sequência consenso buscada. Com isso foi possível prospectar diferentes consensos de fosforilação para diferentes PKs. Para o nosso caso em específico, realizamos a busca utilizando o consenso completo, com resíduos básicos em P-2 e P-3, mas também optamos por realizar uma busca menos restringente, utilizando como sequência a ser buscada sítios com pelo menos um resíduo básico nas posições P-2 ou P-3.

4.2 Geração da estrutura tridimensional de substratos por modelagem por homologia

Um dos objetivos do trabalho foi estudar o contexto estrutural dos sítios fosforilados que não apresentam o consenso de fosforilação. Para tal, foi necessário obter uma estrutura completa desses substratos, o que foi concretizado por meio de modelagem por homologia.

A opção de utilizar modelagem por homologia foi feita com a finalidade de homogenizar o processo de obtenção da estrutura tridimensional dos substratos. O identificador dos substratos utilizados pelo *PhosphositePlus* é código UNIPROT. No UNIPROT existem 549.646 entradas revisadas e 52.783.601 entradas geradas por meios de análises computacionais, enquanto no PDB temos 113.494 estruturas resolvidas (consulta realizada no dia 10/11/2015). Sendo assim o número de sequências registradas no UNIPROT é muito maior do que o número de estruturas resolvidas experimentalmente disponíveis no PDB, fazendo com que um número baixo de estruturas de substratos estejam disponíveis.

Para gerar as estruturas por modelagem computacional, primeiramente compilamos a sequência de todos os substratos. Para tanto foi escrito um script em *Python* que dado um código de acesso do UNIPROT realizava uma cópia do arquivo fasta, com a estrutura primária da proteína, diretamente do banco do UNIPROT.

A modelagem por homologia foi feita utilizando o software YASARA (KRIEGER et al., 2009). O processo seguido para a modelagem por homologia é apresentado abaixo, com comentários sobre os parâmetros utilizados.

Inicialmente todas as sequências são submetidas a uma busca utilizando 6 iterações do PSI-BLAST (ALTSCHUL, 1997) contra o banco de sequências UniRef90, gerando uma matriz de valores posição-específica (PSSM). Esse perfil é então utilizado para buscar no banco de estruturas PDB por potenciais moldes. Casos em que o melhor molde encontrado apresentava baixa relevância estatística (e-value > 0,5) não foram modelados.

O número máximo de moldes utilizados foi definido como cinco, com o intuito de evitar a construção de um número elevado de modelos desnecessários. Os moldes são classificados com base no escore de alinhamento e baseado na qualidade da estrutura molde, de acordo com os parâmetros do WHAT_CHECK obtido da base de dados PDBFinder2 (HOOFT et al., 1996).

Uma predição da estrutura secundária é feita utilizando o PSIPRED (MCGUFFIN; BRYSON; JONES, 2000) e o método de discriminação de classe da estrutura secundária (DSC) (KING; STERNBERG, 1996). Também é feito um perfil da sequência por meio de um alinhamento dos resultados do BLAST da sequência contra o UniRef90. O perfil das sequências dos moldes também é construído. Para tal o YASARA utiliza o PSSP (Profiles from Sequence and Structuraly related Proteins). Nesse método, realiza-se uma busca no PDB por moldes relacionados, que são alinhados estruturalmente utilizando os algorítimos CE (SHINDYALOV; BOURNE, 1998), FatCat (YE; GODZIK, 2003) e SHEBA (JUNG; LEE, 2000). O melhor resultado é mantido e utilizado para gerar um alinhamento múltiplo da sequência alvo contra o molde, ancorado em regiões de alinhamento estrutural bem definido.

O alinhamento do perfil da sequência e do molde é realizado utilizando o matriz

30 Capítulo 4. Método

de valores SSALN (QIU; ELBER, 2005), logo considerando as informações estruturais contidas nos moldes (por exemplo evitando lacunas nos elementos de estrutura secundária e mantendo resíduos polares expostos), bem como mantendo a coerência entre a estrutura secundária predita e assinalada pelos moldes. As regiões sem estrutura secundária definidas passam por uma otimização que consiste na geração de 50 conformações diferentes e otimização das cadeias laterais.

Para casos em que o alinhamento contra o molde não é preciso, foram gerados até 5 alinhamentos alternativos de alta pontuação, utilizando métodos estocásticos (MüCKS-TEIN; HOFACKER; STADLER, 2002). Modelos foram então construídos para todos os alinhamentos possíveis.

Todos os modelos gerados passam por uma minimização de energia, com solvatação implícita e utilizando o campo de força YASARA (KRIEGER et al., 2009). A minimização é realizada em dois passos. Inicialmente é gerado o modelo meio-refinado, que é a minimização feita com os átomos da cadeia principal dos resíduos conservados fixados. No segundo passo, a minimização é feita sem restrições, gerando um modelo totalmente refinado. O Z escore de qualidade dos dois modelos é comparada, sendo preservado o melhor. O Z escore engloba uma média da avaliação dos diedros e do empacotamento de uma e três dimensões. Uma correlação entre os valores de Z escore e o julgamento do YASARA é dado na tabela 1.

Tabela 1 – Correlação entre o Z escore calculado para o modelo e o julgamento dado pelo software YASARA.

Z escore	Descrição
>0	Ótimo
<0	Bom
<-1	Satisfatório
<-2	Pobre
<-3	Ruim
<-4	Terrível
<-5	Repugnante

Esse processo é realizado para os modelos gerados utilizando todos os moldes selecionados bem como todos os alinhamentos considerados. Com isso é gerado um perfil de qualidade por resíduo para todos os modelos. Por fim, é gerado um modelo híbrido, que partindo de um modelo, vai substituindo iterativamente regiões de baixo escore por fragmentos da região correspondente de outros modelos. Se o Z escore de qualidade do modelo híbrido for superior ao dos modelos gerados pelos moldes individuais, esse é considerado o modelo final. Caso contrário, o melhor modelo gerado pelos moldes individuais é utilizado como modelo final.

4.3 Prospecção estrutural dos sítios sem consenso linear na estrutura dos substratos

Na estrutura cristalográfica da PKA em complexo com o PKI (Figura 1) é possível observar claramente uma organização local no SBS que favorece a interação das duas estruturas, com os resíduos R18 e R19 (posições P-3 e P-2) do PKI interagindo com resíduos de carga complementar que formam os bolsões S-3 (E127) e S-2 (E170, E203, E230).

O caráter eletrostático de S-2 e S-3 sugere que o substrato para PKA apresente âncoras em P-2 e P-3 carregadas positivamente, para compensar o grande número de cargas negativas dos subsítios. Dessa forma, se essas âncoras não são localizadas próximas ao resíduo a ser fosforilado na sequência de aminoácidos, esses resíduos deveriam ser providos por regiões distantes do sítio de fosforilação, mas estruturalmente compatíveis com as requisições estruturais impostas pela PK. Estes sítios de fosforilação com tais características são chamados neste trabalho de sítios de fosforilação conformacionais. Portanto, encontrar uma coerência na organização local para os sítios conformacionais envolve uma varredura da superfície dos substratos nos entornos dos sítios de fosforilação em busca de uma conformação com características favoráveis para interagir com a PK.

A heurística utilizada assume que o substrato PKI de PKA tem uma configuração em relação ao SBS da PK considerada ideal. A partir dessa conformação ideal pode-se extrair vários vetores que caracterizam relações espaciais favoráveis a interação com a PK. Assim, foram extraídas informações da disposição dos resíduos das posições P0 (Ala), P-2 (Arg) e P-3 (Arg) de PKI. Considerando os $C\alpha$ desses resíduos, temos que a distância entre o P0 e P-2 é de 6.616 Å, entre P0 e P-3 é de 9.958 Å e entre P-2 e P-3 é de 3.891 Å.

Em seguida, estabelecida a métrica de um substrato ideal, foi realizada uma busca na região ao redor dos sítios sem consenso de fosforilação linear por resíduos básicos (arginina ou lisina) que apresentassem uma disposição estrutural semelhante ao do PKI. Como estão sendo considerados dois resíduos de cadeias laterais longas, em ambos os casos a distância entre o $C\alpha$ e o último átomo da cadeia lateral é de cerca de 5 Å, e ainda contando a eventual flexibilidade da estrutura do substrato, concluiu-se ser aceitável a realização de uma busca com flexibilidade em relação ao PKI. A busca foi feita então por resíduos básicos a uma distância menor que 15 Å do resíduo fosforilado. Foram registrados tanto os casos que apresentavam apenas um resíduo básico dentro da restrição de distância, com potencial portanto para formar um motivo de fosforilação degenerado com apenas uma âncora, quanto os casos em que haviam dois resíduos básicos dentro do raio de busca. Para determinar os casos em que dois resíduos básicos poderiam ser âncoras simultâneas formando o motivo de fosforilação completo, foi imposta a condição que um resíduo âncora precisava estar a menos de 5,5 Å de distância do outro. Um exemplo de busca no entorno

32 Capítulo 4. Método

de um sítio pode ser visto na figura 3. Para cada sítio de fosforilação cadastrado no banco de dados consultado, foi registrada uma listagem de todas possibilidades de configuração de resíduos âncoras encontradas.

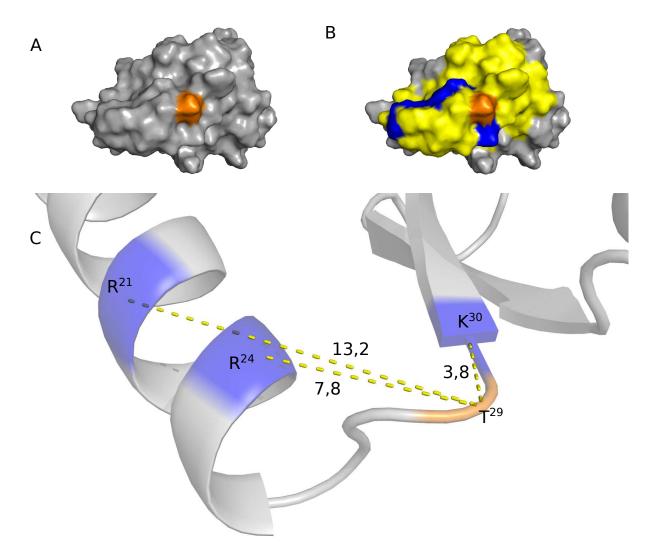


Figura 3 – Exemplo da busca por possíveis âncoras para a proteína Microtubule-associated proteins 1A/1B light chain 3B. A - Visão geral da proteína com o resíduo fosforilado T29 destacado em laranja. B - Em amarelo, resíduos expostos há uma distância de 15 Å e em azul os resíduos básicos encontrados nessa região. C - Visão detalhada na região em torno do sítio, mostrando as distâncias entre o $C\alpha$ do resíduo fosforilado e os $C\alpha$ dos três resíduos básicos encontrados como potenciais âncoras.

Uma restrição essencial foi excluir das buscas, tanto para o sítio quando potenciais âncoras, resíduos que apresentavam uma acessibilidade ao solvente nula, isso é, resíduos enterrados na estrutura, uma vez que seria difícil inferir como os mesmos poderiam ser utilizados para interagir com a PK.

Todo esse processo foi realizado através de um *script* em *Python*. As posições dos átomos foram extraídas diretamente dos arquivos de coordenadas do tipo PDB da

estrutura dos substratos, sendo o cálculo de distância todo feito utilizando funções nativas do *PythonLL*. O cálculo das áreas de acessibilidade ao solvente utilizou a biblioteca *Python* disponível no YASARA.

4.4 Gerando modelos de interação com a PK

4.4.1 Visão geral

Uma vez obtidas as estruturas dos substratos que apresentam sítios sem o consenso de fosforilação na estrutura primária e tendo sido catalogado possíveis resíduos para serem usados como âncoras no processo de fosforilação desses sítios, foram criados modelos estruturais de complexos PK-substrato utilizando um processo de docking proteína-proteína com restrições. O objetivo da construção desse complexo é compreender a viabilidade estrutural desses substratos de interagirem com a PK seguindo a orientação da interação adotada como modelo (interação PKA-PKI), ou seja, avaliar se é possível para o resíduo fosforilado acessar o bolsão S0 do SBS enquanto os resíduos âncoras preditos podem ocupar os bolsões S-2 e S-3, de maneira análoga ao que ocorre em uma estrutura de substrato contendo o consenso de fosforilação. Como esse é um problema muito específico, com alto grau de restrição, em que se deseja que a interação siga uma conformação conhecida, optou-se pelo desenvolvimento de um algorítimo próprio para a realização dessa análise.

Todo o protocolo foi desenvolvido na linguagem *Python*. Todas as operações de manipulação espacial foram desenvolvidas utilizando conceitos de geometria analítica. Para tal, foi utilizado a biblioteca NumPy, que conta com um vasto conjunto de operações matriciais. Apenas para a operação de minimização de energia foi utilizado um protocolo já disponibilizado pelo *software* YASARA.

4.4.2 *Docking* geométrico de PKA-substrato guiado pelo modelo experimental PKA-PKI

A prospecção estrutural feita no passo anterior resulta em uma lista contendo, para cada sítio de fosforilação, uma combinação de todos os resíduos passíveis de servirem de âncora na interação. O primeiro passo do protocolo de gerar modelos estruturais dos complexos é encontrar uma conformação em que os sítios e as âncoras se posicionem em relação ao SBS da PK, o mais próximo possível da conformação ideal de referência, o peptídeo PKI ancorado em PKA.

O algorítimo desenvolvido para tal consiste em inicialmente ler e guardar as coordenadas dos átomos da PK, do peptídeo referência e do substrato. Para facilitar as operações no espaço, o primeiro movimento realizado é transladar o substrato no espaço de modo que o $C\alpha$ do resíduo fosforilado (origem da translação) se posicione na posição de

34 Capítulo 4. Método

coordenadas [0,0,0] (destino da translação). Essas operações são explicitadas nas equações 4.1 e 4.2.

$$[\Delta_x, \Delta_y, \Delta_z] = [RefOri_x - RefDest_x, RefOri_y - RefDest_y, RefOri_z - RefDest_z]$$

$$(4.1)$$

$$\forall \text{\'Atomo} \in \text{Estrutura}, Nova Posição \\ \text{\'Atomo}_{z} - \Delta_{z}, \\ \text{\'Atomo}_{y} - \Delta_{y}, \\ \text{\'Atomo}_{z} - \Delta_{z}]$$
 (4.2)

Na sequência, a estrutura é rotacionada em torno do eixo z até que o $C\alpha$ do resíduo a ser utilizado como âncora fique posicionado no plano xz, sendo então rotacionado no eixo y até que o $C\alpha$ do resíduo âncora esteja posicionado sobre o eixo x. O processo de rotação está descrito pelas equações 4.3 e 4.4. Assim, temos o vetor $\overline{C\alpha_{sitio}C\alpha_{ancora}}$ posicionado sobre o eixo x. Mesmo para casos em que haviam duas âncoras disponíveis, o algorítimo manipulava uma de cada vez. A opção por tratar o problema com apenas uma âncora de referência deu pelo fato que a maioria dos casos encontrados possuem apenas uma âncora possível. Além disso, mesmo quando haviam duas âncoras possíveis, como garantimos a proximidade desses resíduos na busca inicial por âncoras, ao alinhar um na região do SBS, garantimos que o outro se posicione próximo deste, ainda na região do SBS.

Considerando o vetor formado pelo $C\alpha$ do resíduo âncora e sendo \overrightarrow{PA} a projeção desse vetor no plano perpendicular ao eixo de rotação, com \overrightarrow{ER} o eixo de rotação e \overrightarrow{VO} o vetor objetivo no qual queremos sobrepor o nosso $C\alpha$, que no caso é o eixo X.

Definição do ângulo θ de rotação:

$$\theta = \arccos\left(\frac{\overrightarrow{PA} \cdot \overrightarrow{VO}}{\|\overrightarrow{PA}\| \|\overrightarrow{VO}\|}\right) \tag{4.3}$$

Então, \forall átomo \in estrutura, considerando apenas as coordenadas não relacionadas ao eixo de rotação, que são mantidas:

$$NovasCoordenadas_{\acute{a}tomo} = Coordenadas_{\acute{a}tomo} \cdot \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$
(4.4)

Em seguida a estrutura do complexo PKA com o PKI é movida para o mesmo plano espacial do substrato. Para tanto, as estruturas do complexo são transladadas de modo que o $C\alpha$ do resíduo da posição P0 do PKI (A21) se posicione na origem do espaço [0,0,0], ficando assim sobreposto ao $C\alpha$ do resíduo fosforilado do substrato. A PKA é transladada mantendo a posição relativa ao PKI. Em seguida, o conjunto de coordenadas das duas estruturas são rotacionadas de forma análoga a feita ao substrato, de modo que o $C\alpha$ da âncora R19 do PKI se posicione também sobre o eixo x.

Ao final desses passos, os resíduos relevantes na interação estão espacialmente posicionados na região do SBS da PKA. Porém ainda precisamos ajustar o posicionamento da PK de modo a tornar coerente a interação do ponto de vista da estrutura completa e não só localmente. Para isso é feita uma rotação da PK em torno do eixo x. A meta é encontrar a configuração que minimiza a quantidade de átomos sobrepostos entre a PK e o substrato, ou seja, uma configuração o mais realista possível que mantenha a organização local favorável à ocorrência de fosforilação. A configuração encontrada nesse passo é definida como a posição final do docking geométrico, sendo o complexo salvo em um arquivo do tipo PDB. Todos esses passos estão visualmente ilustrados na figura 4.

4.4.3 Remoção de impedimentos estéricos do complexo PK-substrato

Como no passo anterior buscava-se apenas acertar o posicionamento dos resíduos na região do SBS de acordo com a disposição modelo do PKI, utilizando apenas operações geométricas, os complexos gerados podem apresentar problemas de sobreposições entre átomos da PK e do substrato. Esses problemas variam desde pequenas sobreposições de átomos de cadeias laterais, facilmente solucionáveis, até casos mais complexos em que elementos de estrutura secundária de uma proteína se apresentam totalmente sobrepostos (enterrados) na outra proteína.

Para corrigir esses choques que certamente inviabilizam a qualidade do modelo estrutural do complexo, foi desenvolvido um protocolo em que busca-se recriar a estrutura gerada nos passos da seção 4.4.2, porém introduzindo um elemento de flexibilidade torcional, visando ajustar e evitar essas sobreposições formadas. Esse protocolo consiste em definir uma posição inicial, anterior a interação, isso é, em que ambas as estruturas estão próximas, mas ainda sem interagir e então aproximar as duas estruturas, guiando-se pela estrutura final do docking geométrico, permitindo ajustes nas estruturas individuais, visando evitar choques.

Para tal definimos um vetor entre o centro geométrico da estrutura da PK e o ponto médio entre o $C\alpha$ do sítio de fosforilação e o $C\alpha$ do resíduo âncora do substrato. A estrutura da PK é então transladada na direção desse vetor iterativamente, em intervalos de 1 Å. A iteração termina quando não existirem mais átomos sobrepostos. Por questão de eficiência computacional, apenas os $C\alpha$ foram considerados na verificação de colisões.

36 Capítulo 4. Método

4.4.4 Geração de conformações termodinamicamente viáveis para o complexo modelado PK-substrato

O último passo do protocolo consiste em aproximar as estruturas da PK e do substrato, partindo da conformação inicial, em que não há interação direta, seguindo a orientação definida pelo *docking* rígido baseado no modelo do PKI. O principal objetivo desse último passo é ajustar as estruturas formadoras do complexo, que podem apresentar conformações locais termodinamicamente desfavoráveis.

A seleção de conformações termodinamicamente viáveis foi realizada aplicando-se uma minimização de energia enviesada, partindo-se da estrutura sem impedimentos estéricos, tentando atingir a estrutura do complexo com interação. Para esse procedimento, foi utilizada a função morphing do YASARA (Figura 5). Como na definição da estrutura inicial somente foi avaliada a sobreposição dos Cα das duas estruturas, podem existir eventuais sobreposições envolvendo os demais átomos das proteínas. Portanto o primeiro passo foi realizar uma minimização de energia utilizando o algorítimo de descida mais íngreme (Steepest Descent) (DEBYE, 1909), que é computacionalmente eficiente para remover essas sobreposições iniciais. Nessa primeira minimização não é utilizada a componente eletrostática da interação, para evitar que o método fique preso em regiões de mínimos artificiais, causados pela superestimação energética da componente eletrostática em interações de curto alcance. Em todo o processo é utilizado o campo de força YAMBER3 (KRIEGER et al., 2004), cuja parametrização visa a manutenção estrutural de proteínas em dinâmicas longas e o aperfeiçoamento de estruturas geradas por modelagem comparativa.

Um caminho de baixa energia entre as duas conformações é definido por meio de uma minimização de energia realizada com o algorítimo Simulated Annealing (KIRKPATRICK; GELATT; VECCHI, 1983). Foram utilizadas as configurações padrão definidas pelo YASARA, com o passo de 2 fs e as velocidades dos átomos diminuídas por um fator de 0.9 a cada 10 passos. Uma força de 5 pN é aplicada em todos os átomos em direção á posição definida como final. Foi utilizada uma caixa não periódica, uma vez que surgem dificuldades em definir a direção da força a ser aplicada no átomo que atravessa a barreira periódica. O RMSD da estrutura era então comparada com a posição final do complexo a cada 10 passos. Sempre que a melhora do RMSD fosse menor do que 0,01 Å, a força então era aumentada em 50% até um máximo de 25 pN. Esses valores de força foram definidos empiricamente por serem prováveis de não causarem grandes distorções na estrutura, uma vez que são compatíveis com valores exercidos em interações in vivo entre proteínas, que alcançam valores de até algumas centenas de picoNewtons (SIMON; PESKIN; OSTER, 1992).

4.5 Avaliação dos modelos estruturais dos complexos gerados

O principal objetivo da etapa de montagem da estrutura do complexo PK-substrato é avaliar se o substrato é capaz de acomodar o sítio de fosforilação no SBS da PK em uma conformação análoga ao do PKI enquanto ainda mantém uma coerência estrutural. Para avaliar a manutenção da integridade estrutural das proteínas isoladas do complexo, foi realizada uma comparação individual das estruturas envolvidas na interação em sua conformação original e na conformação após a montagem do complexo. A métrica escolhida para a comparação foi o desvio da raiz do valor quadrático médio, conhecido pelo seu nome em inglês *Root Mean Square Deviation* (RMSD). Essa é uma métrica usual na comparação de estruturas proteicas, podendo ser definida pela equação 4.5. Os cálculos de RMSD foram realizados utilizando funções do YASARA..

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \tag{4.5}$$

n = número de átomos

 $d_i = \text{distância}$ entre dois átomos i correspondentes nas estruturas sobrepostas.

As sobreposições estruturais foram realizadas utilizando métodos implementados na biblioteca YASARA do *Python*. O alinhamento das estruturas utiliza o método próprio MUSTANGPP, que é uma extensão do método MUSTANG (KONAGURTHU et al., 2006), adicionando um pós-processamento aos resultados. Partindo da sobreposição gerada pelo MUSTANG, o YASARA extraí os resíduos que podem ser considerados estruturalmente alinhados e realiza a sobreposição apenas nesses resíduos. Então uma nova iteração do processo é realizada, até o número de resíduos alinhados convergir para um máximo. Com isso, a sobreposição é focada em resíduos que são estruturalmente equivalentes.

Para uma análise completa da capacidade desses substratos em formar os complexos previstos, foram realizadas três comparações com distintos objetivos. Primeiramente calculou-se o RMSD de toda estrutura, tanto da PK quanto do substrato, por ser a medida comparativa mais usual e por ser capaz de determinar a similaridade global das estruturas obtidas. Calculou-se também o RMSD focado nas regiões propostas como responsáveis por orientar a interação com a PK. Para isso, restringiu-se o cálculo do RMSD nas estruturas sobrepostas em janelas de 10 resíduos em torno do sítio fosforilado e em torno dos resíduos propostos de atuar como âncoras. Isolar a comparação a apenas essas regiões, permitiu avaliar se a montagem do complexo exigia uma demasiada modificação estrutural na região de interação. Por fim, utilizando o posicionamento do PKI na fenda catalítica, foi feito um cálculo de RMSD apenas com os átomos $C\alpha$ entre os resíduos das posições P0, P-2 e P-3 do PKI e os resíduos fosforilados e os propostos a atuarem como âncoras do substrato, para

38 Capítulo 4. Método

avaliar o quão bem esses resíduos conseguem emular o posicionamento de uma sequência linear na conformação final.

4.6 Conservação dos resíduos âncoras em sítios conformacionais

Alguns sítios de fosforilação apresentam múltiplos resíduos que atendem aos critérios de carga e distância estabelecidos como mínimos para a atuação deste como âncoras. Considerando as combinações possíveis desses resíduos âncoras, uma vez que as PKs estudadas admitem dois resíduos âncoras atuando simultaneamente, é possível propor um elevado número de configurações de interações desses sítios com a PK.

Como uma forma de avalizar a relevância biológica desses resíduos, foi feita uma avaliação da conservação evolutiva. Para tal, foi realizado uma busca por sequências de proteínas homologas a do substrato estudado, utilizando o algoritmo de alinhamento blastp (CAMACHO et al., 2009) buscando sequências no banco de sequências não redundantes RefSeq (PRUITT, 2004). Foram selecionadas todas as sequências com e-value menor que 10^{-10} , removendo sequências provenientes de construções sintéticas existentes no banco de dados consultado.

Uma vez já tendo sido feita a prospecção estrutural nos substratos, listando sítios sem o consenso de fosforilação com os possíveis resíduos âncoras formadores de sítios conformacionais, foi feita uma análise desses resíduos no contexto do alinhamento com as sequências homologas encontradas. Foram gerados arquivos de sequências contendo, para cada sequência encontrada, apenas os resíduos participantes dos sítios conformacionais. Para efeito de comparação, foram gerados arquivos análogos contendo apenas informações de conservação dos resíduos serina, treonina, lisinas e argininas presentes no substrato e não participantes de nenhum sítio de fosforilação conhecido, o que consideramos um grupo de controle. Esses arquivos foram utilizados para a avaliação qualitativa da conservação desses resíduos, que consistiu em gerar weblogos (CROOKS, 2004) com as sequências dos residuos dos sítios conformacionais e do grupo controle.

A avaliação quantitativa foi feita utilizando o serviço web de cálculo de conservação de resíduos em proteínas, disponibilizado pelo departamento de ciência da computação da Universidade de Princeton (CAPRA; SINGH, 2007).

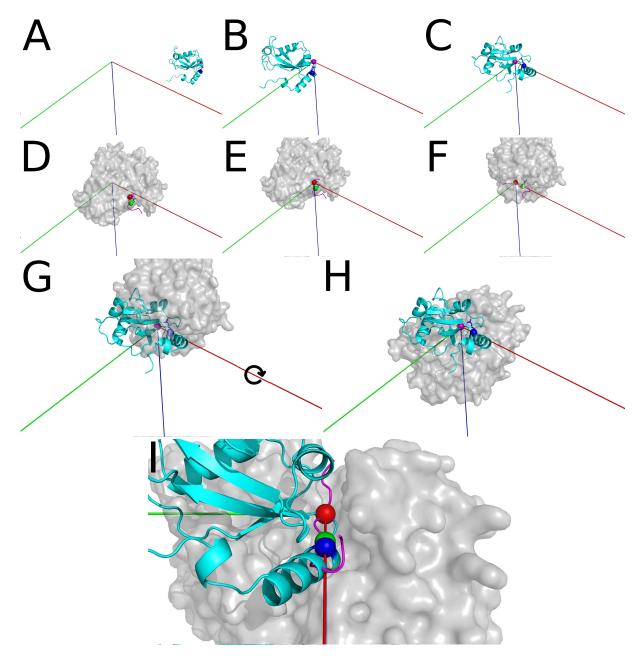


Figura 4 – A - Posição inicial do substrato Microtubule-associated proteins 1A/1B light chain 3B em relação aos eixos x (vermelho), y (azul) e z (verde). Em destaque o $C\alpha$ em representação esférica do resíduo fosforilado (roxo) e do resíduo âncora (azul). B - O substrato transladado sobrepondo o $C\alpha$ do sítio de fosforilação na origem do espaço. C - O substrato rotacionado para sobrepor o $C\alpha$ da âncora sobre o eixo x. D - Complexo PKA (cinza) e PKI (cartoon magenta) na posição original. No PKI, destacados na representação esférica os $C\alpha$ dos resíduos P0 (vermelho) e P-2 (verde). E - O complexo transladado colocando o P0 do PKI na origem do espaço. F - O Complexo rotacionado para orientar o $C\alpha$ da âncora sobre o eixo x. G - Estruturas da PK e do substrato com suas regiões de interação orientadas sobre o eixo x. Agora a PK é rotacionada no eixo x em busca da conformação com o menor número de colisões. H - A posição final encontrada. I - Visão detalhada para comparação da posição final do substrato e do PKI sobrepostos.

40 Capítulo 4. Método

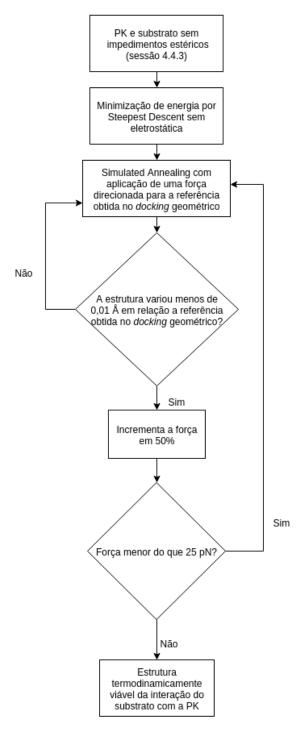


Figura 5 – Workflow do algoritmo utilizado para determinar uma conformação termodinamicamente viável do complexo da PK com o substrato partindo dos modelos gerados nas sessões 4.4.2 e 4.4.3

5 Resultados e Discussão

5.1 Abundância de sítios de fosforilação conformacionais permite estudos estruturais em larga escala

Na prospecção da base de dados experimentais de sítios de fosforilação de proteínas, foram identificados 1106 sítios registrados como alvos da proteína PKA. Esses sítios se distribuem por 625 proteínas diferentes.

Analisando a sequência de aminoácidos desses sítios, constatamos que 492 (44,5%) correspondem ao motivo canônico, com resíduos básicos em P-2 e P-3 ([RK][RK]X[ST]), enquanto 474 (42,9%) apresentam o motivo menos estringente, com um resíduo básico em, pelo menos, uma das duas posições de ancoragem consideradas. Com isso, temos que 966 (87,3%) apresentam algum resíduo básico em sua sequência nas posições de ancoragem, restando então um contingente de 140 sítios (12,7%) sem nenhum resíduo na sequência linear dentre os esperados para ser utilizado como âncora na interação.

Dos substratos relatados, 578 (92,5%) foram modelados por homologia. Dos 1106 sítios para PKA, 523 (47,3%) tiveram suas estruturas tridimensional elucidadas. A diferença é explicada pela falta de estruturas para serem utilizadas como modelo comparativo, e nos casos modelados, devido à cobertura incompleta de algumas sequências de substratos por moldes. Desses sítios modelados, 486 apresentavam algum tipo de motivo linear (50,3% dos sítios lineares disponíveis no banco de dados) e 37 não apresentavam a conformação consensual em sua estrutura primária (26,5% dos sítios sem motivo linear disponíveis no banco de dados). Informações sobre a modelagem dos substratos contendo sítios sem consenso linear podem ser vistos na tabela 6.

5.2 O conceito de sítio conformacional

A estrutura do SBS da PKA é bem caracterizada, tendo sido esse dividido em subsítios para fins de estudos de afinidades com peptídeos (KOBE et al., 2005). Considerando o consenso canônico de fosforilação dessa quinase ([RK][RK]X[ST]), constatamos que os subsítios essenciais para a interação são os S-3, formados pelos resíduos E127 e Y330, e o subsítio S-2, que conta com os resíduos T201, E203, Y204, E170 e E230. É possível constatar que ambos apresentam um predominante teor de resíduos com cargas negativas. Essa constatação é coerente com o consenso de fosforilação da PKA, que apresenta resíduos com cargas positivas nas posições de interação P-2 e P-3 (figura 6A).

O PKI, um inibidor natural da PKA, forma um complexo termodinamicamente

muito favorável ($K_i = 0.2$ nM) (PAULUCCI-HOLTHAUZEN; O'CONNOR, 2006). Consequentemente a disposição dos resíduos de P0, P-2 e P-3 nesse peptídeo apresenta uma configuração ideal para interagir com a PKA. Essa configuração pode ser vista na figura 6. Não por acaso, a maioria dos sítios de fosforilação apresenta algum resíduo básico em sua estrutura primária nas posições P-2 e P-3. Sítios nessas condições apresentam naturalmente uma estrutura muito similar à do PKI.

Um sítio conformacional se caracteriza por apresentar uma estrutura tridimensional que mimetiza a disposição espacial do PKI, pelo menos em relação aos resíduos considerados mais importantes na interação com a quinase. Ou seja, desconsiderando a disposição na estrutura primária, avalia-se o posicionamento espacial do resíduo fosforilado e dos resíduos carregados na superfície da proteína, com base na distância dos $C\alpha$ desses resíduos. Se houver semelhança na configuração desses resíduos com a do PKI, pode-se supor que essa disposição espacial faria com que o sítio de fosforilação, independente da sequência na região em que se encontra, apresentasse a estrutura do consenso de fosforilação. Consequentemente esse sítio apresentaria uma configuração favorável para interagir com uma quinase que interaja com o PKI, gerando portanto um potencial sítio conformacional.

Dessa foma, para buscar sítios conformacionais, todos os sítios sem consenso de fosforilação modelados foram submetidos a uma prospecção estrutural por resíduos âncoras (carregados positivamente). Dos 37 sítios conformacionais, 34 apresentam ao menos uma configuração que atende os critérios definidos como mínimos para definir um sítio conformacional. Desses sítios, 16 apresentam a possibilidade de formar o motivo de fosforilação com um ou dois resíduos âncoras, enquanto os demais apresentam possibilidades com um único resíduo âncora.

Dentre os três sítios que não apresentam possibilidades de resíduos âncoras na estrutura do substrato, notamos que um dos substratos, a proteína *Myosin regulatory light*

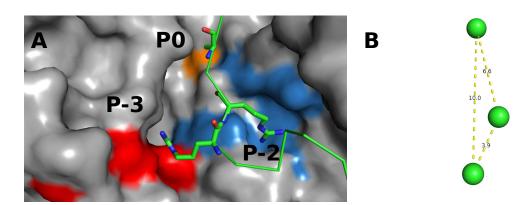


Figura 6 – A - SBS da PKA (cinza) com os subsítio S-3 destacado em vermelho, S-2 em azul e S0 em laranja. Em verde o PKI com destaque para os resíduos que interagem nesses subsítios. B - Visão isolada do $C\alpha$ dos resíduos de P0, P-2 e P-3 do PKI, com as suas respectivas distâncias.

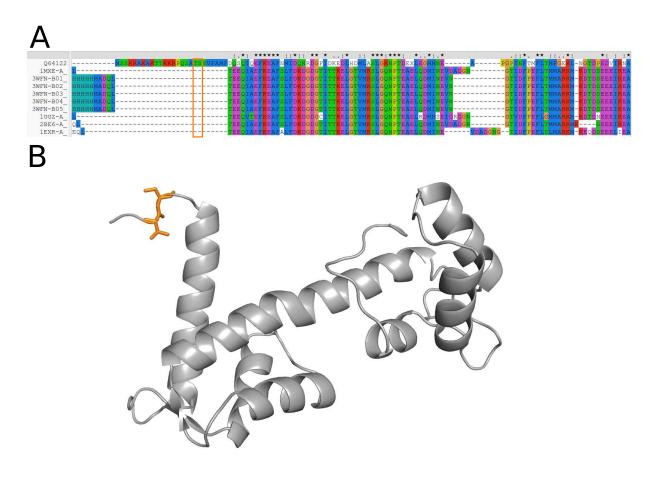


Figura 7 – A - Alinhamento da sequência do substrato com os moldes utilizados. Destacado com um retângulo laranja os dois resíduos fosforilados (T19 e S20), que já se encontra em uma região que não se alinha em nenhum molde e portanto não apresentam uma modelagem confiável. Na região entre o N-terminal e os sítios, podemos ver a existência de sete resíduos básicos próximos aos sítios de fosforilação. B - Modelo gerado com os dois resíduos fosforilados destacados em *sticks* laranja dispostos no limite da região modelada.

polypeptide 9 (UNIPROT: Q64122), contendo dois sítios de fosforilação juntos na sequência (T19 e S20), não foi plenamente modelado, tendo 90,6% da sequência modelada (figura 7). Os sítios ficaram no limiar da região modelada sendo que na região não modelada, próxima ao N-terminal, existem uma série de resíduos básicos que possivelmente poderiam servir como âncoras no processo de atracamento (K4, R5, K7, K9, K12, K13, R14). Com isso, podemos considerar que apenas um dos sítios sem motivo de fosforilação explicito na estrutura primária definitivamente não apresenta uma configuração plausível de motivo estrutural de fosforilação pelos nossos critérios. Esse sítio, S182 no substrato *Vitamin D3 receptor* (UNIPROT: P11473), apresentou uma modelagem de qualidade baixa na região do sítio de fosforilação. Portanto, a análise da estrutura nessa região do modelo apresenta um resultado pouco confiável, tornando impossível excluir esse caso como um sítio conformacional (figura 8).

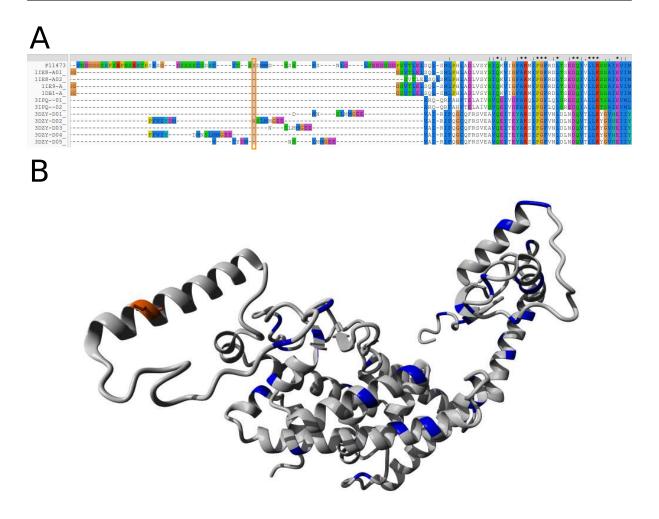


Figura 8 – A - Alinhamento da sequência do substrato com os moldes utilizados. Destacado com um retângulo laranja o resíduo fosforilado (S182). B - Modelo gerado com o resíduo fosforilado destacados em *sticks* laranja e em azul todas as possíveis âncoras. O substrato apresenta muitos resíduos básicos em sua superfície. Mesmo que em nosso modelo nenhum esteja disposto de forma a ser utilizado como âncora para o sítio conhecido, a baixa qualidade da modelagem nessa região torna pouco confiável a estrutura nessa região e portanto as distâncias calculadas, tornando impossível descartar esse sítio como um possível sítio conformacional.

Para comparação e validação dos critérios utilizados na busca estrutural de âncoras, efetuamos a mesma busca para 486 sítios que apresentam o motivo de fosforilação linear. Para 455 casos, foi encontrado na busca estrutural os resíduos que constituem o consenso sequencial. Dos 31 casos restantes, 27 apresentaram alguma conformação de sítio estrutural plausível. Desses dados podemos concluir que os critérios de busca utilizados estão de acordo com o esperado da estrutura de um consenso de fosforilação, uma vez que a busca estrutural encontrou 93,1% dos consensos lineares existentes nos sítios modelados. Outra conclusão desse estudo é que em determinados substratos, os resíduos básicos que estão próximos ao sítio de fosforilação na sequência de aminoácidos, podem se encontrar inacessíveis na

estrutura do substrato, sendo que mesmo nesses casos, podem existir resíduos distantes na estrutura primária, porém próximas na terciária, e que podem ser essenciais para a ancoragem do substrato na PK. Isso pode implicar que sítios conformacionais sejam mais abundantes do que os encontrados neste trabalho.

5.3 Analise estrutural dos complexos PK-substratos modelados com sítios conformacionais

Para os 34 sítios sem motivo linear, foram gerados 175 modelos de complexo de ligação com a PKA, utilizando diferentes configurações das âncoras encontradas. A relação completa de modelos gerados, bem como os dados de análise na íntegra, podem ser vistos na tabela 7. Analisando o RMSD da estrutura completa, tanto a PK quanto o substrato apresentam pouca variação, com uma média de 1,38 Å para a PKA e 1,61 Å para o substrato. Já o RMSD da região do sítio e das âncoras apresentam distorções maiores, com média de 4,67 Å para a janela em torno do sítio e 4,27 para as âncoras. Analisando o desvio padrão dessas medidas, notamos que o RMSD das estruturas como um todo apresentam uma distribuição mais uniforme, com os valores calculados para a PK apresentando um valor de 0,40 Å e para o substrato de 0,52 Å. Já para as janelas, o desvio padrão foi bem mais elevado, com o RMSD em torno do sítio apresentando um desvio padrão de 4,62 Å e em torno das âncoras 5,04 Å. Já quando comparamos a conformação final do sítio conformacional com a estrutura referência do PKI, o RMSD médio encontrado é de 10,14 Å com um desvio padrão de 4,06 Å. A distribuição desses dados está ilustrada na figura 9.

Os maiores valores de RMSD encontrados quando comparado isoladamente as regiões do sítio e das âncoras, é algo esperado, uma vez que essas são as regiões que estão na interface do complexo, e portanto as que apresentam real necessidade de ajuste na conformação com a PK. Com isso o RMSD nessas regiões se torna bastante sensível, gerando valores muito maiores se comparados ao cálculo do RMSD global.

Os valores de RMSD são um indicativo esparso, servindo apenas como uma indicação da viabilidade do modelo definido, porém insuficiente para definitivamente excluir ou aceitar um modelo. Modelos com elevado RMSD podem indicar que a conformação proposta de âncoras não está correta, uma vez que o substrato ou não foi capaz de se ajustar adequadamente no SBS, indicado por um alto RMSD na comparação com o PKI, ou necessitou demasiadas alterações na estrutura no substrato, resultando em um alto RMSD nas comparações com a estrutura original. Porém mesmo esses modelos não podem ser completamente descartados, uma vez que tais complexos podem ser viáveis em contexto biológico, por exemplo, viabilizado por mudanças conformacionais não triviais de serem computacionalmente consideradas. Por outro lado, existem sítios que são capazes de formar um complexo com os resíduos essenciais, isso é, o resíduo a ser fosforilado e os

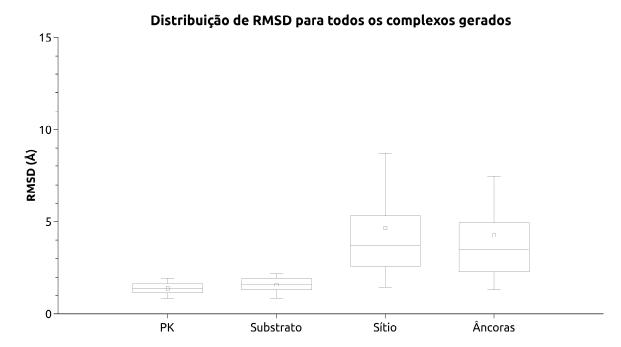


Figura 9 – Distribuição de RMSD calculado para todos os modelos de complexos gerados, considerando a avaliação global da PK, global do substrato, uma janela de 10 resíduos em torno do sítio de fosforilação conhecido e janelas de 10 resíduos em torno dos resíduos positivos passíveis de auxiliar no atracamento da PK, denotados âncoras.

âncoras, se aproximando do posicionamento dos resíduos equivalentes no PKI, sofrendo apenas pequenas alterações na estrutura do substrato. Dessa observação é possível concluir que esses substratos podem se atracar com a PKA na conformação predita. Esse é um forte indicativo de que o consenso estrutural predito pode ser efetivamente essencial no reconhecimento do substrato pela PK.

Três exemplos de substratos de PKA encontrados neste estudo, ilustram bem a significância de sítios conformacionais nos processos de fosforilação por PKA:

A proteína Microtubule-associated proteins 1A/1B light chain 3B (UNIPROT: Q9GZQ8) é fosforilada no resíduo T29, cuja estrutura primária no entorno do sítio é LEU-ILE-ARG-GLU-GLN-HIS-PRO-THR29-LIS-ILE-PRO-VAL-ILE-ILE-GLU. Na nossa prospecção estrutural foram encontrados três resíduos possíveis de serem utilizados como âncoras por PKA: R21, R24 e R30. As duas primeiras possíveis âncoras (R21 e R24) se encontram próximas (5,37 Å), podendo então ser utilizadas como âncoras concomitantemente. Os dados de RMSD para esse substrato podem ser encontrados na tabela 2. Nela podemos observar que as conformações com duas âncoras apresentam RMSD menores e portanto melhores do que a conformação com uma âncora. Todos os modelos apresentaram valores de RMSD aceitáveis para a prospecção das estruturas completas e das janelas. Porém o RMSD em relação ao PKI ficou no melhor caso em 4,61 Å para o modelo 1 com duas âncoras, enquanto para o modelo 3 com uma só âncora esse valor foi 9,76Å

. Como apresentado na sessão de métodos, utilizamos apenas um resíduo para orientar o posicionamento do substrato na fenda catalítica, sendo esse nomeado sempre como P-2. Então a ordem dos resíduos em P-2 e P-3 na nomenclatura dos modelos é apenas um identificador da referência utilizada na construção do modelo, não necessariamente condizente com a estrutura final encontrada. Nesse caso, é possível observar isso, haja visto que a melhor configuração calculada é identificado com o resíduo R21 na posição P-2 e R24 na P-3, mas por meio de uma inspeção visual do modelo formado, pode-se avaliar que o resíduo R24 se ajusta melhor no bolsão S-2 enquanto o resíduo R21 se ajusta bem no bolsão S-3, conformação essa exposta na figura 10.

Tabela 2 – Análise de RMSD para o complexo da PKA com o substrato *Microtubule-associated proteins 1A/1B light chain 3B* para o sítio T29. Para cada possível configuração de âncoras encontradas foi gerado um complexo e realizado o cálculo de RMSD das estruturas antes e depois do processo de atracamento. A comparação foi feita para a estrutura completa da PK, estrutura completa do substrato, para a região do sítio, das âncoras e em referência ao PKI no SBS.

					ı	KMSD		
Modelo	Sítio	P-2	P-3	PΚ	Substrato	Sítio	Âncoras	PKI
1	T29	R21	R24	0,98	0,91	1,10	0,95	4,61
2	T29	R24	R21	0,99	0,99	2,31	$2,\!22$	5,75
3	T29	K30	_	1,39	1,44	2,83	2,85	9,76

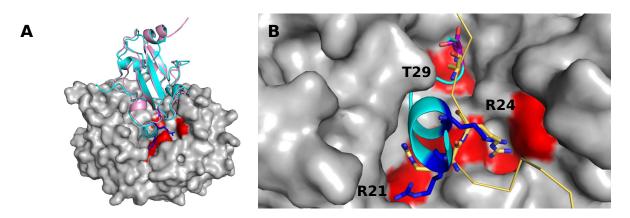


Figura 10 – A - Uma visão geral do complexo formado entre a PKA e o substrato *Microtubule-associated proteins 1A/1B light chain 3B*. Em representação de superfície, a PKA, com os resíduos ácidos que formam as regiões de interação S0, S-2 e S-3 em vermelho. Em *cartoon* azul, a estrutura ajustada no processo de minimização de energia, com os resíduos chaves na interação destacados com as cadeias laterais explicitadas. Em rosa, a estrutura original do substrato após o *docking* rígido, para comparação da deformação. B - Visão em destaque da região de interação no SBS. Em amarelo, o PKI sobreposto, para comparação da conformação do sítio estrutural com esse modelo linear. Identificação dos resíduos do substrato para auxilio na comparação.

A proteína Regulator of G-protein signaling 13 (UNIPROT: O14921), cujo resíduo fosforilado é a T41 tem a seguinte estrutura primária no entorno do sítio: SER-PHE-

GLU-GLN-LEU-MET-ALA-THR41-LYS-TYR-GLY-PRO-VAL-VAL-TYR. A prospecção estrutural apresentou quatro possibilidades de âncoras: R13, K17, R18 e K42. Novamente o melhor resultado considerando o RMSD em comparação com o PKI é uma configuração com duas âncoras, 4,5 Å contra 7,09 Å e 9,56 Å das configurações com uma âncora. Assim como no exemplo anterior, a prospecção visual do complexo gerado (figura 11) apresenta um ajuste melhor dos resíduos âncoras invertidos ao da referência do protocolo, isto é, K17 na posição P-2 e R18 na posição P-3.

Tabela 3 – Análise de RMSD para o complexo da PKA com o substrato Regulator of G-protein signaling 13 para o sítio T41. Para cada possível configuração de âncoras encontradas foi gerado um complexo e realizado o cálculo de RMSD das estruturas antes e depois do processo de atracamento. A comparação foi feita para a estrutura completa da PK, estrutura completa do substrato, para a região do sítio, das âncoras e em referência ao PKI no SBS.

					RMSD			
Modelo	Sítio	P-2	P-3	PΚ	Substrato	Sítio	Âncoras	PKI
1	T41	R13	-	1,28	1,40	2,74	5,62	7,09
2	T41	R18	K17	1,65	1,63	1,73	5,03	4,50
3	T41	K17	R18	1,71	1,82	$2,\!57$	5,20	9,10
4	T41	K42	_	1,35	1,16	1,24	1,15	9,56

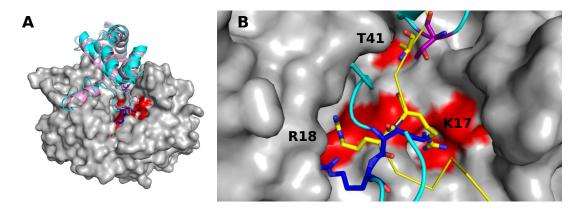


Figura 11 – A - Uma visão geral do complexo formado pela PKA com o substrato Regulator of G-protein signaling 13. Em representação de superfície, a PKA, com os resíduos ácidos que formam as regiões de interação S0, S-2 e S-3 em vermelho. Em cartoon azul, a estrutura ajustada no processo de minimização de energia, com os resíduos chaves na interação destacados com as cadeias laterais explicitadas. Em rosa, a estrutura do modelo original do substrato para comparação. B - Visão em destaque da região de interação no SBS. Em amarelo, o PKI sobreposto, para comparação do conformação do sítio estrutural com esse modelo linear. Identificação dos resíduos do substrato para auxilio na comparação.

A proteína 28S ribosomal protein S29, mitochondrial (UNIPROT: P51398) apresenta fosforilação no resíduo S185, cuja sequência é dada por PHE-ASP-GLN-PRO-LEU-GLU-ALA-SER185-THR-TRP-LEU-LIS-ASN-PHE-LIS. Seis âncoras foram encontradas

na prospecção estrutural. Todas as conformações encontradas apresentam duas âncoras e todos os complexos gerados apresentaram valores de RMSD próximos (considerando o calculo em relação ao PKI, entre 9,46 Å e 11,36 Å), dificultando a escolha da conformação mais favorável para ancoragem de PKA. Para ilustrar o ajuste dessa estrutura, escolhemos o modelo 5 com as âncoras K189 e K192 (figura 12).

Tabela 4 – Análise de RMSD para o complexo da PKA com o substrato *ribosomal protein S29, mitochondrial* para o sítio S185. Para cada possível configuração de âncoras encontradas foi gerado um complexo e realizado o cálculo de RMSD das estruturas antes e depois do processo de atracamento. A comparação foi feita para a estrutura completa da PK, estrutura completa do substrato, para a região do sítio, das âncoras e em referência ao PKI no SBS.

					RMSD			
Modelo	Sítio	P-2	P-3	PK	Substrato	Sítio	Âncoras	PKI
1	S185	R177	K175	1,30	1,92	7,32	7,15	10,51
2	S185	R213	K212	1,29	$1,\!65$	7,69	$6,\!35$	11,36
3	S185	K175	R177	1,68	$2,\!11$	5,75	5,48	10,52
4	S185	K189	K192	1,30	1,56	4,23	3,59	9,85
5	S185	K192	K189	$1,\!53$	1,62	3,31	2,17	9,46
6	S185	K212	R213	1,33	1,30	8,71	6,78	10,75

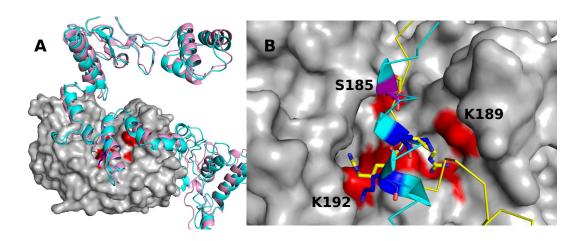
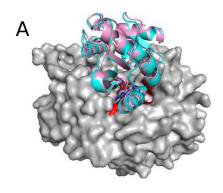


Figura 12 – A

- Uma visão geral do complexo formado pela PKA com o substrato ribosomal protein S29, mitochondrial. Em representação de superfície, a PKA, com os resíduos ácidos que formam as regiões de interação S0, S-2 e S-3 em vermelho. Em cartoon azul, a estrutura ajustada no processo de minimização de energia, com os resíduos chaves na interação destacados com as cadeias laterais explicitadas. Em rosa, a estrutura original do modelo de substrato. B - Visão em destaque da região de interação no SBS. Em amarelo, o PKI sobreposto, para comparação da conformação do sítio estrutural com esse modelo linear. Identificação dos resíduos do substrato para auxílio na comparação.

5.4 Analise estrutural dos complexos quinase-substratos modelados com sítios com o consenso linear

Sítios que possuem a sequência do consenso linear, evidentemente apresentam uma possibilidade de encaixe natural com o SBS. Isso é esperado, já que estruturalmente esses sítios apresentam forte semelhança com o PKI, visto que considerando o espaçamento entre as âncoras e o sítio de fosforilação (isso é, uma distância de um ou dois resíduos), existe um limitado grau de liberdade, fazendo com que essas estruturas apresentem características semelhantes. Um caso ideal, em que todos os resíduos envolvidos na interação estão expostos na estrutura do substrato e todos estão em uma região de *coil* pode ser vista na figura 13, na qual é possível constatar a alta similaridade dessa configuração com o PKI.



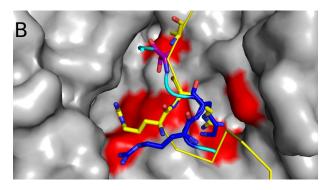


Figura 13 – A

- Visão geral do docking do substrato Glia maturation factor beta, em azul, com a PKA (cinza). Em rosa, a estrutura original do modelo para comparação. B - Visão focada na região de interação das duas estruturas com o PKI (amarelo) sobreposto para referência. Em roxo o resíduo fosforilado, T27, com os resíduos K25 e R24 em azul.

Porém, nem todos os sítios que apresentam essa sequência possuem as características estruturais ideais para a interação com a PK. Um exemplo disso seria no caso em que o sítio de fosforilação e os resíduos âncoras se encontram em uma α -hélice. Na figura 14, vemos o sítio da proteína Bcl2-associated agonist of cell death (UNIPROT: Q92934), que é fosforilada no resíduo S118. A sequência no entorno desse sítio é dada por GLY-ARG-GLU-LEU-ARG-ARG-MET-SER118-ASP-GLU-PHE-VAL-ASP-SER-PHE, ou seja existem resíduos propícios para atuar como âncoras do ponto de vista tradicional. Porém, mesmo o SBS sendo largo o suficiente para acomodar uma α -hélice, a disposição dos resíduos nessa estrutura secundária não favorece a interação de um consenso linear com a PK. Isso porque na disposição da α -hélice, sendo o sítio a posição P0, o resíduo em P-2 (R115) apresentará a sua cadeia lateral na face oposta da α -hélice, tornando impossível uma interação simultânea dos dois resíduos com o SBS da PK. Baseado nessa observação, pode-se esperar que em casos de sítios dispostos em α -hélice com as âncoras dispostas na mesma alpha-hélice, a disposição sequencial desses resíduos mude, apresentando um espaçamento sequencial maior em relação ao consenso linear, de forma a acomodar o

resíduo fosforilado e as âncoras na mesma face da α -hélice. Esse comportamento pode ser observado na figura 12, que apresenta um encaixe muito bom comparado com o modelo linear tendo os resíduos na α -hélice dispostos em uma configuração P0, P+4 e P+7.

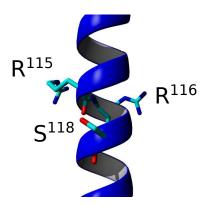


Figura 14 – Sítio da estrutura do substrato Bcl2-associated agonist of cell death. Mesmo apresentando o consenso linear, pode-se observar na estrutura que em uma α -hélice, o sítio e o resíduo em P-2 dificilmente conseguiriam interagir simultaneamente com a PK, por estarem expostos em faces opostas da estrutura

Outra situação que não favorece a interação com o consenso linear é quando as âncoras próximas do sítio na sequência da proteína, se encontram inacessíveis na estrutura do substrato. Um exemplo desse tipo pode ser visto no substrato G1/S-specific cyclin-D1 (UNIPROT: P24385), fosforilado no resíduo S90. Esse sítio, apresenta em sua estrutura primária uma arginina na posição P-3 (R87). Esse resíduo apresenta pouca exposição ao solvente, tendo somente 19,8 Å² de área exposta (figura 15A). Por outro lado, os resíduos K46 e K50, ambos a menos de 15Å de distância em relação ao sítio, se encontram em uma região altamente flexível e estão totalmente expostos (169 Å² e 175 Å² de exposição ao solvente). O modelo de interação da PKA com o substrato, utilizando o resíduo R87 como âncora gerou uma configuração do substrato mal ajustado ao SBS. Comparado com o PKI atracado a PKA, o RMSD dos C α dos resíduos âncora calculado foi de 7,34 Å. No complexo, é possível observar que tanto a âncora quanto o sítio se encontram distantes do SBS. Para aproximar mais esses resíduos na fenda catalítica, seria necessário que a α -hélice que contém esses resíduos se aproximasse mais da fenda catalítica, mas é possível observar que existe um impedimento estérico com a estrutura do próprio substrato, que impossibilita essa aproximação (figura 15B). Já o docking com as âncoras do possível sítio conformacional, gerou uma configuração mais próxima a do PKI (RMSD de 4,5 Å) sem distorcer a estrutura do substrato (figura 15 C e D).

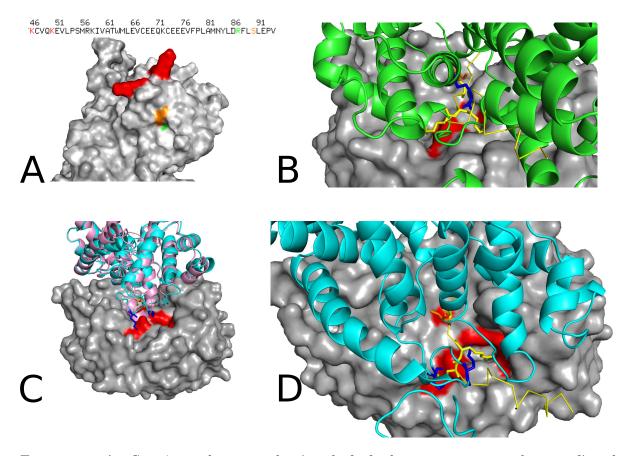


Figura 15 – A - Sequência da região do sítio de fosforilação e uma visão da superfície da proteína, com o resíduo fosforilado em laranja, as possíveis âncoras estruturais em vermelho e a âncora linear disponível em verde. B - Complexo gerado do docking do substrato (verde) coma PKA (cinza) utilizando o sítio (sticks laranja) e a âncora linear R87 (sticks azul). Em amarelo, o PKI sobreposto para referência. C - Visão geral do docking utilizando as âncoras estruturais K46 e K50 (em sticks azul). D - Visão em destaque da interface formada pelo docking de C, com o PKI sobreposto como referência.

5.5 Possíveis aspectos da relevância biológica de sítios conformacionais

A partir da prospecção por possíveis configurações para gerar os modelos das interações, foi possível constatar que para um mesmo sítio de fosforilação podem existir variadas configurações de âncoras capazes de gerar a conformação de um sítio conformacional. No estudo dos substratos de PKA, foram encontrados 34 sítios conformacionais, que geraram 175 configurações de âncoras possíveis. Com isso temos que cada sítio apresenta em média 5 configurações distintas de atracagem com a quinase. Enquanto alguns sítios apresentam apenas uma única configuração, outros apresentam um número bem elevado de configurações possíveis, sendo o máximo encontrado 15.

O sítio com o maior número de possíveis configurações de atracamento foi o resíduo S67 do substrato *Protein phosphatase 1 regulatory subunit 1A*. A fosforilação desse resíduo

foi observada em experimento in vivo, tendo sido provado que a sua fosforilação é crucial na regulação da atividade da fosfatase PP1, essencial em diversos processos celulares (HUANG; PAUDEL, 2000). O segundo sítio com mais conformações, 11 no total, foi o sítio S286 presente na proteína Syndecan-1. Essa é uma proteína de membrana fundamental para a proliferação e migração celular, atuando como um importante receptor na matriz celular. A expressão dessa proteína é induzida pelo $TGF\beta$, sendo regulada pela fosforilação do resíduo em questão (HAYASHIDA et al., 2006). Com 9 conformações possíveis, temos a proteína Microtubule-associated protein light chain 3, fosforilada no resíduo T6. Essa proteína é um marcador específico para autofagia em células mamíferas. A presença de uma quinase capaz de fosforilar esse sítio demonstrou um efeito inibitório no processo autofágico (JIANG et al., 2010).

Considerando brevemente a literatura disponível sobre os três sítios com mais configurações de âncoras possíveis, foi possível constatar que todos possuem uma funcionalidade crucial no ciclo celular. Outros sítios existentes no banco de dados prospectado, apresentam dados bem menos contundentes sobre a importância associada a fosforilação dos mesmos, como, por exemplo, somente evidências de ensaios de larga escala.

Tendo em vista a importância da fosforilação desses resíduos, é possível levantar a hipótese de que esse é um mecanismo evolutivo para conferir robustez a funcionalidade do sítio. Isso porque, mesmo se alguma configuração de âncora for perdida, como, por exemplo, por mutação de algum resíduo âncora, ainda sobram variadas configurações de atracamento da PK, garantindo assim a fosforilação do resíduo.

Para esses três casos expostos, foi feita uma avaliação da conservação evolutiva dos resíduos envolvidos no sítio de fosforilação, juntamente com os resíduos âncoras propostos. Para efeito de comparação, para esses três substratos foi feita a análise da conservação das serinas e treoninas sem dado de fosforilação conhecido, bem como dos resíduos arginina e lisina que não apresentavam nenhum indicativo de atuar como âncora para sítios conhecidos, considerando os nossos critérios de busca espacial. A comparação da conservação encontrada para esses resíduos em proteínas homólogas está ilustrada na forma de weblogos percentuais na figura 16. Uma mera inspeção visual já revela uma maior conservação dos resíduos pertencentes aos sítios conformacionais estudados do que o grupo de controle. Numericamente, a taxa de conservação média dos resíduos dos sítios conformacionais ficou em 80,6%, com desvio padrão de 18,0 Resíduos do grupo controle apresentaram uma média de 39,1%, com desvio padrão de 20,9.

Resíduos com alta taxa de conservação presentes no grupo de controle podem apresentar variadas explicações. Por exemplo, podem apresentar uma função crucial ainda não descrita, podem apresentar funções não relacionadas a fosforilação pelas PKs consideradas em nossos estudos, ou pode ser simplesmente parte da sequência que não obteve cobertura para ser modelada, e portanto não estudada no contexto de sítios

conformacionais. Por outro lado, os resíduos com as menores taxas de conservação no conjunto de resíduos pertencentes ao sítios conformacionais estudados, podem significar que a predição de uma determinada âncora não está correta, por exemplo por estar em uma disposição menos favorável que as demais âncoras possíveis, ou, quando apresenta um padrão de convergência tendendo a favorecer a fosforilação predita, ser resultado de uma divergência evolutiva recente.

Protein phosphatase 1 regulatory subunit 1A

A PRESERVA Resource R

Figura 16 – Padrão de conservação dos resíduos pertencentes aos sítios conformacionais com o maior número de configurações possíveis para cada um dos substratos considerados (em A, C e E) e o padrão de conservação dos resíduos do grupo controle desses substratos (B, D e F).

5.6 Os sítios conformacionais em ação: Uma visão in vivo da fosforilação de alpha-tubulina

A avaliação dos sítios da família PKC foi restringida a substratos da PKC β . A princípio, como não existem consensos de fosforilação diferenciados para cada membro da família das PKCs, seria possível prospectar todos os substratos fosforilados por qualquer PKC de forma indistinta. Porém essa restrição foi feita pois havia um interesse especial nessa PKC, objeto de estudo do grupo de pesquisa da Dra. Deborah Schechtman do Instituto de Química da USP. Substratos de PKC β apresentam uma taxa maior de sítios que não apresentam o consenso de fosforilação comparada à PKA. Para PKC β foram encontrados 105 sítios dos quais 46 não apresentam enquanto 59 apresentam o consenso

linear (53 com o consenso simples e apenas 6 o consenso completo). Assim como para PKA, a busca estrutural nesses sítios revelou uma elevada taxa de possíveis sítios estruturais.

Os sítios de PKC β estão distribuídos em 58 substratos. Desses 51 foram modelados sendo que foi possível modelar a região de interação com a quinase para apenas 40 sítios de fosforilação. Dos 40 sítios modelados, 19 não continham o consenso estrutural, sendo que a nossa busca revelou que 18 deles apresentavam alguma possibilidade de consenso estrutural. Para os 40 sítios modelados, a busca estrutural encontrou um consenso de fosforilação para 39 deles, mostrando que a mesma está adequada também para encontrar os motivos lineares.

Essa elevada taxa de sítios sem consenso linear e consequentemente um possível número elevado de sítios com consenso estrutural, pode ser explicado pela própria estrutura da PKC β . Quando comparamos a estrutura da fenda catalítica dessa quinase com o nosso modelo inicial de PKA, observamos que a fenda catalítica da PKC β é mais larga (figura 17). Na nossa análise volumétrica, utilizando o software KVFinder (OLIVEIRA et al., 2014), encontramos que a fenda de PKA apresenta um volume de 537 Å³, o que é consideravelmente inferior à da PKC β com 803 Å³. Com isso é possível conjecturar que para a PKC β , os sítios possuem um espaço maior para se ajustar e portanto mais possibilidade de se conformar estruturalmente.

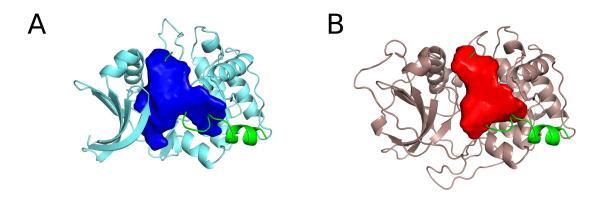


Figura 17 – A - A estrutura da PKC β com o sítio catalítico definido. B - A estrutura da PKA com o sítio catalítico definido. Em ambas as estruturas temos o PKI em verde, como referência para o posicionamento da fenda.

A despeito de nosso trabalho de prospecção computacional de sítios conformacionais apresentar resultados importantes, não existiam até esta data uma comprovação de que estes sítios realmente funcionassem em sistemas biológicos reais. De fato poucas estruturas de complexos de quinases com substratos estão disponíveis e de nosso conhecimento todos os complexos são formados com peptídeos apresentando consenso linear. Diante deste cenário, foi realizada uma colaboração com a Dra. Schechtman do laboratório de sinalização celular do IQ-USP, que tem interesse no entendimento de sinalização mediada pela família PKC.

A validação experimental da existência dos consensos estruturais foi realizado com o substrato α -Tubulina (PDB: 1JFF). Essa é uma proteína globular muito estudada, por ser uma das componentes dos microtúbulos e portanto desempenhar papel fundamental numa série de eventos celulares. Essa proteína apresenta uma fosforilação no resíduo T253, realizada pela PKC β (Figura 19A). Estudos comprovaram que essa fosforilação é realizada pela PKC β durante a mitose (DUARTE et al., 2014).

Esse sítio não apresenta em sua estrutura primária o consenso de fosforilação comum à família das PKC. Na verdade, esse sítio apresenta um ácido aspártico na posição P-2, gerando portanto uma configuração desfavorável para interagir com o SBS da PKC. Porém nossa prospecção estrutural em torno desse sítio, revelou a presença de resíduos plausíveis de serem utilizados como âncoras no processo de atracamento. Das cinco configurações encontradas, três eram configurações com apenas uma âncora, com as outras duas sendo permutações de um par de resíduos capazes de formar um consenso estrutural com duas âncoras. Dados da análise de RMSD dos complexos gerados podem ser vistos na tabela 5. Partindo da premissa que duas âncoras apresentam um maior potencial de estabilização do complexo e avaliando que a configuração com duas âncoras apresentou a melhor conformação quando comparada com o PKI, chegou-se à conclusão que uma configuração utilizando as âncoras K163 e K164 é a mais favorável, sendo esses resíduos essenciais para a interação. O modelo do complexo gerado utilizando essa configuração encontrada pode ser observado na figura 18.

Tabela 5 – Avaliação dos complexos gerados para o substrato α -Tubulina, indicando os resíduos considerados âncora e o RMSD, medido em Å³, para a conformação final da PK, do substrato todo e das regiões de interesse para a fosforilação.

P-2	P-3	PK	substrato	sítio	âncoras	PKI
R2	-	1,41	1,14	1,95	2,33	16,74
K163	K164	1,39	1,26	2,84	1,31	$5,\!33$
K164	K163	1,50	$1,\!17$	2,87	1,92	4,02
K166	-	1,43	$1,\!47$	3,06	$2,\!52$	6,69
K352	-	1,35	1,16	2,30	1,20	8,94

Para confirmar a importância das âncoras no processo de formação do complexo com a PKC β foram realizadas mutações individuais nos resíduos K163 e K164, que acarretaram numa significante queda na taxa de fosforilação da α -Tubulina, quando comparada com o tipo selvagem (DUARTE et al., 2014). As mutações independentes realizada foram T253A, K163A e K164A. A α -Tubulina com essas mutações foram capazes de manter a sua funcionalidade, sendo que as proteínas mutadas foram incorporadas nos microtúbulos (figura 19D), sendo que todas as mutações afetaram de forma contundente a taxa de fosforilação da α -Tubulina (figura 19E). O artigo completo desse trabalho está disponível como o anexo C.

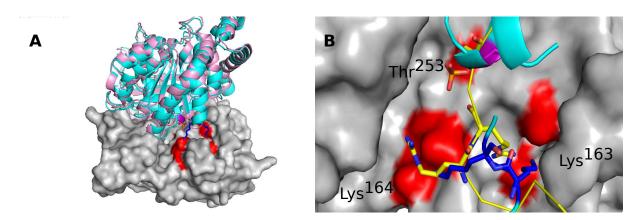


Figura 18 – A - Uma visão geral do complexo formado. Em representação de superfície, a PKC, com os resíduos ácidos que formam as regiões de interação S0, S-2 e S-3 em vermelho. Em cartoon, em azul a estrutura ajustada no processo de minimização de energia, com os resíduos chaves na interação destacados com as cadeias laterais explicitadas. Em rosa, a estrutura original do substrato após o atracamento rígido, para comparação da deformação. B - Visão em destaque da região de interação no sítio catalítico. Em amarelo, o PKI sobreposto, para comparação da conformação do sítio estrutural com esse modelo linear. Identificação dos resíduos do substrato para auxílio na comparação.

5.7 Considerações sobre o método de docking PK-substrato desenvolvido neste trabalho

Enquanto que o levantamento e triagem de dados sobre fosforilação, apesar de essenciais para esse estudo, representam uma etapa tecnicamente mais simples, a análise da estrutura dos sítios selecionados representa um real desafio. Gerar as estruturas tridimensionais dos sítios e prospectá-los já é uma tarefa de alta demanda computacional, considerando a quantidade de informações sobre fosforilação disponível na literatura e a recorrência encontrada de sítios sem a conformação ideal em sua estrutura primária.

Mas o maior desafio técnico encontrado foi estabelecer um método para gerar as conformações dos complexos PK-substrato. O principal objetivo dessa etapa é avaliar se o substrato é capaz de acomodar o sítio de fosforilação na fenda catalítica da PK em uma conformação análoga ao do PKI enquanto ainda mantém uma coerência estrutural. As dificuldades dessa parte consiste no fato de que esse é um protocolo voltado para aplicação em larga escala, para prospecção estrutural de bancos de fosforilação. Logo, a eficiência computacional é um elemento chave a ser considerado. Por outro lado, objetiva-se estudar a viabilidade da conformação de um complexo de forma bastante detalhada, buscando ajustar a interação de resíduos específicos de acordo com um modelo experimental selecionado (PKI).

A abordagem natural para um problema desse tipo seria utilizar algorítimos já publicados de *docking* proteína-proteína. Um primeiro problema dessa abordagem é que esses

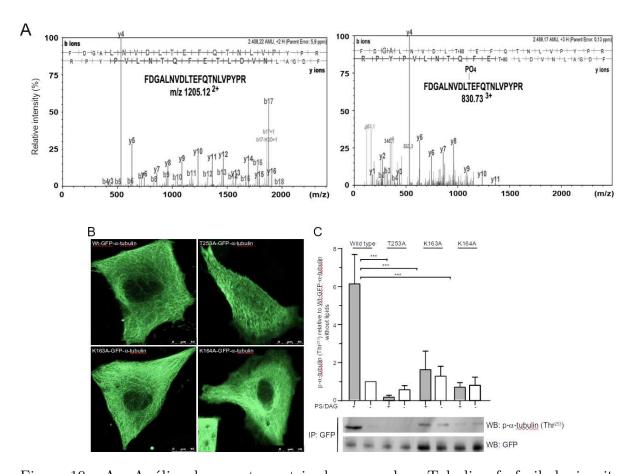


Figura 19 – A - Análise de espectrometria de massa da α -Tubulina fosforilada $in\ vitro$ pela PKC β . A esquerda o peptídeo FDGALNVDL(T253)EQTNLVPYPR sem fosforilação (massa monoisotópica de 2408,22 Da) e a direita o peptídeo fosforilado no resíduo T253 (massa de 2489,17 Da). B - Detecção por microscopia confocal de fluorecência de células HeLa transfectadas com mutantes da α -Tubulina associadas ao GFP. C - Moléculas de GFP- α -Tubulina imunoprecipitadas de células HEK 293T foram fosforiladas por PKC β $in\ vitro$. A análise de $Western\ blot$ mostra a média e o desvio padrão da quantidade das variantes fosforiladas e não fosforilada da α -Tubulina que foram encontradas, detectadas pelo anticorpo que reconhece a fosforilação da T253. Adaptado de (DUARTE et al., 2014).

métodos costumam ter um elevado custo computacional, uma vez que são desenvolvidos para resolver problemas mais genéricos e portanto exploram mais conformações do que era interessante para esse trabalho, focado em avaliar uma condição bem específica. Para diminuir a complexidade dos algorítimos de docking tradicionais, seria possível utilizar as restrições de distância como as utilizadas nesse trabalho, baseadas no PKI. Mas como foi explicitado na sessão 4, os substratos quando conformados no SBS apresentam quase que na totalidade algum tipo de impedimento estérico. Esse tipo de flexibilização na configuração não é permitida pelos programas tradicionais de docking. Para esse passo, foram considerados insuficientes os algorítimos PatchDock (SCHNEIDMAN-DUHOVNY et al., 2005), HEX (MACINDOE et al., 2010) e ZDock (PIERCE; HOURAI; WENG,

2011), todos programas bem estabelecidos na literatura, mas limitados a manipulação das estruturas das proteínas rígidas.

A solução ideal desse problema seria realizar um docking flexível utilizando restrições. Apesar de vários trabalhos na literatura abordarem o problema de interação proteína-proteína com flexibilidade (BONVIN, 2006; WANG; BRADLEY; BAKER, 2007; TORCHALA et al., 2013), esses foram considerados inadequados. Vários motivos justificam essa inadequabilidade, desde indisponibilidade de programa descrito na literatura, ineficiência nos resultados apresentados até a dificuldade de operação em larga escala por só estar disponível na forma de um serviço web. Porém o principal fator limitante dessas abordagens é o custo computacional proibitivo para a solução de um problema de larga escala. A abordagem desse trabalho para a introdução de flexibilidade na geração do complexo do substrato com a PK, utilizando minimização de energia guiada, uma abordagem computacionalmente minimalista que explora uma restrição biológica conhecida, não é uma abordagem comum. Porém a estratégia utilizada se mostrou computacionalmente eficiente e, o mais importante, capaz de atingir um resultado satisfatório, uma vez que a avaliação dos modelos gerados forneceu informações suficientes para a completude dos objetivos científicos propostos.

6 Conclusão

Esse estudo visou dar mais um passo no entendimento das condições que coordenam os eventos de fosforilação em proteínas. Esse projeto trouxe uma ideia nova que pode ser muito relevante ao tema, já que como demonstramos pela avaliação do banco de dados de fosforilação, a ocorrência de sítios que são fosforilados por uma quinase sem conter o consenso de fosforilação, é bastante comum. Por meio de estudos computacionais, tanto para os substratos de PKA, quanto de PKC β , constatamos que em ambos os casos, mais de 90% dos sítios sem consenso de fosforilação na estrutura primária apresentavam em sua estrutura terciária uma disposição de resíduos factível de se ajustar espacialmente de forma semelhante à de uma estrutura proteica com o consenso de fosforilação tradicional. Sítios com essa características foram chamados de sítios conformacionais, por apresentarem o padrão sequencial reconhecido pelas PKs não em sua sequência primária, mas sim na conformação estrutural da proteína dobrada. Nossos estudos computacionais para os substratos de PKA revelam que grande parte desses sítios conformacionais poderiam se ajustar na fenda catalítica e assim auxiliar a formação do complexo com a quinase, sendo portanto possivelmente um elemento chave no processo de reconhecimento desses substratos. Com a validação experimental obtida para o substrato α -Tubulina, confirmamos a existência de sítios conformacionais e a coerência dos modelos computacionais propostos. Com isso, passa a ser necessário uma revisão de todo o conhecimento sobre o tema sob essa nova ótica. O entendimento do processo do reconhecimento de substratos pelas PKs, analisado por tanto tempo de forma unidimensional, passa a requerer uma análise tridimensional. Essa abordagem pode, não só ajudar a esclarecer o processo de fosforilação, agregando valor em áreas de ciências básicas, mas também pode ser usado de maneira aplicada no desenvolvimento de novas tecnologias.

- ALTO, N. M. et al. Bioinformatic design of a-kinase anchoring protein-in silico: A potent and selective peptide antagonist of type II protein kinase a anchoring. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 100, n. 8, p. 4445–4450, apr 2003. Disponível em: http://dx.doi.org/10.1073/pnas.0330734100. Citado na página 20.
- ALTSCHUL, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, Oxford University Press (OUP), v. 25, n. 17, p. 3389–3402, sep 1997. Disponível em: http://dx.doi.org/10.1093/nar/25.17.3389. Citado na página 29.
- APWEILER, R. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, Oxford University Press (OUP), v. 32, n. 90001, p. 115D–119, jan 2004. Disponível em: http://dx.doi.org/10.1093/nar/gkh131. Citado na página 27.
- BLOM, N. et al. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *PROTEOMICS*, Wiley-Blackwell, v. 4, n. 6, p. 1633–1649, jun 2004. Disponível em: http://dx.doi.org/10.1002/pmic.200300771. Citado na página 20.
- BONVIN, A. M. Flexible protein–protein docking. *Current Opinion in Structural Biology*, Elsevier BV, v. 16, n. 2, p. 194–200, apr 2006. Disponível em: http://dx.doi.org/10.1016/j.sbi.2006.02.002. Citado na página 59.
- BRINKWORTH, R. I.; BREINL, R. A.; KOBE, B. Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proceedings of the National Academy of Sciences*, v. 100, n. 1, p. 74–79, 2003. Disponível em: http://www.pnas.org/content/100/1/74.abstract. Citado 2 vezes nas páginas 15 e 19.
- BYLUND, D. B.; KREBS, E. G. Effect of denaturation on the susceptibility of proteins to enzymic phosphorylation. *Journal of Biological Chemistry*, v. 250, n. 16, p. 6355–6361, 1975. Disponível em: http://www.jbc.org/content/250/16/6355.abstract. Citado na página 18.
- CAMACHO, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics*, Springer Science + Business Media, v. 10, n. 1, p. 421, 2009. Disponível em: http://dx.doi.org/10.1186/1471-2105-10-421. Citado na página 38.
- CAPRA, J. A.; SINGH, M. Predicting functionally important residues from sequence conservation. *Bioinformatics*, Oxford University Press (OUP), v. 23, n. 15, p. 1875–1882, may 2007. Disponível em: http://dx.doi.org/10.1093/bioinformatics/btm270. Citado na página 38.
- COHEN, P.; RYLATT, D. B.; NIMMO, G. A. The hormonal control of glycogen metabolism: The amino acid sequence at the phosphorylation site of protein phosphatase inhibitor-1. *FEBS Letters*, Wiley-Blackwell, v. 76, n. 2, p. 182–186, apr 1977. Disponível em: http://dx.doi.org/10.1016/0014-5793(77)80147-6. Citado na página 18.

COHEN, P.; WATSON, D. C.; DIXON, G. H. The hormonal control of activity of skeletal muscle phosphorylase kinase. amino-acid sequences at the two sites of action of adenosine-3: 5-monophosphate-dependent protein kinase. *Eur J Biochem*, Wiley-Blackwell, v. 51, n. 1, p. 79–92, feb 1975. Disponível em: http://dx.doi.org/10.1111/j.1432-1033.1975.tb03909.x. Citado na página 16.

- CROOKS, G. E. WebLogo: A sequence logo generator. *Genome Research*, Cold Spring Harbor Laboratory Press, v. 14, n. 6, p. 1188–1190, may 2004. Disponível em: http://dx.doi.org/10.1101/gr.849004. Citado na página 38.
- DAILE, P.; CARNEGIE, P. R. Peptides from myelin basic protein as substrates for adenosine 31, 51-cyclic monophosphate-dependent protein kinases. *Biochemical and Biophysical Research Communications*, Elsevier BV, v. 61, n. 3, p. 852–858, dec 1974. Disponível em: http://dx.doi.org/10.1016/0006-291X(74)90234-4. Citado na página 17.
- DAILE, P.; CARNEGIE, P. R.; YOUNG, J. D. Synthetic substrate for cyclic AMP-dependent protein kinase. *Nature*, Nature Publishing Group, v. 257, n. 5525, p. 416–418, oct 1975. Disponível em: <http://dx.doi.org/10.1038/257416a0>. Citado 2 vezes nas páginas 18 e 19.
- DEBYE, P. Näherungsformeln für die zylinderfunktionen für große werte des arguments und unbeschränkt veränderliche werte des index. *Math. Ann.*, Springer Science + Business Media, v. 67, n. 4, p. 535–558, dec 1909. Disponível em: http://dx.doi.org/10.1007/BF01450097. Citado na página 36.
- DUARTE, M. L. et al. Protein folding creates structure-based, noncontiguous consensus phosphorylation motifs recognized by kinases. *Science Signaling*, American Association for the Advancement of Science (AAAS), v. 7, n. 350, p. ra105—ra105, nov 2014. Disponível em: http://dx.doi.org/10.1126/scisignal.2005412. Citado 2 vezes nas páginas 56 e 58.
- DUREK, P. et al. Detection and characterization of 3d-signature phosphorylation site motifs and their contribution towards improved phosphorylation site prediction in proteins. *BMC Bioinformatics*, Springer Science + Business Media, v. 10, n. 1, p. 117, 2009. Disponível em: http://dx.doi.org/10.1186/1471-2105-10-117. Citado na página 20.
- ECHOLS, N. Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. *Nucleic Acids Research*, Oxford University Press (OUP), v. 30, n. 11, p. 2515–2523, jun 2002. Disponível em: http://dx.doi.org/10.1093/nar/30.11.2515. Citado na página 15.
- ENNOR, A. H.; ROSENBERG, H.; ARMSTRONG, M. D. Specificity of creatine phosphokinase. *Nature*, Nature Publishing Group, v. 175, n. 4446, p. 120–120, jan 1955. Disponível em: http://dx.doi.org/10.1038/175120a0. Citado na página 16.
- HANKS, S. K.; HUNTER, T. Protein kinases 6. the eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *The FASEB Journal*, v. 9, n. 8, p. 576–96, 1995. Disponível em: http://www.fasebj.org/content/9/8/576.abstract. Citado na página 20.
- HAYASHIDA, K. et al. Syndecan-1 expression in epithelial cells is induced by transforming growth factor beta through a PKA-dependent pathway. *Journal of Biological Chemistry*,

American Society for Biochemistry & Molecular Biology (ASBMB), v. 281, n. 34, p. 24365–24374, jun 2006. Disponível em: http://dx.doi.org/10.1074/jbc.M509320200. Citado na página 53.

- HJELMQUIST, G. et al. Amino acid sequence of a (32p)phosphopeptide from pig liver pyruvate kinase phosphorylated by cyclic 31, 51-AMP-stimulated protein kinase and (32p)ATP. Biochemical and Biophysical Research Communications, Elsevier, v. 61, n. 2, p. 559–563, nov 1974. Disponível em: http://dx.doi.org/10.1016/0006-291X(74)90993-0. Citado na página 18.
- HJERRILD, M. et al. Identification of phosphorylation sites in protein kinase a substrates using artificial neural networks and mass spectrometry. *J. Proteome Res.*, American Chemical Society (ACS), v. 3, n. 3, p. 426–433, jun 2004. Disponível em: http://dx.doi.org/10.1021/pr0341033. Citado na página 20.
- HOOFT, R. W. W. et al. Errors in protein structures. *Nature*, Nature Publishing Group, v. 381, n. 6580, p. 272–272, may 1996. Disponível em: http://dx.doi.org/10.1038/381272a0. Citado na página 29.
- HORNBECK, P. V. et al. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *PROTEOMICS*, Wiley-Blackwell, v. 4, n. 6, p. 1551–1561, jun 2004. Disponível em: http://dx.doi.org/10.1002/pmic.200300772. Citado na página 22.
- HORNBECK, P. V. et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Research*, Oxford University Press (OUP), v. 40, n. D1, p. D261–D270, dec 2011. Disponível em: http://dx.doi.org/10.1093/nar/gkr1122. Citado 2 vezes nas páginas 22 e 27.
- HOUSE, C.; KEMP, B. Protein kinase c contains a pseudosubstrate prototope in its regulatory domain. *Science*, American Association for the Advancement of Science (AAAS), v. 238, n. 4834, p. 1726–1728, dec 1987. Disponível em: http://dx.doi.org/10.1126/science.3686012. Citado na página 21.
- HUANG, K. x.; PAUDEL, H. K. Ser67-phosphorylated inhibitor 1 is a potent protein phosphatase 1 inhibitor. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 97, n. 11, p. 5824–5829, may 2000. Disponível em: http://dx.doi.org/10.1073/pnas.100460897>. Citado na página 53.
- HUBBARD, S. R. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and atp analog. *The EMBO Journal*, EMBO Press, v. 16, n. 18, p. 5572–5581, 1997. ISSN 0261-4189. Citado na página 19.
- HUMBLE, E. et al. Non-dependence on native structure of pig liver pyruvate kinase when used as a substrate for cyclic $3\prime$, $5\prime$ -AMP-stimulated protein kinase. Biochemical and Biophysical Research Communications, Elsevier BV, v. 66, n. 2, p. 614–621, sep 1975. Disponível em: http://dx.doi.org/10.1016/0006-291X(75)90554-9. Citado na página 18.
- IAKOUCHEVA, L. M. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, Oxford University Press (OUP), v. 32, n. 3, p. 1037–1049, feb 2004. Disponível em: http://dx.doi.org/10.1093/nar/gkh253. Citado na página 20.

JIANG, H. et al. Protein kinase c inhibits autophagy and phosphorylates LC3. *Biochemical and Biophysical Research Communications*, Elsevier BV, v. 395, n. 4, p. 471–476, may 2010. Disponível em: http://dx.doi.org/10.1016/j.bbrc.2010.04.030. Citado na página 53.

- JOHNSON, L. N. et al. The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nature Cell Biology*, Nature Publishing Group, v. 1, n. 7, p. 438–443, oct 1999. Disponível em: http://dx.doi.org/10.1038/15674>. Citado na página 19.
- JR, J. E. F.; UBERSAX, J. A. Mechanisms of specificity in protein phosphorylation. *Nature Reviews Molecular Cell Biology*, v. 8, p. 530–541, 2007. Citado na página 15.
- JUNG, J.; LEE, B. Protein structure alignment using environmental profiles. *Protein Engineering Design and Selection*, Oxford University Press (OUP), v. 13, n. 8, p. 535–543, aug 2000. Disponível em: http://dx.doi.org/10.1093/protein/13.8.535. Citado na página 29.
- KEMP, B. E.; BENJAMINI, E.; KREBS, E. G. Synthetic hexapeptide substrates and inhibitors of 3':5'-cyclic amp-dependent protein kinase. *Proceedings of the National Academy of Sciences*, v. 73, n. 4, p. 1038–1042, 1976. Disponível em: http://www.pnas.org/content/73/4/1038.abstract. Citado na página 18.
- KEMP, B. E. et al. Substrate specificity of the cyclic amp-dependent protein kinase. *Proceedings of the National Academy of Sciences*, v. 72, n. 9, p. 3448–3452, 1975. Disponível em: http://www.pnas.org/content/72/9/3448.abstract. Citado 2 vezes nas páginas 18 e 19.
- KEMP, B. E. et al. Role of multiple basic residues in determining the substrate specificity of cyclic amp-dependent protein kinase. *Journal of Biological Chemistry*, v. 252, n. 14, p. 4888–94, 1977. Disponível em: http://www.jbc.org/content/252/14/4888.short. Citado na página 18.
- KING, R. D.; STERNBERG, M. J. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Science*, Wiley-Blackwell, v. 5, n. 11, p. 2298–2310, nov 1996. Disponível em: http://dx.doi.org/10.1002/pro.5560051116. Citado na página 29.
- KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. *Science*, American Association for the Advancement of Science (AAAS), v. 220, n. 4598, p. 671–680, may 1983. Disponível em: http://dx.doi.org/10.1126/science.220.4598.671. Citado na página 36.
- KNIGHTON, D. et al. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, v. 253, n. 5018, p. 407–414, 1991. Disponível em: http://www.sciencemag.org/content/253/5018/407.abstract. Citado 2 vezes nas páginas 21 e 27.
- KNIGHTON, D. et al. Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, American Association for the Advancement of Science (AAAS), v. 253, n. 5018, p. 414–420, jul 1991. Disponível em: http://dx.doi.org/10.1126/science.1862343. Citado na página 21.

KOBE, B. et al. Substrate specificity of protein kinases and computational prediction of substrates. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, Elsevier BV, v. 1754, n. 1-2, p. 200–209, dec 2005. Disponível em: http://dx.doi.org/10.1016/j.bbapap.2005.07.036. Citado 2 vezes nas páginas 27 e 41.

- KONAGURTHU, A. S. et al. MUSTANG: A multiple structural alignment algorithm. *Proteins*, Wiley-Blackwell, v. 64, n. 3, p. 559–574, may 2006. Disponível em: http://dx.doi.org/10.1002/prot.20921. Citado na página 37.
- KREBS, E. G.; BEAVO, J. A. Phosphorylation-dephosphorylation of enzymes. *Annu. Rev. Biochem.*, Annual Reviews, v. 48, n. 1, p. 923–959, jun 1979. Disponível em: <http://dx.doi.org/10.1146/annurev.bi.48.070179.004423>. Citado 2 vezes nas páginas 16 e 18.
- KRIEGER, E. et al. Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins*, Wiley-Blackwell, v. 57, n. 4, p. 678–683, dec 2004. Disponível em: http://dx.doi.org/10.1002/prot.20251. Citado na página 36.
- KRIEGER, E. et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins*, Wiley-Blackwell, v. 77, n. S9, p. 114–122, 2009. Disponível em: http://dx.doi.org/10.1002/prot.22570. Citado 2 vezes nas páginas 29 e 30.
- KUROCHKIN, S. et al. Study of the specificity of histone kinase using synthetic substrates. FEBS Letters, Wiley-Blackwell, v. 88, n. 1, p. 59–61, apr 1978. Disponível em: <http://dx.doi.org/10.1016/0014-5793(78)80606-1>. Citado na página 18.
- LANGAN, T. A. Cyclic amp and histone phosphorylation. *Annals of the New York Academy of Sciences*, Wiley-Blackwell, v. 185, n. 1, p. 166–180, dec 1971. Disponível em: http://dx.doi.org/10.1111/j.1749-6632.1971.tb45246.x. Citado na página 17.
- LANGAN, T. A. Protein kinases and protein kinase substrates. *Adv Cyclic Nucleotide Res*, v. 3, p. 99–153, 1973. Citado na página 17.
- MACINDOE, G. et al. HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Research*, Oxford University Press (OUP), v. 38, n. Web Server, p. W445–W449, may 2010. Disponível em: http://dx.doi.org/10.1093/nar/gkq311. Citado na página 58.
- MANNING, G. The protein kinase complement of the human genome. *Science*, American Association for the Advancement of Science (AAAS), v. 298, n. 5600, p. 1912–1934, dec 2002. Disponível em: http://dx.doi.org/10.1126/science.1075762. Citado na página 20.
- MCGUFFIN, L. J.; BRYSON, K.; JONES, D. T. The PSIPRED protein structure prediction server. *Bioinformatics*, Oxford University Press (OUP), v. 16, n. 4, p. 404–405, apr 2000. Disponível em: <http://dx.doi.org/10.1093/bioinformatics/16.4.404>. Citado na página 29.
- MüCKSTEIN, U.; HOFACKER, I.; STADLER, P. Stochastic pairwise alignments. *Bioinformatics*, v. 18, n. suppl 2, p. S153–S160, 2002. Disponível em: http://bioinformatics.oxfordjournals.org/content/18/suppl_2/S153.abstract. Citado na página 30.

MONTENEGRO, M. et al. Comparative study of the prereactive protein kinase a michaelis complex with kemptide substrate. *Journal of Computer-Aided Molecular Design*, Springer Netherlands, v. 21, n. 10-11, p. 603–615, 2007. ISSN 0920-654X. Disponível em: http://dx.doi.org/10.1007/s10822-007-9143-x. Citado na página 15.

- OLIVEIRA, S. H. et al. KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinformatics*, Springer Science + Business Media, v. 15, n. 1, p. 197, 2014. Disponível em: http://dx.doi.org/10.1186/1471-2105-15-197>. Citado na página 55.
- PAULUCCI-HOLTHAUZEN, A. A.; O'CONNOR, K. L. Use of pseudosubstrate affinity to measure active protein kinase a. *Analytical Biochemistry*, Elsevier BV, v. 355, n. 2, p. 175–182, aug 2006. Disponível em: http://dx.doi.org/10.1016/j.ab.2006.06.002. Citado na página 42.
- PIERCE, B. G.; HOURAI, Y.; WENG, Z. Accelerating protein docking in ZDOCK using an advanced 3d convolution library. *PLoS ONE*, Public Library of Science (PLoS), v. 6, n. 9, p. e24657, sep 2011. Disponível em: http://dx.doi.org/10.1371/journal.pone.0024657. Citado na página 59.
- POMERANTZ, A. H. et al. Studies on the mechanism of phosphorylation of synthetic polypeptides by a calf thymus cyclic amp-dependent protein kinase. *Proceedings of the National Academy of Sciences*, v. 74, n. 10, p. 4261–4265, 1977. Disponível em: http://www.pnas.org/content/74/10/4261.abstract. Citado na página 18.
- PRUITT, K. D. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, Oxford University Press (OUP), v. 33, n. Database issue, p. D501–D504, dec 2004. Disponível em: http://dx.doi.org/10.1093/nar/gki025. Citado na página 38.
- QIU, J.; ELBER, R. SSALN: An alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs. *Proteins*, Wiley-Blackwell, v. 62, n. 4, p. 881–891, dec 2005. Disponível em: http://dx.doi.org/10.1002/prot.20854. Citado na página 30.
- RABINOWITZ, M.; LIPMANN, F. Reversible phosphate transfer between yolk phosphoprotein and adenosine triphosphate. *Journal of Biological Chemistry*, v. 235, n. 4, p. 1043–1050, 1960. Disponível em: http://www.jbc.org/content/235/4/1043.short. Citado na página 16.
- ROBERTS, R. J. PubMed central: The GenBank of the published literature. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 98, n. 2, p. 381–382, jan 2001. Disponível em: http://dx.doi.org/10.1073/pnas.98.2.381. Citado na página 27.
- SCHNEIDMAN-DUHOVNY, D. et al. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research*, Oxford University Press (OUP), v. 33, n. Web Server, p. W363–W367, jul 2005. Disponível em: http://dx.doi.org/10.1093/nar/gki481. Citado na página 58.
- SHABB, J. B. Physiological substrates of cAMP-dependent protein kinase. *Chemical Reviews*, American Chemical Society (ACS), v. 101, n. 8, p. 2381–2412, aug 2001. Disponível em: http://dx.doi.org/10.1021/cr000236l. Citado na página 18.

SHINDYALOV, I. N.; BOURNE, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering Design and Selection*, Oxford University Press (OUP), v. 11, n. 9, p. 739–747, sep 1998. Disponível em: http://dx.doi.org/10.1093/protein/11.9.739. Citado na página 29.

- SHLYAPNIKOV, S. et al. Investigation of the sites phosphorylated in lysine-rich histones by protein kinase from pig brain. *FEBS Letters*, Wiley-Blackwell, v. 53, n. 3, p. 316–319, may 1975. Disponível em: http://dx.doi.org/10.1016/0014-5793(75)80045-7. Citado na página 18.
- SIMON, S. M.; PESKIN, C. S.; OSTER, G. F. What drives the translocation of proteins? *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 89, n. 9, p. 3770–3774, may 1992. Disponível em: http://dx.doi.org/10.1073/pnas.89.9.3770. Citado na página 36.
- SODERLING, T. R. et al. Inactivation of glycogen synthetase and activation of phosphorylase kinase by muscle adenosine 3',5'-monophosphate-dependent protein kinases. *Journal of Biological Chemistry*, v. 245, n. 23, p. 6317–6328, 1970. Disponível em: http://www.jbc.org/content/245/23/6317.abstract. Citado na página 17.
- SOLS, A.; CRANE, R. K. Substrate specificity of brain hexokinase. *Journal of Biological Chemistry*, v. 210, n. 2, p. 581–595, 1954. Disponível em: http://www.jbc.org/content/210/2/581.short. Citado na página 16.
- TORCHALA, M. et al. SwarmDock: a server for flexible protein-protein docking. *Bioinformatics*, Oxford University Press (OUP), v. 29, n. 6, p. 807–809, jan 2013. Disponível em: http://dx.doi.org/10.1093/bioinformatics/btt038>. Citado na página 59.
- WANG, C.; BRADLEY, P.; BAKER, D. Protein-protein docking with backbone flexibility. *Journal of Molecular Biology*, Elsevier BV, v. 373, n. 2, p. 503–519, oct 2007. Disponível em: http://dx.doi.org/10.1016/j.jmb.2007.07.050. Citado na página 59.
- WOODGETT, J. R.; GOULD, K. L.; HUNTER, T. Substrate specificity of protein kinase c. use of synthetic peptides corresponding to physiological sites as probes for substrate recognition requirements. Eur J Biochem, Wiley-Blackwell, v. 161, n. 1, p. 177–184, nov 1986. Disponível em: $\frac{\text{http:}}{\text{dx.doi.org}}$ Citado na página 21.
- YAFFE, M. B. et al. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol.*, Nature Publishing Group, v. 19, n. 4, p. 348–353, apr 2001. Disponível em: http://dx.doi.org/10.1038/86737. Citado na página 20.
- YE, Y.; GODZIK, A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, Oxford University Press (OUP), v. 19, n. Suppl 2, p. ii246–ii255, sep 2003. Disponível em: http://dx.doi.org/10.1093/bioinformatics/btg1086. Citado na página 29.
- YEAMAN, S. J. et al. The substrate specificity of adenosine 3/:5/-cyclic monophosphate-dependent protein kinase of rabbit skeletal muscle. *Biochem. J.*, Portland Press Ltd., v. 162, n. 2, p. 411–421, feb 1977. Disponível em: http://dx.doi.org/10.1042/bj1620411. Citado na página 18.

ZANZONI, A. et al. Phospho3d 2.0: an enhanced database of three-dimensional structures of phosphorylation sites. *Nucleic Acids Research*, Oxford University Press (OUP), v. 39, n. Database, p. D268–D271, oct 2010. Disponível em: http://dx.doi.org/10.1093/nar/gkq936. Citado na página 20.

ZHANG, J.; YANG, P. L.; GRAY, N. S. Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer*, Nature Publishing Group, v. 9, n. 1, p. 28–39, jan 2009. Disponível em: http://dx.doi.org/10.1038/nrc2559. Citado na página 15.



ANEXO A – Tabela dos substratos de PKA gerados por homologia

Tabela 6 – Substratos de PKA obtidos no PhosphoSitePlus que não apresentavam o consenso de fosforilação em sua sequência e que foram última coluna, a porcentagem da sequência que foi modelada. pelo YASARA. Seguem listados os PDBs utilizados como moldes acompanhado da cobertura e a identidade desses moldes. Na modelados por meio de técnicas de modelagem comparativa. A qualidade dos modelos foi avaliada utilizando métricas definidas

P14598 P51398	F14598	111700	P02687	Q13522	Q6J4I0	P27732	P23441	P12839	P02522	O54724	Q9GZQ8	Q61411	P63259	P00517	O88767	O14921	Q66H98	Q64122	Q06210	P68136	P63269	P60711	P26955	P11473	P07335	B5DEN5	P18828	UNIPROT
	bom	ótimo	bom	$\acute{ m o}{ m timo}$	$\acute{ m otimo}$	satisfatório	$\acute{ m o}{ m timo}$	satisfatório	$\acute{ m o}{ m timo}$	satisfatório	$\acute{ m o}{ m timo}$	ótimo	$\acute{ m o}{ m timo}$	$\acute{ m o}{ m timo}$	$\acute{ m o}{ m timo}$	ótimo	ótimo	$\acute{ m o}{ m timo}$	satisfatório	[Diedros								
1								pobre																			ruim	3D
	satisfatório	satisfatório	satisfatório	pobre	pobre	pobre	pobre	pobre	pobre	pobre	$\acute{ m otimo}$	$\acute{ m o}{ m timo}$	$_{\rm bom}$	$_{ m bom}$	$_{ m bom}$	$_{ m bom}$	$_{ m bom}$	$_{ m bom}$	$_{ m bom}$	$_{ m bom}$	$_{ m bom}$	$_{ m bom}$	ruim	Geral				
14000	$3\mathrm{HU}3$	1NG2	3PL6	4Q2C-A	3IOX	2R9R	3RKQ	2FKJ	1YTQ	2I1J	2ZJD	4LDJ	$1 \mathrm{WUA}$	4022	20R3	2BV1	1C1G	1MXE	2ZJ3	2FXU	2FXU	2FXU	2 GYS	1IE8	3DRB	1IJF	3IOX	molde
20	57,8	45,1	33,7	70,8	41,0	14,5	15,6	27,9	88,3	78,8	96,0	88,9	95,2	96,3	98,4	79,2	65,5	77,2	52,2	95,8	96,0	95,7	44,1	57,4	98,7	39,6	56,3	
00 /	18,3	99,4	42,1	20,7	17,9	21,6	81,0	26,3	97,8	14,9	100,0	92,9	93,8	96,2	91,4	49,2	15,4	47,0	100,0	99,7	98,3	93,9	61,0	99,6	96,5	53,9	19,4	
MaCe	$3\mathrm{HU}1$	$1 \mathrm{UEC}$	306F		3IPK	3G43-E	1FTT	3SWK	2BB2	2I1K	,	40BE	,	4DG2	1P5F	2ODE	$2\mathrm{EFR}$	3WFN	2ZJ4	1QZ5	1QZ5	1QZ5	3CXE	1IE9	3B6R	1F60	3IPK	molde
π 0	58,8	45,4	62,7		41,0	03,0	15,1	10,0	88,3	79,3		88,9		99,7	98,4	82,4	36,5	77,2	52,2	95,2	95,5	95,5	45,0	58,5	98,7	39,6	49,8	
ο π	20.1	99,4	11,3		17,9	92,5	100,0	38,8	100,0	16,4		93,5		96,9	91,4	47,3	09,9	45,5	100,0	99,7	98,3	93,9	60,5	99,6	96,5	53,9	20,0	
OI 10	3D8B	$1 \mathrm{KQ} 6$	1FV1	,	,	4EKW	1MH3	3UF1	1BLB	2YPG	,	4LPK		4HPU	2RK3	2IHD	$2 \mathrm{EFS}$	4DJC	,	1WU.	1WU.	1WU.	1GH7	1DB1	3DRE	2B7C	1	molde
			10,7			,		12,4				87,8																Θ
n / n	14,5	99,3	100,0			25,8	19,2	41,9	100,0	19,4																		
	1ZM	10V:	3T0E	,				3TNU																			1	molde
			72,2					J 11,2				V = 84,1											7 22,8					е
27 0	16,9	97,0	22,1			21,9	19,6	43,2	55,2			90,6																
11DE	1R7F	1K4U	1YMI	ı				3TNU																				molde
			M = 08,3					J = 11,2				87,8											07,9		1			e
			100,0					43,2				100,0																
-	_	•			100,0			62,5																			100,0	Modelado

ANEXO B – Tabela com a avaliação por RMSD dos complexos gerados para os substratos de PKA

Tabela 7 – Avaliação dos complexos gerado para todos os sítios de PKA sem consenso sequencial e que possui possibilidade de consenso estrutural. Para cada substrato apresentamos os sítios fosforilados bem como as possíveis âncoras propostas, seguido pelo RMSD da estrutura completa da PK e do substrato. Por fim, o RMSD das regiões de interação do substrato, focada em janelas de 10 resíduos em torno do sítio de fosforilação e das âncoras e o RMSD do $C\alpha$ dos resíduos chaves do substrato comparado com o posicionamento do PKI em sua posição na fenda catalítica.

RMSD

Substrato	Sítio	P-2	P-3	PK	Substrato	Região Sítio	Região âncoras	Referência PKI
B5DEN5	S106	K66	-	1,32	1,92	$7,\!47$	$2,\!22$	$9,\!41$
B5DEN5	S106	K73	K74	1,21	1,65	6,98	5,81	9,01
B5DEN5	S106	K74	K73	1,39	1,62	8,75	4,93	10,34
B5DEN5	S106	K96	-	1,46	2,17	4,08	6,03	11,38
B5DEN5	S140	K139	-	1,15	1,74	5,33	5,54	$9,\!53$
B5DEN5	S140	K197	-	1,23	1,51	5,63	2,38	8,26
B5DEN5	S140	K224	-	1,19	1,85	4,34	$5,\!22$	11,83
O14921	T41	R13	-	1,28	1,40	2,74	$5,\!62$	7,09
O14921	T41	R18	K17	1,65	1,63	1,73	5,03	4,50
O14921	T41	K17	R18	1,71	1,82	$2,\!57$	5,20	9,10
O14921	T41	K42	-	1,35	1,16	1,24	1,15	$9,\!56$
O54724	S171	K167	-	1,45	1,55	$2,\!25$	2,08	6,67
O54724	S171	K172	-	1,14	1,42	1,84	1,98	7,06
O54724	S171	K175	-	1,39	1,63	2,59	3,56	8,80
O54724	S171	K319	-	1,22	1,60	2,10	2,43	8,85
O88767	S155	R145	-	1,63	1,48	3,75	4,94	$9,\!25$
O88767	S155	R156	-	1,36	1,42	1,93	1,83	8,09
O88767	S155	K122	-	1,98	1,54	1,43	3,88	8,15
O88767	S155	K130	-	1,48	1,16	1,74	3,09	8,05

O88767	S155	K148	-	1,68	1,26	1,18	$2,\!45$	7,88
O88767	S155	K377	-	1,62	1,46	1,30	2,83	9,14
O88767	T154	R145	-	1,45	1,37	2,46	3,71	8,12
O88767	T154	R156	-	1,45	1,31	2,67	2,51	9,83
O88767	T154	K122	-	2,01	1,91	1,98	4,69	8,80
O88767	T154	K130	-	1,44	1,15	1,61	2,08	9,37
O88767	T154	K148	-	1,70	1,31	1,68	2,81	9,30
P00517	S15	K17	-	1,19	0,89	4,31	4,42	8,41
P00517	S15	K22	-	1,01	0,60	1,39	1,09	8,44
P00517	S15	K9	-	1,02	0,80	2,40	2,86	6,03
P00517	S213	R195	-	1,24	0,85	1,05	1,18	4,32
P00517	S213	K214	-	1,19	0,80	1,30	1,30	8,50
P00517	S326	R135	-	1,22	0,91	3,60	0,78	5,82
P00517	S326	K318	-	1,14	0,82	3,88	3,75	4,62
P00517	S326	K320	-	1,27	0,98	$2,\!17$	2,03	3,81
P00517	S326	K48	-	1,12	0,94	3,57	1,73	$5,\!22$
P00517	S54	R57	-	2,82	2,00	$32,\!47$	$32,\!55$	9,08
P00517	S54	K73	-	1,59	1,41	2,88	3,46	11,33
P00517	S54	K77	-	1,29	1,04	1,66	$1,\!27$	9,30
P00517	S54	K79	K82	1,43	1,27	$1,\!17$	1,67	9,60
P00517	S54	K82	K79	1,49	1,48	1,95	1,38	11,11
P00517	S54	K84	-	1,52	1,38	2,31	1,73	8,61
P02522	S204	K11	-	$0,\!27$	2,29	36,73	43,86	10,28
P02687	S10	R23	-	1,15	5.99	16,03	12,28	9,88
P02687	S10	R5	K4	1,09	1,70	3,16	3,23	$4,\!65$
P02687	S10	R9	K11	1,04	1,27	9,83	9,58	9,18
P02687	S10	K11	R9	1,00	1,19	10,35	10,18	10,39
P02687	S10	K4	R5	1,06	1,49	5,75	6,05	$6,\!52$
P07335	S178	R172	-	1,07	0,79	2,92	1,96	5,46
P07335	S178	R215	-	1,66	1,15	4,69	4,28	10,37
P07335	S178	K177	-	1,99	1,27	$1,\!42$	1,16	10,64
P12839	S609	K591	K603	1,61	1,98	3,94	5,03	9,08
P12839	S609	K603	K591	1,60	1,79	4,74	6,12	8,52
P12839	S609	K608	-	1,93	2,07	3,63	3,62	11,26
P12839	S609	K637	-	1,60	2,14	5,90	3,76	$15,\!36$
P12839	S609	K641	K648	1,68	1,84	2,67	2,66	10,05
P12839	S609	K645	K648	$1,\!47$	$2,\!17$	3,58	$2,\!25$	14,68
P12839	S609	K648	K641	1,59	1,82	3,21	1,99	10,03
P12839	S609	K648	K645	1,59	1,82	3,21	1,96	10,66
P12839	S609	K787	-	1,62	1,86	2,25	2,11	9,50
P12839	S767	R777	-	1,40	1,85	3,55	3,36	9,98
P12839	S767	K726	K727	1,48	1,88	5,53	5,37	10,63

P12839	S767	K727	K726	1,61	1,77	$6,\!25$	3,09	6,38
P12839	S767	K739	-	1,29	1,82	4,62	4,61	7,83
P12839	S767	K781	-	1,36	2,03	5,64	4,87	12,46
P12839	S767	K801	-	1,26	1,79	6,89	3,58	10,75
P12839	S767	K823	-	1,31	1,75	5,20	5,13	9,21
P14598	S359	R301	R302	1,79	1,95	2,51	3,18	13,56
P14598	S359	R302	R301	1,80	2,05	2,94	5,01	17,67
P14598	S359	R354	-	1,47	1,93	4,54	4,40	11,68
P14598	S359	R358	-	1,92	1,94	2,52	2,53	16,18
P14598	S359	K235	-	1,50	1,80	3,24	2,47	9,02
P14598	S359	K360	-	1,50	1,97	4,29	4,29	8,73
P18828	S286	R278	K280	1,13	1,43	2,62	2,48	5,58
P18828	S286	R278	K281	1,13	1,43	2,62	2,48	5,70
P18828	S286	K280	R278	1,04	1,36	2,40	2,58	5,80
P18828	S286	K280	K281	1,04	1,36	2,40	2,60	4,44
P18828	S286	K280	K282	1,04	1,36	2,40	2,58	3,48
P18828	S286	K281	R278	1,04	1,42	2,76	3,29	3,67
P18828	S286	K281	K280	1,04	1,42	2,76	3,34	4,74
P18828	S286	K281	K282	1,04	1,42	2,76	3,28	3,72
P18828	S286	K282	K280	0,81	1,17	1,75	1,46	3,41
P18828	S286	K282	K281	0,81	1,17	1,75	1,48	4,30
P18828	S286	K293	-	1,02	1,37	2,82	2,73	8,17
P23441	S337	R257	R258	1,37	2,03	3,36	2,34	8,55
P23441	S337	R258	R257	1,66	1,94	3,97	2,49	7,79
P26955	S603	R210	-	1,67	1,38	$2,\!15$	$3,\!55$	8,36
P26955	S603	R593	K592	1,63	1,63	1,69	4,63	12,83
P26955	S603	R627	-	1,63	1,44	3,19	5,45	14,07
P26955	S603	R629	-	1,54	$1,\!27$	2,88	5,10	13,16
P26955	S603	R631	-	1,69	1,52	2,63	4,19	10,78
P26955	S603	K592	R593	1,61	1,58	3,04	5,75	11,65
P26955	S603	K616	-	1,95	1,36	$3,\!57$	6,64	13,68
P27732	T443	R727	-	1,84	1,38	$5,\!53$	4,82	11,68
P27732	T443	R730	-	1,79	1,40	6,83	3,96	13,48
P27732	T443	R737	-	1,58	1,26	$4,\!14$	$2,\!54$	9,16
P27732	T443	K733	-	1,97	1,45	$4,\!37$	3,99	14,14
P51398	S185	R177	K175	1,30	1,92	$7,\!32$	7,15	10,51
P51398	S185	R213	K212	1,29	1,65	7,69	6,35	11,36
P51398	S185	K175	R177	1,68	2,11	5,75	5,48	10,52
P51398	S185	K189	K192	1,30	1,56	$4,\!23$	3,59	9,85
P51398	S185	K192	K189	1,53	1,62	3,31	2,17	9,46
P51398	S185	K212	R213	1,33	1,30	8,71	6,78	10,75
P51398	T186	R177	K175	1,31	1,60	8,68	7,76	11,58

P51398	T186	R213	K212	1,67	1,95	5,00	3,48	11,91
P51398	T186	K175	R177	1,72	2,10	$6,\!64$	$7,\!64$	$9,\!33$
P51398	T186	K189	K192	1,59	1,93	2,74	1,98	$11,\!57$
P51398	T186	K192	K189	1,46	1,86	3,37	2,75	$9,\!25$
P51398	T186	K207	-	1,79	1,93	$9,\!56$	9,83	18,45
P51398	T186	K212	R213	1,40	1,58	6,41	7,39	12,31
P60711	S33	R183	-	1,59	1,57	7,67	1,39	12,13
P60711	S33	R37	-	1,89	1,79	3,62	4,35	15,91
P60711	S33	R62	-	1,80	1,62	2,37	2,39	13,78
P60711	S33	K18	-	1,58	1,66	$5,\!52$	2,31	11,64
P60711	S33	K68	-	1,63	1,99	4,19	3,96	13,10
P60711	S33	K84	-	1,94	1,49	6,38	2,03	11,84
P63259	S33	R183	-	1,56	1,52	13,61	1,15	13,20
P63259	S33	R37	-	1,64	1,77	4,87	6,18	15,12
P63259	S33	K18	-	1,68	1,46	5,11	$2,\!23$	9,84
P63259	S33	K68	-	1,78	2,04	3,69	4,49	13,58
P63259	S33	K84	-	1,97	1,32	3,85	2,13	10,40
P63269	S34	R184	-	1,57	1,76	8,24	1,58	13,32
P63269	S34	R38	-	1,81	1,85	$5,\!27$	5,49	17,51
P63269	S34	R63	-	1,85	1,79	2,61	2,72	14,13
P63269	S34	K19	-	1,64	1,48	5,51	1,84	9,86
P63269	S34	K69	-	1,82	1,84	4,60	5,54	12,66
P63269	S34	K85	-	1,93	1,34	6,08	1,76	11,63
P68136	S35	R185	-	1,78	1,50	$6,\!54$	1,46	11,81
P68136	S35	R39	-	1,67	1,86	6,04	5,41	18,97
P68136	S35	R64	-	1,67	1,84	2,82	3,29	14,33
P68136	S35	K20	-	1,81	1,44	6,76	2,62	10,78
P68136	S35	K70	-	1,79	2,10	3,16	3,96	13,32
P68136	S35	K86	-	2,05	1,42	$6,\!14$	3,00	13,77
Q06210	S235	R230	-	0,16	2,38	3,45	3,33	17,49
Q06210	S235	R239	-	0,18	$2,\!37$	3,57	3,98	34,63
Q06210	S235	R246	-	0,16	$2,\!29$	3,38	4,17	24,21
Q06210	S235	R258	-	$0,\!25$	$2,\!16$	$36,\!45$	$41,\!56$	17,20
Q06210	S235	K216	-	$0,\!22$	$2,\!56$	3,78	4,01	$19,\!22$
Q06210	S235	K236	-	0,21	$2,\!55$	3,75	$4,\!24$	16,94
Q06210	S235	K248	-	$0,\!17$	$2,\!28$	3,21	3,89	26,77
Q13522	S67	R30	R31	1,24	2,03	6,23	4,13	$7,\!25$
Q13522	S67	R30	R32	1,24	2,03	$6,\!23$	4,07	6,88
Q13522	S67	R31	R30	1,18	2,02	5,10	3,59	7,71
Q13522	S67	R31	R32	1,18	2,02	5,10	3,59	6,83
Q13522	S67	R32	R30	1,23	1,82	3,97	2,96	8,98
Q13522	S67	R32	R31	1,23	1,82	3,97	3,00	8,47

Q13522	S67	R69	-	1,56	2,08	2,61	2,59	11,98
Q13522	S67	R71	K72	1,34	1,98	$4,\!46$	5,03	10,02
Q13522	S67	R71	K73	1,34	1,98	$4,\!46$	4,96	8,87
Q13522	S67	R76	-	1,06	1,90	$4,\!26$	$4,\!17$	7,05
Q13522	S67	K61	-	1,08	1,92	4,66	3,25	8,09
Q13522	S67	K72	R71	1,29	2,31	4,73	5,24	9,06
Q13522	S67	K72	K73	1,29	2,31	4,73	5,37	7,92
Q13522	S67	K73	R71	1,42	1,99	$4,\!64$	4,77	10,77
Q13522	S67	K73	K72	1,42	1,99	$4,\!64$	4,88	10,77
Q61411	S177	R164	-	1,07	1,05	4,99	3,63	7,40
Q61411	S177	R169	-	0,91	1,00	3,97	3,16	$7,\!42$
Q66H98	S332	R354	-	1,01	1,02	3,16	3,08	10,34
Q66H98	S332	K290	-	1,09	1,13	$5,\!49$	3,51	$6,\!52$
Q66H98	S332	K350	-	1,18	1,10	4,94	5,59	8,97
Q6J4I0	S137	K143	-	1,05	1,63	7,16	5,71	8,54
Q91XW0	S263	R226	-	0,87	0,78	2,60	1,10	$5,\!53$
Q91XW0	S263	K270	K271	1,10	0,79	4,07	3,12	10,13
Q91XW0	S263	K271	K270	1,15	0,80	7,34	5,62	9,70
Q9GZQ8	T29	R21	R24	0,98	0,91	1,10	0,95	4,61
Q9GZQ8	T29	R24	R21	0,99	0,99	2,31	2,22	5,75
Q9GZQ8	T29	K30	-	1,39	1,44	2,83	2,85	9,76
Q9GZQ8	T6	R10	R11	1,21	$1,\!23$	4,29	3,80	9,49
Q9GZQ8	T6	R11	R10	1,16	1,19	3,02	2,68	8,86
Q9GZQ8	T6	R16	-	1,12	1,23	$2,\!56$	1,33	8,54
Q9GZQ8	T6	R37	-	1,20	1,01	2,92	1,73	$7,\!51$
Q9GZQ8	T6	K39	-	1,20	1,14	2,98	1,39	$7,\!56$
Q9GZQ8	T6	K49	-	1,30	1,46	$3,\!57$	1,66	9,88
Q9GZQ8	T6	K51	-	1,24	1,43	3,05	1,41	10,06
Q9GZQ8	T6	K5	-	1,77	1,56	5,74	5,91	10,90
Q9GZQ8	T6	K8	-	1,28	1,39	3,73	3,60	6,63

ANEXO C – Artigo: Protein folding creates structure-based, noncontiguous consensus phosphorylation motifs recognized by kinases

PROTEIN MOTIFS

Protein folding creates structure-based, noncontiguous consensus phosphorylation motifs recognized by kinases

Mariana Lemos Duarte,¹* Darlene Aparecida Pena,¹* Felipe Augusto Nunes Ferraz,²* Denise Aparecida Berti,¹ Tiago José Paschoal Sobreira,² Helio Miranda Costa-Junior,¹ Munira Muhammad Abdel Baqui,³ Marie-Hélène Disatnik,⁴ José Xavier-Neto,² Paulo Sérgio Lopes de Oliveira,² Deborah Schechtman^{1†}

Linear consensus motifs are short contiguous sequences of residues within a protein that can form recognition modules for protein interaction or catalytic modification. Protein kinase specificity and the matching of kinases to substrates have been mostly defined by phosphorylation sites that occur in linear consensus motifs. However, phosphorylation can also occur within sequences that do not match known linear consensus motifs recognized by kinases and within flexible loops. We report the identification of Thr^{253} in α -tubulin as a site that is phosphorylated by protein kinase C β I (PKC β I). Thr²⁵³ is not part of a linear PKC consensus motif. Instead, Thr^{253} occurs within a region on the surface of α -tubulin that resembles a PKC phosphorylation site consensus motif formed by basic residues in different parts of the protein, which come together in the folded protein to form the recognition motif for PKC β I. Mutations of these basic residues decreased substrate phosphorylation, confirming the presence of this "structurally formed" consensus motif and its importance for the protein kinase—substrate interaction. Analysis of previously reported protein kinase A (PKA) and PKC substrates identified sites within structurally formed consensus motifs in many substrates of these two kinase families. Thus, the concept of consensus phosphorylation site motif needs to be expanded to include sites within these structurally formed consensus motifs.

INTRODUCTION

One reason that elucidation of signal transduction pathways is challenging is the difficulty in detecting and predicting specific targets for protein kinases. Among the factors that determine the recognition of a particular substrate by a specific protein kinase is the interaction of the catalytic site of the protein kinase with consensus sequences in the substrates. These consensus sequences are composed of amino acids flanking the phosphorylated residue. These residues often contribute to the specific recognition of the substrate by the protein kinase and aid substrate anchoring, increasing the interaction energy between the substrate and protein kinase, thereby favoring the formation of the complex (1). Within the consensus sequence, the phosphorylated residue on the substrate is referred to as P0 (the phosphorylation site), and the three residues flanking the phosphorylation site on the N and C-terminal sides are referred to as P-3, P-2, P-1 and P+1, P+2, and P+3, respectively. Subsites in the protein kinase that accommodate these seven residues of the peptide substrate are referred to as S-3, S-2, S-1, S0, S+1, S+2, and S+3. Interactions between substrates and protein kinases within these regions help determine protein kinase specificity and increase the interactions between protein kinases and their substrates. Structural modeling confirms that the amino acids flanking the phosphorylated residue are important for determining protein kinase specificity and increasing the interaction energy between a specific substrate and protein kinase (2).

¹Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo 05508000, Brazil. ²Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas 13083-970, Brazil. ³Departamento de Biologia Celular e Molecular, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo 14049-900, Brazil. ⁴Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA.

Screens of peptide libraries for consensus motifs can predict protein kinase specificity (3-5), and phosphorylation sites are frequently predicted by inspecting primary amino acid sequences for linear consensus motifs (6). Computational methods are being used to understand the specificity of binding between substrates and protein kinases. Some of the strategies are based on the use of amino acid sequences of known substrates to identify conserved residues that would reveal a linear consensus motif (7, 8). With this linear consensus motif in hand, similarity searches performed against databases of protein sequences predict substrate candidates. However, when the number of specific known substrates for the target kinase is low, establishing a linear consensus motif is difficult. Similarly, upon identification of previously unknown phosphorylation sites, protein kinases responsible for the phosphorylation events are commonly predicted on the basis of a linear consensus sequence containing the newly identified site (6).

Charge distribution (9) and frequency of disorder (10) have been suggested as important requisites for recognition of protein phosphorylation sites. Electrostatic interactions are likely involved in anchoring substrates to protein kinases, and these may be important for particular sets of kinases, such as the AGC family of kinases (11). However, overall differences in charge distribution cannot distinguish between substrates and nonsubstrates (9). On the basis of an increasing number of reported phosphorylation sites in databases and resolved structures, predictive tools that determine phosphorylation events using three-dimensional (3D) structures are emerging (12, 13). For instance, DISPHOS uses information from disordered regions of the structures to predict the most probable phosphorylated residues (10), whereas NetPhos (14) groups the structures of substrates [available in the Protein Data Bank (PDB)] that were phosphorylated by similar kinases and structurally superimposes the substrates to predict new substrates on the basis of structural similarity. Another search for phosphorylation patterns involves calculating the relative frequencies of the 20 amino acids within

^{*}These authors contributed equally to this work.

[†]Corresponding author. E-mail: deborah@iq.usp.br

radial distances ranging from 2 to 10 Å of the phosphorylated residue. The most frequently occurring amino acids at these distances in substrates of specific protein kinases may contribute to kinase recognition (12, 15). Screening surfaces for kinase interactions using docking methods is an uncommon approach to matching kinases with their substrates. Further knowledge on the nature of the 3D interaction between the kinase and its substrate can aid in the development of predictive tools.

With the advances in chemical biology (16), mass spectrometry (MS), and proteomics (6), new strategies have identified protein kinase–specific substrates in cells (17). Surprisingly, many substrates are phosphorylated on residues that are not within linear consensus motifs (6, 16, 17). Instead, the phosphorylated residues are often found in flexible structures, such as loops, which have been proposed to conformationally adapt to the catalytic site (17). Substrate docking sites in the kinase that are distant from the catalytic site are also important in enhancing substrate-kinase interactions (18, 19).

Members of the protein kinase C (PKC) family of serine/threonine kinases recognize linear consensus motifs with basic residues (Lys or Arg) usually at the P-3 and P-2 positions (20). Indeed, a regulatory mechanism of the PKC family is the presence of a pseudosubstrate region within the enzyme's regulatory domain. This pseudosubstrate region contains an Ala rather than a phosphorylatable Ser or Thr at position P0 (21).

Pharmacological inhibition of PKC in embryonic stem cells decreases α -tubulin phosphorylation (22); however, whether this phosphorylation

event is direct or indirect was not determined. Here, we reported that Thr²⁵³ of α-tubulin is a previously unknown PKC phosphorylation site. Thr²⁵³ is not present within a linear PKC consensus motif formed by flanking basic residues but rather has an acidic amino acid at position P-2. By analyzing structural models, we found that Thr²⁵³ is present within a "structurally formed" consensus motif in which basic residues (Lys¹⁶³ and Lys¹⁶⁴) present in a distant part of the primary sequence are adjacent to the phosphorylation site in the folded protein, thereby producing a structurally formed PKC consensus motif. Mutation of these basic residues diminished α -tubulin phosphorylation at Thr²⁵³ by PKCβI. Previous studies reported that phosphorylated PKC is found at the mitotic spindle of mouse eggs (23). Here, we determined that PKCBI localized at the mitotic spindle of HeLa cells and that phosphorylation of αtubulin at Thr²⁵³ occurred in dividing cells.

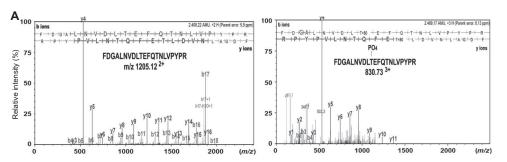
Like PKC, protein kinase A (PKA) also recognizes phosphorylation sites preceded by basic residues (4). By inspecting previously reported PKA or PKC substrates that do not contain linear consensus motifs and by performing molecular docking simulations, we found that phosphorylated sites within structurally formed consensus motifs were also predicted to exist in these substrates: In the folded protein substrates, basic residues distant in the linear sequence come close to the phosphorylated residue and were predicted to interact with residues in the protein kinase, thus promoting a func-

tional interaction between the substrate and the kinase. Our results showed that structurally formed consensus motifs are common and that analysis of linear sequences surrounding phosphorylation sites may reveal only a subset of substrates.

RESULTS

PKC phosphorylates α -tubulin in a site that is not part of a linear consensus motif

α-Tubulin is a highly conserved protein (99.3% identical among 12 vertebrate species), and residue Thr^{253} is among the conserved amino acids (fig. S1). We performed an in vitro protein kinase assay with recombinant rat PKCβI and α-tubulin purified from bovine brain to determine whether α-tubulin was a substrate of PKC. MS analysis of the in vitro kinase reactions indicated that Thr^{253} was phosphorylated by PKCβI (Fig. 1A). To investigate Thr^{253} phosphorylation by PKC in cellular contexts, we generated polyclonal antibodies against a linear peptide, containing the sequence surrounding phosphorylated Thr^{253} in α-tubulin (Fig. 1B). Serum from one mouse (animal 4) [anti–p-α-tubulin (Thr^{253})] had the best relative response for the phosphorylated peptide containing p Thr^{253} (Fig. 1B) and was used for subsequent experiments. In Western blots, anti–p-α-tubulin (Thr^{253}) was significantly more effective at recognizing purified α-tubulin



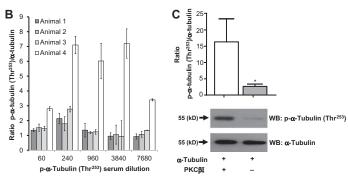


Fig. 1. Thr²⁵³ of α -tubulin is phosphorylated by PKC β I. (A) Purified α -tubulin phosphorylated in vitro by PKC β I in the presence of the activating phospholipids was digested with trypsin, and the peptides were subjected to LC-MS/MS analysis. Left: MS/MS spectrum of the peptide FDGALNVDLTEFQT-NLVPYPR from α -tubulin

without phosphorylation modification (monoisotopic mass of the unprotonated peptide = 2408.22 daltons). Right: MS/MS spectrum of the same peptide with phosphorylation in Thr²⁵³ (monoisotopic mass of the unprotonated peptide = 2489.17 daltons). (B) Recognition of the phosphorylated peptide (GALNVDLpTEFQTNLVP) relative to the nonphosphorylated peptide by antibodies in serum from mice immunized with the phosphorylated peptide. Peptide binding was detected by enzyme-linked immunosorbent assay (ELISA). (C) In vitro phosphorylation of purified α -tubulin by PKC β I detected by Western blot analysis with an antibody recognizing α -tubulin or the serum from animal 4 recognizing p- α -tubulin (Thr²⁵³). Equal amounts of PKC β I-phosphorylated α -tubulin were loaded on separate gels and reacted with either anti-p- α -tubulin (Thr²⁵³) or the antibody recognizing α -tubulin. Quantified data from three independent experiments shown as mean and SD are shown along with representative blots. Statistical significance was determined using Mann-Whitney test; *P = 0.0286.

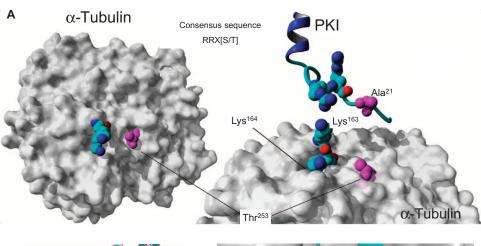
phosphorylated in vitro by PKCβI than nonphosphorylated α -tubulin (Fig. 1C). The phosphorylated peptide also blocked anti–p- α -tubulin (Thr²⁵³) from recognizing PKCβI-phosphorylated α -tubulin more effectively than did the nonphosphorylated peptide (fig. S2). The MS data indicated that the phosphorylated Thr is not flanked by basic amino acids commonly found in PKA (20) and PKC (4) consensus site motifs (R/K R/K X S/T), but rather by an acidic Asp (D) at position P-2 (Fig. 1A).

lpha-Tubulin contains a structurally formed PKC consensus motif

Using the structure of *Bos taurus* (PDB ID: 1JFF) (24), we identified the location of the phosphorylated Thr²⁵³ and found that Lys¹⁶³ and Lys¹⁶⁴, which are also conserved in several vertebrates (fig. S1), were spatially close to the phosphorylated site (Fig. 2A). The distance between the α carbon (C α) of Lys¹⁶⁴ and C α of Thr²⁵³ is 8.12 Å. The structure of PKI, a

peptide inhibitor of PKA (PDB ID: 1FMO) (25), has a similar distance; the distance between C α of Arg¹⁹ and C α of Ala²¹ in PKI is 6.61 Å (Fig. 2A).

To understand at the molecular level how PKC β I may recognize this phosphorylation site in α -tubulin, we modeled the interaction between α -tubulin and PKC β I (Fig. 2, B and C), setting Lys¹⁶³ as P-3, Lys¹⁶⁴ as P-2, and Thr²⁵³ as P0 (see Materials and Methods for the details of the modeling and models S1). Comparison of the interaction of the refined protein kinase model containing the PKI peptide (representing a linear consensus motif) with the structurally formed consensus motif of α -tubulin considering these three key residues of α -tubulin (P-3, P-2, and P0) relative to PKI produced low root mean square deviation (RMSD) values (Table 1), indicating that the structure and the model were similar. Molecular modeling and docking thus suggested that Lys¹⁶³ and Lys¹⁶⁴ could functionally substitute for basic residues in a linear PKC consensus motif, anchoring α -tubulin to PKC.



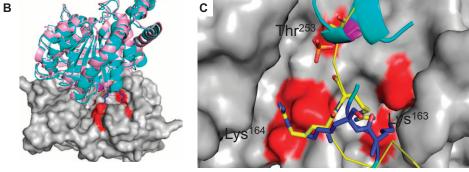


Fig. 2. Analyses of the structurally formed PKC consensus motif containing Thr 253 in α -tubulin. (A) Comparison of PKI, representing a linear consensus motif, and the structurally formed consensus motif comprising Lys 163 , Lys 164 , and Thr 253 . In sphere representation, residues suitable to act as P0 (Thr 253 on α -tubulin and Ala 21 on PKI) are pink. Anchoring residues (Lys 163 and Lys 164 on α -tubulin and Arg 18 and Arg 19 on PKI) are blue. (B) Proposed model for the α -tubulin and PKC complex. PKC is shown in gray, with acid residues forming the S-3, S-2, and S0 substrate interaction regions in red. Docked α -tubulin is shown in cyan cartoon representation, with the anchoring residues (Lys 163 and Lys 164) represented in blue sticks and the phosphorylated Thr 253 in purple. The nondocked original structure of α -tubulin of *B. taurus* [PDB ID: 1JFF] (24) is shown in pink and aligned over the docked one for conformational change comparison. (C) Enlarged image of the interaction region on the docked model shown in (B), with PKI (yellow) for positional reference. Key residues on PKI are represented as sticks (yellow). Blue and pink residues in stick are the key residues on α -tubulin.

Phosphorylated sites are present in structurally formed consensus motifs in multiple PKC and PKA substrates

Analysis of the PhosphositePlus database demonstrated that of 1095 PKA substrates, 140 lack a linear consensus motif for PKA: a basic amino acid at either P-3 or P-2. Of the 140 substrates that did not contain a linear consensus motif, we generated 27 substrate homology models containing 36 nonlinear consensus motifs (table S1). Among these, we used 33 structurally formed consensus motifs to simulate the interaction with PKA. Three of the phosphorylation sites modeled did not display anchoring basic residues: T19 and S2 Q64122 had basic residues in regions that could not be modeled, and S¹⁸² on P11473 did not have anchoring residues. Three representative substrates with two basic residues as anchors and low local RMSD (modeled substrate relative to docked substrate) are shown (Fig. 3 and models S1). Also, three different substrates each with a linear consensus motif were modeled (table S1 and models S1).

Microtubule-associated protein light chain 3 (LC3), a PKA and PKC substrate, has phosphorylated Thr29 (P0) (26) and anchoring basic residues Arg24 (P-2) and Arg²¹ (P-3) (Fig. 3A). Regulator of G protein (heterotrimeric guanine nucleotidebinding protein) signaling (RGS13) (27) is phosphorylated by PKA and has phosphorylated Thr⁴¹ (P0) and anchoring residues Arg¹⁸ (P-2) and Lys¹⁷ (P-3) (Fig. 3B). Death-associated protein 3 (DAP3) (28) is also phosphorylated by PKA and has phosphorylated Ser¹⁸⁵ (P0) and anchoring residues Lys¹⁹² and Lys¹⁸⁹ (Fig. 3C). We aligned the structural consensus with the peptide from PKI, which is a linear consensus motif, to compare the interaction

Table 1. RMSD calculation for PKA-docked models with structurally formed consensus sites, compared to the linear peptide from PKI and the nondocked substrate model. Substrates were divided into two different groups: α -tubulin (Thr²⁵³) and phosphorylated sites without a linear consensus motif as obtained from PhosphositePlus data bank and shown in Fig. 3. The docked substrate represents the model of the substrate docked with PKA and the nondocked substrate starting model, derived from the published crystal structure (table S1). RMSD values are shown for comparisons of the

PKA-docked modeled substrate and modeled nondocked substrate (whole substrate), the 10 residues flanking each side of the phosphorylated residue (p-residue) in the docked and nondocked models, the 10 residues flanking each side of the pair of anchoring residues in docked and nondocked models, and the p-site and anchor residues in the docked models (key residues) compared to the analogous key residues in PKI (P0, P-2, and P-3). For the comparison of the key residues, only $C\alpha$, $C\beta$, and $C\gamma$ when present in the p-site and anchor residues in the substrate were used.

				D	Docked and nondocked substrates (Å)				
Models	Protein	p-Site	Anchors	Whole substrate	Ten residues around the p-residue	Ten residues around anchoring residues	Key residues and PKI		
Tubulin	α-Tubulin	T ²⁵³	K ¹⁶⁵ , K ¹⁶⁴	1.16	2.27	3.18	1.83		
Lacking a linear	LC3	T ²⁹	R^{21} , R^{24}	0.77	1.72	1.66	2.08		
consensus motif	RGS13	T ⁴¹	K ¹⁷ , R ¹⁸	1.55	2.81	3.70	2.38		
	DAP3	S ¹⁸⁵	K ¹⁸⁹ , K ¹⁹²	1.63	1.93	3.60	2.72		

of the kinase with PKI peptide to that of the kinase with the structurally formed consensus motif in LC3, RGS13, and DAP3 (Table 1). RMSD values of PKI and the key residues of the substrates docked with PKA are low, indicating that other substrates besides α-tubulin can adopt a structure similar to a linear consensus sequence. The modeled substrate structures docked to kinase structures were not substantially different than the nondocked substrate structures as determined by RMSD values obtained from comparing the PKA-docked models and nondocked models of the substrates in the regions that interact with the protein kinase (17-19) (Table 1). We performed the same type of analysis with substrates containing linear PKA consensus motifs docked to PKA (fig. S3, A to F), which produced RMSD values relative to PKI that were less than 1 (table S2). These data indicated that the docking process did not cause structural disruption in either the structures of the kinase or the substrate.

Anchoring residues Lys¹⁶³ and Lys¹⁶⁴ are key for Thr²⁵³ phosphorylation by PKCβI

We created a fusion protein between green fluorescent protein (GFP) and α -tubulin (GFP- α -tubulin) to assess the importance of the structurally formed consensus motif in α -tubulin for phosphorylation by PKC and function in cells. We individually mutated Lys¹⁶³ and Lys¹⁶⁴ to alanine in GFP- α -tubulin to confirm that these residues are necessary for PKC phosphorylation of α -tubulin. As a control, we also generated a T253A mutant. When expressed in HeLa cells, the GFP- α -tubulin proteins, including the T253A, K163A, and K164A mutants,

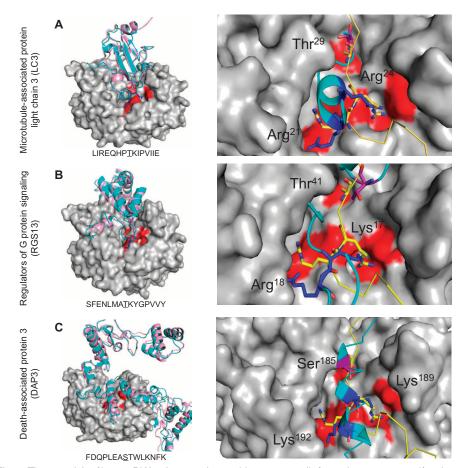


Fig. 3. Three models of known PKA substrates that could use structurally formed consensus motifs to interact with PKA. (A to C) Modeled PKA substrates containing structurally formed consensus motifs: LC3 (A), RGS13 (B), and DAP3 (C). Linear sequences flanking the phosphorylated site (p) obtained from PhosphositePlus data bank are displayed. Left: PKA is shown in gray, with acid residues forming the S-3, S-2, and S0 substrate interaction regions in red; the docked substrates are shown in cyan cartoon representation, with interacting residues represented as sticks in blue and the phosphorylated residue in purple stick. The original structure of the substrate aligned over the docked one is shown in pink. Right: Enlarged images of the interaction region on the docked models, with PKI (yellow) for positional reference. Labels identify the key residues (phosphorylated site and anchor residues) on the substrate.

were incorporated into tubulin fibers (Fig. 4A). To assess the importance of the key residues for PKC-mediated phosphorylation, we immunoprecipitated the GFP- α -tubulin wild type and mutants from human embryonic

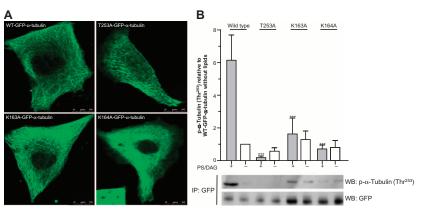


Fig. 4. Lys¹⁶³ and Lys¹⁶⁴ contribute to PKC phosphorylation of α -tubulin. (A) HeLa cells were transfected with the indicated GFP- α -tubulin constructs. Fixed cells were stained with the antiserum recognizing GFP, and fluorescence was detected by confocal microscopy. Images are representative of all cells that were GFP-positive. (B) Immunoprecipitated GFP- α -tubulin proteins from transfected HEK 293T cells were phosphorylated by PKC β I in vitro. Samples were analyzed by Western blotting with the antibodies or serum recognizing p- α -tubulin (Thr²⁵³) or α -tubulin. The amount of GFP- α -tubulin immunoprecipitated was determined by running 10% of the reaction on a separate gel and blotting with an antibody recognizing GFP. A representative experiment (n = 3) is shown. The graph showing the average and SD indicates the amount of p- α -tubulin (Thr²⁵³) relative to the amount detected with anti-p- α -tubulin (Thr²⁵³) in assays with the kinase and wild-type (WT) GFP- α -tubulin in the absence of lipids. Statistical analysis was performed using two-way analysis of variance (ANOVA); ****P < 0.01.

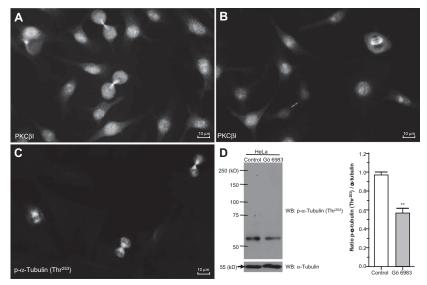


Fig. 5. PKC β I Thr²⁵³ phosphorylation is enriched at the mitotic spindles. (A and B) PKC β I detected at the mitotic spindle in HeLa cells at different stages of cell division. (C) Phosphorylated α -tubulin (Thr²⁵³) detected in dividing HeLa cells with anti–p- α -tubulin (Thr²⁵³). (D) Effect of inhibition of PKC on α -tubulin Thr²⁵³ phosphorylation in HeLa cells exposed to 10 μ M Gö 6983 for 6 hours. Lysates were assayed by Western blot by probing for p- α -tubulin (Thr²⁵³) and α -tubulin. The plot shows the average and SD of the ratio between p- α -tubulin (Thr²⁵³) and α -tubulin (n = 3). Blots from a representative experiment are also shown. Statistical analysis was determined using the nonparametric Mann-Whitney test; **P < 0.05.

kidney (HEK) 293T cells with an antibody recognizing GFP and then used the immunoprecipitated proteins for in vitro kinase assays with recombinant rat PKC β I. Anti–p- α -tubulin (Thr²⁵³) detected phosphorylated α -tubulin in

wild-type GFP- α -tubulin (Fig. 4B). However, phosphorylation was undetectable for the T253A mutant and was significantly decreased by mutation of either Lys¹⁶³ or Lys¹⁶⁴, confirming the importance of these lysine residues for Thr²⁵³ phosphorylation by PKC β I (Fig. 4).

$\alpha\text{-Tubulin Thr}^{253}$ phosphorylation occurs in cells during mitosis

Direct protein substrates have to physically interact with protein kinases in the context of living cells. We found that PKC β I localized mainly at the mitotic spindle during different stages of cell division in HeLa cells (Fig. 5, A and B). In particular, PKC β I accumulated on the spindle of cells in the later stages of mitosis. Similarly, staining cells with anti–p α -tubulin (Thr²⁵³) indicated that phosphorylation of Thr²⁵³ occurred mainly at the mitotic spindle (Fig. 5C). Exposing HeLa cells to Gö 6983, a general PKC inhibitor, significantly decreased α -tubulin phosphorylation at Thr²⁵³ detected by Western blot with anti–p- α -tubulin (Thr²⁵³), providing additional support for PKC-mediated phosphorylation of this site on α -tubulin in cells (Fig. 5D).

DISCUSSION

Few PKC substrates have been experimentally identified thus far, which relates to difficulties imposed by the biological complexity of these kinases and the low abundance of some of the substrates. The computational strategies commonly used to identify PKC targets search for linear PKC consensus motifs (29), despite reports of many protein kinase substrates that lack linear consensus motifs (17). Here, our results demonstrated that the strategy of looking for phosphorylation sites only within linear consensus motifs may exclude several substrates.

We identified a previously unknown PKC phosphorylation site in α -tubulin, Thr²⁵³. This site is present within a motif formed by residues that, although placed at distant parts of the primary sequence, in the folded protein are arranged such that they mimic the recognition sequence of a linear consensus motif-containing peptide. This PKC phosphorylation site and others that are part of structurally formed consensus motifs cannot be identified by searching for linear PKC consensus motifs composed of basic residues flanking the phosphorylated residue. Indeed, in the linear sequence containing Thr²⁵³ at P-2, there is a negatively charged Asp, which is incompatible with the electrostatic requirements of a PKC substrate. Furthermore, inspection of the 3D structure of α -tubulin showed that basic residues Lys¹⁶³ and Lys¹⁶⁴ are in close proximity to Thr²⁵³ and that, structurally, this environment resembles the PKA peptide inhibitor, PKI, which contains a linear PKA/C consensus motif (25). We detected PKCβI and Thr²⁵³-phosphorylated α-tubulin at the mitotic spindle in dividing cells, and pharmacological inhibition of PKC in HeLa cells decreased Thr²⁵³ phosphorylation, which is consistent with another study that also reported phosphorylation of Thr²⁵³ in α -tubulin of synchronized HeLa cells (6). Thus, this phosphorylation

event occurs in cells during mitosis. Future studies need to evaluate the physiological importance of this phosphorylation.

Phosphorylation in flexible regions that would adapt to fit into the catalytic site of a kinase has been suggested as a reason for phosphorylation occurring in the absence of a linear consensus motif (17). However, comprehensive structural analysis of experimentally identified phosphorylation sites demonstrated that 37% of the phosphorylation sites mapped to protein regions displaying a well-defined secondary structure (13).

In addition to promoting the binding of the substrate to the catalytic site, substrate docking sites located at a distance from the catalytic site are also important for defining substrate specificity (18, 19). On the basis of the structure of the kinase domain (2) and analysis of phosphorylated substrates deposited in databases (15), researchers have suggested that kinases could recognize substrates in a structural manner. Indeed, our data validated the importance of consensus sequences in substrate binding to protein kinases and extended this notion by showing that tertiary structure contributes to structurally formed consensus motifs. Therefore, both linear consensus motifs and structurally formed consensus motifs should be considered when analyzing and predicting protein kinase phosphorylation sites.

Our data indicate that phosphorylation events could be mediated by recognition of a structurally formed consensus motif. Because usually protein kinases interact with substrates in a very similar fashion and substrates can be phosphorylated on sites not within linear consensus motifs, we expect that the recognition of substrates with structurally formed consensus motif extends beyond the PKC and PKA families.

MATERIALS AND METHODS

Cell culture

HEK 293T cells, obtained from B. Malnic (University of São Paulo, Instituto de Química), and HeLa cells, obtained from R. Tonelli (Federal University of São Paulo), were maintained in high-glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and penicillin/streptomycin (Gibco-BRL) at 37°C under 5% CO₂.

In vitro phosphorylation of α -tubulin and identification of the phosphorylated residue

Purified bovine tubulin (TL238, Cytoskeleton) was phosphorylated in vitro with rat recombinant PKCβI (Santa Cruz Biotechnology) as previously described (*30*) but without [³²P]adenosine triphosphate (ATP). The phosphorylation site was identified by liquid chromatography–tandem MS (LC-MS/MS) performed at Stanford University MS facility (https://mass-spec.stanford.edu/). Phosphorylated protein was precipitated with 4 volumes of −80°C acetone. Protein pellets were reconstituted in 15 μl of 8 M urea (pH 7.8) and 20 μl of 0.2% Protease Max (Promega) per manufacturer's protocol. Samples were solubilized by shaking and sonication at room temperature. Solubilized material was reduced for a half hour at 55°C using dithiothreitol (5 mM); samples were cooled to room temperature, alkylated with iodoacetamide [10 mM and 0.1 μg of sequencing-grade trypsin (Promega)], and digested overnight at 37°C. Samples were then acidified with formic acid and purified using in-house packed STAGE tips.

For LC-MS/MS analysis and identification of phosphorylated residues, liquid chromatography was done using Eksigent nano2D LC coupled to an in-house packed C18 analytical column. A Michrom source was used for nano–electrospray ionization at a flow rate of 750 nl/min. A 1-hour linear gradient from 2% mobile phase B (acetonitrile with 0.1% formic acid) to 40% B was used. Data were acquired using an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher) in a data-dependent fashion (MS/MS via higher-energy C-trap dissociation) of the top 10 most intense

precursor ions. RAW files were converted to mzXML format and searched on a Sorcerer (SageN) processing system using Sequest against the appropriate SwissProt and NCBI (National Center for Biotechnology Information) databases. Search results were uploaded into a Scaffold workstation

Generation and characterization of antibodies recognizing $p-\alpha$ -tubulin (Thr²⁵³)

Mice were immunized with p- α -tubulin (Thr²⁵³) peptides obtained from Proteimax. Coupling of both phosphorylated (GALNVDLpTEFQTNLVP) and nonphosphorylated (GALNVDLTEFQTNLVP) peptides to keyhole limpet hemocyanin (KLH) (Sigma-Aldrich) or to bovine serum albumin (BSA) was performed with formaldehyde according to the manufacturer's instruction. We immunized four 8-week-old female BALB/c mice from Biotério de Produção e Experimentação da Faculdade de Ciências Farmacêuticas e do Instituto de Química da USP (housed according to the specifications of the Guide for the Care and Use of Laboratory Animals). Emulsion of the phosphorylated peptide coupled to KLH (50 µg) and complete Freund's adjuvant was administered intraperitoneally, followed by three booster immunizations at 7-day intervals with the emulsion prepared using incomplete Freund's adjuvant. Pre- and postimmunization blood samples were collected, and the antibody titer was determined by ELISA. One hundred microliters of nonphosphorylated or phosphorylated peptide coupled to BSA (100 ng peptide, assuming that all 2.5 mg of peptide efficiently coupled to 5 mg BSA) diluted in phosphate-buffered saline (PBS; pH 7.4) was added to each well of 96-well ELISA plates and incubated overnight at 4°C. Plates were washed three times with PBS plus 0.05% Tween 20 (PBS/Tween) and blocked with 5% nonfat milk in PBS for 1 hour at 37°C. After the addition of 100 µl of serial serum dilutions, the plates were incubated for 2 hours at 37°C, then washed three times with PBS/Tween. After the addition of 100 µl of horseradish peroxidase conjugated goat anti-mouse immunoglobulin G (Sigma-Aldrich), the plates were incubated for 1 hour at 37°C and washed three times in PBS/Tween; 100 μl of TMB (3,3′,5,5′-tetramethylbenzidine) substrate (BD Biosciences) was subsequently added, and absorbance was read at 655 nm. To identify the serum that best differentially recognized the phosphorylated peptide, we calculated the ratio of the absorbance in the presence of the nonphosphorylated to that in the presence of the phosphorylated peptide, and the serum from a mouse that recognized the phosphorylated peptide at lower dilutions than those required to detect the nonphosphorylated peptide [anti-p-\alpha-tubulin (Thr²⁵³)] was determined and used for Western blot and immunofluorescence experiments at 1:1000 and 1:100, respectively. The specificity of anti–p-αtubulin (Thr²⁵³) was evaluated by preadsorption of this antibody with 100 µM immunization peptide containing a phosphorylated Thr²⁵³ or the nonphosphorylated peptide. After 2 hours of incubation at 37°C, blocked antibodies were used in Western blot assays with in vitro PKC-phosphorylated α-tubulin.

Docking, structural alignments, alignments, energy minimization, and molecular dynamics

To explore the effects of 3D structure surrounding the phosphorylation sites in proteins, we obtained validated PKA substrates from the phosphorylation data bank PhosphositePlus (search performed on 2 July 2014). We chose PKA substrates because the crystal structure of this protein kinase has been solved in complex with the inhibitor peptide PKI (PDB ID: 1FMO) (25), which we used as reference to model all kinase-substrate interactions. All substrate structures were generated through homology modeling using YASARA software (31). A geometrical search was performed on all the substrates, taking as starting reference the residue known to be phosphorylated and searching for exposed residues matching the phosphorylation motif,

which we considered to be basic residues within 14 Å of the phosphorylated residue. Distance restrictions were made on the basis of the structure of PKI. An initial kinase-substrate complex model was made by aligning key residues on the protein over the P0 and P-2 onto the structure of PKI in complex with PKA. We moved this initial model by rotating the protein kinase on the vector defined by the $C\alpha$ of these residues, searching for the conformation that minimized atom overlaps. All the structural alignments for model generation and RMSD calculations were made using YASARA (31).

Models that presented a possible structural phosphorylation motif were refined through steered energy minimization. Taking as starting point the noninteracting (independent or unbound) structures of the kinase and the substrate, we tried to achieve the conformation created by overlapping the key residues of the substrate onto PKI. The low-energy pathway between the two conformations was calculated with YASARA through a restrained simulated annealing minimization. We applied a force of 25 pN pointing toward the target coordinates to all atoms. We then evaluated the final model by comparing the docked and nondocked structures. The energy minimization used YAMBER3 force field (31). This steered docking process was used to evaluate whether the region of the substrate containing the phosphorylated residue can access and interact with the binding site of the protein kinase without hindrance or requiring large conformational changes to both structures. Changes in the structures were evaluated by calculating the global and local (around the key residues) RMSD values relative to the structure of the starting model of the substrate.

To generate images, we refined the models using a steered molecular simulation that included springs connecting the side chains of key residues of the substrate protein, the ones proposed to act as anchors and the phosphorylated residue, with residues forming the subsites, the substrate docking residues in the kinase. Connections were made using springs with a stretching constant of 30 N/m and an equilibrium length of 3.5 Å. Atoms on the backbone of the protein kinase were fixed, forcing only the substrate structure to adjust to the docking conformation, and the dynamics ran until the springs reached the equilibrium state. Subsite residues on PKA were extracted from the literature: Tyr³³⁰ and Glu¹²⁷ for S-3; Glu²³⁰, Tyr²⁰⁴, Glu²⁰³, Thr²⁰¹, and Glu¹⁷⁰ for S-2; and Asp¹⁶⁶ for S0 (2).

To generate the α -tubulin–PKC complex, we followed the same process described above, using the α -tubulin structure of B. taurus (PDB ID: 1JFF) (24) as the substrate and a PKC β I structure built through homology modeling, on the basis of the sequence of the protein kinase domain of a rat PKC β I (UniProt: P68403, CDD: 214567) using [PDB ID: 1FMO (25)] as template. The template presented a coverage of 96.1% and an identity of 43.8%. Alignment of the PKA and PKC sequences was performed using Clustal X (32), and the analogous residues forming subsites on the PKC structure were determined.

Plasmids and site-directed mutagenesis

Human α-tubulin from the GFP-α-tubulin vector (Evrogen) was subcloned into the mammalian expression vector pEGFP-CI (Clontech) between the restriction enzyme sites Xho I and Bam HI. Site-directed mutagenesis was performed using the Stratagene QuikChange kit according to the manufacturer's instructions. Primers for construction of the mutants were as follows: T253A mutation forward 5′-TGAATGTTGACCTGGCAGAATTCCAGACCAAC-3′ and reverse 5′-GTTGGTCTGGAATTCTGCCAGGTCAACATTCA-3′; K163A mutation forward 5′-GTTGATTATGGCGCGAAGTCCAAGCTGAGT-3′ and reverse 5′-ACTCCAGCTTGGACTTCGCCATAATCAAC-3′. K164A mutation forward 5′-GTTGATTATGGCAAGGCGTCCAAGCTGGAGTGAGT-3′ and reverse 5′-ACTCCAGCTTGGACCTTGCCATAATCAAC-3′. Selected clones were sequenced with BigDye Terminator v3.1 Cycle Sequencing Kits (Life Technologies) at the sequencing facility at Universidade de São Paulo, Instituto de Química with an ABI 3730 DNA Analyzer. DNA from

clones containing the specific desired mutations was prepared using a Qiagen Midi Prep kit.

Transfection of HeLa and HEK 293T cells

Transfections were performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. For immunofluorescence studies, HeLa cells were transfected for 48 hours, then fixed with 4% paraformal-dehyde, and labeled with an antibody recognizing GFP as described below.

Immunoprecipitation of GFP- α -tubulin and in vitro phosphorylation

For immunoprecipitation assays, HEK 293T cells were used; 1 µl of antiserum against GFP (33), donated by F. Gueiros Filho (University of São Paulo, Instituto de Química), was incubated with 50 µl of packed volume of protein G beads (Invitrogen) for 2 hours at 4°C. Antibody-bound beads were then washed twice with PBS and blocked with 1% BSA for 1 hour at 4°C. Cells transfected with GFP-α-tubulin or the three mutants were lysed in PBS (pH 7.4) containing both protease inhibitor cocktail (Sigma-Aldrich) and phosphatase inhibitor cocktail PhosStop (Roche) by three freeze-thaw cycles. Cells were then sonicated for 30 min at 80 Hz (output) with a probe sonicator (Branson Sonifier 250). Cell lysates were precleared with protein G beads for 1 hour at 4°C, incubated with antibody-bound beads overnight at 4°C, and subsequently washed with PBS. Bound proteins were used in phosphorylation assays with PKCβI. Immunoprecipitated proteins were mixed with 80 ng of PKCBI, 40 µM ATP, 40 mM MgCl₂, 2 mM CaCl₂, phosphatidylserine (PS; 60 μg/ml), dioleoylglycerol (DG; 2 μl/ml) in 20 mM tris-HCl (pH 7.4) at 37°C for 20 min. The reactions were stopped by addition of Laemmli buffer. Phosphorylated immunoprecipitates were analyzed on two different 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE), one in which 15% of the immunoprecipitated amount was transferred and probed for GFP and the other 75% was transferred and probed with anti–p-α-tubulin (Thr²⁵³). Densitometric analyses were performed with ImageJ. The GFP signal was used to normalize the amount of transfected immunoprecipitated GFP-α-tubulin. Because transfection efficiency varied for each construct, normalization was performed separately for each construct. Bands were quantified, and the ratio of p- α -tubulin (Thr²⁵³) to α -tubulin in immunoprecipitates of the wild-type GFP- α -tubulin without the PKC activators CaCl2, PS, and DG was determined for each construct. Statistical analysis was performed on data from at least three experiments using two-way ANOVA.

Immunofluorescence

For immunofluorescence studies, cells were cultured on 13-mm glass coverslips coated with 3% gelatin. Cells at 80% confluence were fixed with 4% paraformaldehyde, permeabilized with PBS–0.1% Triton X-100, and blocked in blocking solution (PBS, 0.1% Triton X-100, 1% normal goat serum) for 40 min at room temperature in a humidified chamber. Cells were incubated overnight at 4°C in a humidified chamber with a mouse antibody recognizing PKC β I (sc-8049, Santa Cruz Biotechnology), at a final concentration of 2 μ g/ml, or the serum with anti–p- α -tubulin-Thr²⁵³ diluted 1:100 in blocking solution, or with rabbit anti-GFP serum (1:1000), followed by incubation in the dark for 1 hour at room temperature with goat anti-mouse conjugated with Alexa 555 (4 μ g/ml) or goat anti-rabbit conjugated with Alexa 488 (4 μ g/ml). Immunofluorescence staining was analyzed using a fluorescence microscope (Nikon Eclipse E600).

Transfected cells were labeled with the antibody recognizing GFP, diluted 1:1000, and visualized by Alexa 488–conjugated goat anti-rabbit (Molecular Probes) (8 µg/ml) diluted in blocking solution. Fluorescence was detected using a Leica TCS SP5 laser scanning confocal microscope (Leica Microsystems) in Laboratório de Microscopia Confocal, FMRP-USP.

Inhibition of PKC in cells

HeLa cells were plated in high-glucose DMEM containing 10% fetal bovine serum (FBS), and 24 hours after plating, medium was changed to serum-free DMEM to partially synchronize the cells. After serum starvation for 24 hours, cells were released into cell cycle by the addition of 10% FBS and then incubated in the presence or absence of 10 µM Gö 6983 (Sigma-Aldrich) for 6 hours. Lysates were prepared in Laemmli buffer, run on 10% SDS-PAGE, transferred, and probed for p- α -tubulin (Thr²⁵³) and α -tubulin. Densitometric analysis was performed using ImageJ v.1.46, and data were plotted as the ratio between p- α -tubulin (Thr²⁵³) and α -tubulin. Statistical significance of at least three experiments was determined using the Mann-Whitney test.

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/7/350/ra105/DC1 Fig. S1. Thr^{253}, Lys^{163}, and Lys^{164} are conserved in vertebrate α -tubulin.

Fig. S2. Specificity of the polyclonal antiserum recognizing α -tubulin phosphorylated on Thr 253 Fig. S3. Models of known PKA substrates phosphorylated within sites in linear consensus motifs. Table S1. Substrates of PKA obtained from PhosphositePlus modeled through homology modeling.

Table S2. RMSD calculation for PKA-docked models with linear consensus motifs, compared to the linear peptide and the nondocked substrate model.

Models S1. PDB files of the models of the three substrates with linear consensus motifs, three substrates with structurally formed consensus motifs, and the α -tubulin-PKC complex.

REFERENCES AND NOTES

- 1. B. E. Kemp, D. B. Bylund, T. S. Huang, E. G. Krebs, Substrate specificity of the cyclic AMP-dependent protein kinase. Proc. Natl. Acad. Sci. U.S.A. 72, 3448-3452
- 2. B. Kobe, T. Kampmann, J. K. Forwood, P. Listwan, R. I. Brinkworth, Substrate specificity of protein kinases and computational prediction of substrates. Biochim. Biophys. Acta 1754, 200-209 (2005).
- 3. J. E. Hutti, E. T. Jarrell, J. D. Chang, D. W. Abbott, P. Storz, A. Toker, L. C. Cantley, B. E. Turk, A rapid method for determining protein kinase phosphorylation specificity. Nat. Methods 1, 27-29 (2004).
- 4. Z. Songyang, S. Blechner, N. Hoagland, M. F. Hoekstra, H. Piwnica-Worms, L. C. Cantley, Use of an oriented peptide library to determine the optimal substrates of protein kinases. Curr. Biol. 4, 973-982 (1994).
- 5. T. B. Trinh, Q. Xiao, D. Pei, Profiling the substrate specificity of protein kinases by on-bead screening of peptide libraries. Biochemistry 52, 5645-5655 (2013).
- 6. A. N. Kettenbach, D. K. Schweppe, B. K. Faherty, D. Pechenick, A. A. Pletnev, S. A. Gerber, Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. Sci. Signal. 4. rs5 (2011).
- 7. N. M. Alto, S. H. Soderling, N. Hoshi, L. K. Langeberg, R. Fayos, P. A. Jennings, J. D. Scott, Bioinformatic design of A-kinase anchoring protein-in silico: A potent and selective peptide antagonist of type II protein kinase A anchoring. Proc. Natl. Acad. Sci. U.S.A. 100, 4445-4450 (2003).
- 8. M. Hjerrild, A. Stensballe, T. E. Rasmussen, C. B. Kofoed, N. Blom, T. Sicheritz-Ponten. M. R. Larsen, S. Brunak, O. N. Jensen, S. Gammeltoff, Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry. J. Proteome Res. 3, 426-433 (2004).
- 9. J. Kitchen, R. E. Saunders, J. Warwicker, Charge environments around phosphorylation sites in proteins. BMC Struct. Biol. 8, 19 (2008).
- 10. L. M. lakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, A. K. Dunker, The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32, 1037-1049 (2004).
- 11. C. S. Gibbs, M. J. Zoller, Identification of electrostatic interactions that determine the phosphorylation site specificity of the cAMP-dependent protein kinase. Biochemistry **30**, 5329–5334 (1991).
- 12. P. Durek, C. Schudoma, W. Weckwerth, J. Selbig, D. Walther, Detection and characterization of 3D-signature phosphorylation site motifs and their contribution towards improved phosphorylation site prediction in proteins. BMC Bioinformatics 10, 117 (2009).
- 13. A. Zanzoni, D. Carbajo, F. Diella, P. F. Gherardini, A. Tramontano, M. Helmer-Citterich, A. Via, Phospho3D 2.0: An enhanced database of three-dimensional structures of phosphorylation sites. Nucleic Acids Res. 39, D268-D271 (2011).
- 14. N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294, 1351-1362 (1999).
- 15. M. Su, K. Huang, C. Tung, Y. Lee, A new scheme to predict kinase-specific phosphorylation sites on protein three-dimensional structures, Int. J. Biosci. Biochem. Bioinformatics 3. 473-478 (2013)

- 16. J. D. Blethrow, J. S. Glavy, D. O. Morgan, K. M. Shokat, Covalent capture of kinasespecific phosphopeptides reveals Cdk1-cyclin B substrates. Proc. Natl. Acad. Sci. U.S.A. **105**. 1442-1447 (2008).
- 17. L. N. Johnson, Substrates of mitotic kinases. Sci. Signal. 4, pe31 (2011).
- 18. R. M. Biondi, A. R. Nebreda, Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. Biochem. J. 372, 1-13 (2003).
- 19. J. A. Ubersax, J. E. Ferrell Jr., Mechanisms of specificity in protein phosphorylation. Nat. Rev. Mol. Cell Biol. 8, 530-541 (2007).
- 20. J. R. Woodgett, K. L. Gould, T. Hunter, Substrate specificity of protein kinase C. Use of synthetic peptides corresponding to physiological sites as probes for substrate recognition requirements. Eur. J. Biochem. 161, 177-184 (1986).
- 21. C. House, B. E. Kemp, Protein kinase C contains a pseudosubstrate prototope in its regulatory domain. Science 238, 1726-1728 (1987).
- 22. H. M. Costa-Junior, N. M. Garavello, M. L. Duarte, D. A. Berti, T. Glaser, A. de Andrade, C. A. Labate, A. T. Ferreira, J. E. Perales, J. Xavier-Neto, J. E. Krieger, D. Schechtman, Phosphoproteomics profiling suggests a role for nuclear BIPKC in transcription processes of undifferentiated murine embryonic stem cells. J. Proteome Res. 9, 6191–6206
- 23. Z. Y. Zheng, Q. Z. Li, D. Y. Chen, H. Schatten, Q. Y. Sun, Translocation of phosphoprotein kinase Cs implies their roles in meiotic-spindle organization, polar-body emission and nuclear activity in mouse eggs. Reproduction 129, 229-234 (2005).
- 24. J. Löwe, H. Li, K. H. Downing, E. Nogales, Refined structure of $\alpha\beta\text{-tubulin}$ at 3.5 Å resolution, J. Mol. Biol. 313, 1045-1057 (2001).
- N. Naravana, S. Cox, S. Shaltiel, S. S. Taylor, N. Xuong, Crystal structure of a polyhistidinetagged recombinant catalytic subunit of cAMP-dependent protein kinase complexed with the peptide inhibitor PKI(5-24) and adenosine. Biochemistry 36, 4438-4448 (1997).
- 26. H. Jiang, D. Cheng, W. Liu, J. Peng, J. Feng, Protein kinase C inhibits autophagy and phosphorylates LC3. Biochem. Biophys. Res. Commun. 395, 471-476 (2010).
- 27. Z. Xie, Z. Yang, K. M. Druey, Phosphorylation of RGS13 by the cyclic AMP-dependent protein kinase inhibits RGS13 degradation. J. Mol. Cell. Biol. 2, 357-365 (2010).
- J. L. Miller, H. Koc, E. C. Koc, Identification of phosphorylation sites in mammalian mitochondrial ribosomal protein DAP3. Protein Sci. 17, 251-260 (2008).
- 29. T. P. Abeyweera, X. Chen, S. A. Rotenberg, Phosphorylation of α 6-tubulin by protein kinase $C\alpha$ activates motility of human breast cells. J. Biol. Chem. 284, 17648–17656 (2009)
- 30. D. Schechtman, M. L. Craske, V. Kheifets, T. Meyer, J. Schechtman, D. Mochly-Rosen, A critical intramolecular interaction for protein kinase Cε translocation. J. Biol. Chem. 279, 15831-15840 (2004).
- 31. E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, G. Vriend, Making optimal use of empirical energy functions: Force-field parameterization in crystal space. Proteins 57, 678-683 (2004).
- 32. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948 (2007).
- 33. D. Z. Rudner, R. Losick, A sporulation membrane protein tethers the pro- ζ^K processing enzyme to its inhibitor and dictates its subcellular localization. Genes Dev. 16. 1007-1018 (2002).

Acknowledgments: We would like to thank C. Adams (Stanford University Mass Spectrometry, Stanford, CA) for mass spectroscopic analysis: B. Malnic and R. Tonelli for cell lines; M. J. Manso Alves, W. Colli, R. I. Schumacher, H. Chaimovich, J. Kobara. and A. Gordon for editing the manuscript; and F. Gueiros Filho for antibodies recognizing GFP and for inspiring discussions. Funding: H.M.C.-J. was a recipient of a FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) postdoctoral fellowship (2006/52062-6); M.L.D. and D.A.P. were recipients of FAPESP doctoral fellowships (2009/51337 and 2011/10321-3, respectively); D.A.B., a FAPESP postdoctoral fellowship (2010/15424-2); and M.-H.D., NIH grant HL52141. This research was supported by Brazilian agency grants FAPESP 2010/18640-8 and 2012/24154-4 to D.S. Author contributions: M.L.D. D.A.P., F.A.N.F., H.M.C.-J., M.-H.D., T.J.P.S., and P.S.L.d.O. performed the experiments. D.S., M.L.D., D.A.P., F.A.N.F., P.S.L.d.O., M.M.A.B., and J.X.-N. planned the experiments and analyzed the data. D.S., P.S.L.d.O., J.X.-N., M.M.A.B., and F.A.N.F. wrote the paper. All authors read and approved the paper. Data and materials availability: Three substrates with linear consensus motifs, three substrates with structurally formed consensus motifs, and the α -tubulin–PKC complex models (discussed in the manuscript) are available as PDB files in models S1.

Submitted 23 April 2014

Accepted 17 October 2014

Final Publication 4 November 2014

10.1126/scisignal.2005412

Citation: M. L. Duarte, D. A. Pena, F. A. N. Ferraz, D. A. Berti, T. J. P. Sobreira, H. M. Costa-Junior, M. M. A. Baqui, M.-H. Disatnik, J. Xavier-Neto, P. S. L. de Oliveira, D. Schechtman, Protein folding creates structure-based, noncontiguous consensus phosphorylation motifs recognized by kinases. Sci. Signal. 7, ra105 (2014).